Pre-diagnostic alterations in circulating bile acid profiles in the development of hepatocellular carcinoma

Short Title: Bile acid profiles and hepatocellular carcinoma risk

Magdalena Stepien, PhD [1], Marina Lopez-Nogueroles, PhD [2], Agustin Lahoz, PhD [2], Tilman Kühn, PhD [3], Gabriel Perlemuter, MD, PhD [4 5 6], Cosmin Voican, MD PhD [4 5 6], Dragos Ciocan, MD [4 5 6], Marie-Christine Boutron-Ruault, MD, PhD [7 8], Eugene Jansen, PhD [9], Vivian Viallon, PhD [1], Michael Leitzmann, MD, PhD [10], Anne Tjønneland, MD, PhD [11], Gianluca Severi, PhD [7 8], Francesca Romana Mancini, PhD [7 8], Catherine Dong, PhD [7 8 12], Rudolf Kaaks, PhD [3], Renee Turzanski Fortner, PhD [3] , Manuela M. Bergmann, PhD [13], Heiner Boeing, PhD [13], Antonia Trichopoulou, MD, PhD [14], Anna Karakatsani, MD, MPH, PhD [14 15], Eleni Peppa, MD [14], Domenico Palli, MD, PhD [16], Vittorio Krogh, MD [17], Rosario Tumino, MD [18], Carlotta Sacerdote, PhD [19], Salvatore Panico, MD, PhD [20], H. Bas Bueno-de-Mesquita, MD, MPH, PhD [21 22 23 24], Guri Skeie, PhD [25], Susana Merino, PhD [26], Raul Zamora Ros, PhD [27], Maria Jose Sánchez, PhD [28 29], Pilar Amiano, PhD [29 30] , Jose Mª Huerta, PhD [29 31], Aurelio Barricarte, PhD [29 32 33], Klas Sjöberg, MD [34], Bodil Ohlsson, MD [35], Hanna Nyström, PhD [36], Marten Werner, MD [37], Aurora Perez-Cornago, PhD [38], Julie A. Schmidt, PhD [38], Heinz Freisling, PhD [1], Augustin Scalbert, PhD [1], Elisabete Weiderpass, MD, MPH, PhD [39], Sofia Christakoudi, MD [40, 41], Marc J. Gunter, PhD [1], Mazda Jenab, PhD [1*]

1. Section of Nutrition and Metabolism, International Agency for Research on Cancer (IARC-WHO), Lyon, France
2. Analytical Unit, Health Research Institute Hospital La Fe, Valencia, Spain
3. Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
4. INSERM U996, Intestinal Microbiota, Macrophages and Liver Inflammation, DHU Hepatinov, Labex LERMIT, Clamart, France
5. Faculté de Médecine Paris-Sud, Univ Paris-Sud, Université Paris-Saclay, Le Kremlin-Bicêtre, France
6. Service d'hépato-gastroentérologie, Hôpital Antoine-Béclère, Hôpitaux Universitaires Paris-Sud, Assistance Publique-Hôpitaux de Paris, Clamart, France

7. CESP, Fac. de médecine - Univ. Paris-Sud, Fac. de médecine - UVSQ, INSERM, Université Paris-Saclay, 94805, Villejuif, France

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

9. Former: National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

10. Department of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany

11. Diet, Genes and Environment Unit, Danish Cancer Society Research Center, Copenhagen, Denmark

12. Hôpital de Bicêtre, department of Gastroenterology, Assistance Publique-Hôpitaux de Paris, F-94275 Le Kremlin-Bicêtre, France

13. Department of Epidemiology, German Institute of Human Nutrition, Potsdam-Rehbrücke, Germany

14. Hellenic Health Foundation, Athens, Greece

15. 2nd Pulmonary Medicine Department, School of Medicine, National and Kapodistrian University of Athens, “ATTIKON” University Hospital, Haidari, Greece

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

17. Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori Milano, Italy

18. Cancer Registry and Histopathology Department, "M.P. Arezzo" Hospital, ASP Ragusa, Ragusa, Italy

19. Unit of Cancer Epidemiology, Città della Salute e della Scienza University-Hospital and Center for Cancer Prevention (CPO), Turin, Italy

20. Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy

21. Former senior scientist, Department for Determinants of Chronic Diseases (DCD), National Institute for Public Health and the Environment (RIVM), PO Box 1, 3720 BA Bilthoven, The Netherlands

22. Former associate professor, Department of Gastroenterology and Hepatology, University Medical Centre, Utrecht, The Netherlands.

23. Former Visiting professor, Department of Epidemiology and Biostatistics, The School of Public Health, Imperial College London, St Mary's Campus, Norfolk Place, London, United Kingdom

24. Former Academic Icon / visiting professor, Department of Social & Preventive Medicine, Faculty of Medicine, University of Malaya, Pantai Valley, 50603, Kuala Lumpur, Malaysia

25. Department of Community Medicine, UIT – the Arctic University of Norway, Tromsø, Norway

26. Public Health Directorate, Asturias, Spain

27. Unit of Nutrition and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology (ICO), Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain

28. Escuela Andaluza de Salud Pública. Instituto de Investigación Biosanitaria ibs. GRANADA, Universidad de Granada. Granada, Spain

29. CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

30. Public Health Division of Gipuzkoa, BioDonostia Research Institute, San Sebastian, Spain

31. Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain

32. Navarra Public Health Institute, Pamplona, Spain

33. Navarra Institute for Health Research (IdiSNA), Pamplona, Spain

34. Lund University, Skåne University Hospital, Department of Gastroenterology and Nutrition, Malmö, Sweden

35. Lund University, Skåne University Hospital, Department of Internal Medicine, Malmö, Sweden

36. Department of Surgery, Department of Surgical and Perioperative Sciences, Umeå University, Sweden

37. Department of Public Health and Clinical Medicine, Umeå University, Sweden

38. Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK

39. Office of the Director, International Agency for Research on Cancer (IARC-WHO), Lyon, France

40. Department of Epidemiology and Biostatistics, Imperial College London, London, UK

41. MRC Centre for Transplantation, King's College London, Great Maze Pond, London, UK

*** Corresponding Author Contact information:** Dr. Mazda Jenab, OncoMetabolomics Team, Nutrition and Metabolism Branch, International Agency for Research on Cancer (IARC-WHO), 150 Cours Albert

Thomas, 69372 Lyon CEDEX 08, France, telephone number: +33(0)472738082, email: jenabm@iarc.fr ; twitter: @mjenab ; ORCID: 0000-0002-0573-1852

Keywords: hepatocellular carcinoma, cancer prevention, biomarkers, bile acid metabolism, obesity

Novelty and Impact

We studied BA metabolism in hepatocellular carcinoma (HCC) development using pre-diagnostically collected plasma samples from a multicentric prospective cohort. We observed perturbed BA metabolism in HCC, apparent from several years before diagnosis. Compared to matched controls, HCC cases showed increased total BAs with a shift towards taurine-conjugation, with adjustment for lifestyle/metabolic confounders. Future studies should explore the potential for modulation of BA metabolism in HCC development and BA profiling for clinical surveillance of high-risk patients.

List of abbreviations: AFP, alpha-fetoprotein; ALD, alcoholic liver disease; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BA, bile acid; BMI, body mass index; CA, cholic acid; CDCA, chenodeoxycholic acid; DCA, deoxycholic acid; EPIC, European Prospective Investigation into Cancer and Nutrition cohort; ESI, electrospray ionization; FLI, fatty liver index; GCA, glycocholic acid; GCDCA, glycochenodeoxycholic acid; GDCA, glycodeoxycholic acid; GGT, gamma-glutamyltransferase; GHCA, glycohyocholic acid; GUDCA, glycoursodeoxycholic acid; HbA1c, glycated haemoglobin; HCC, hepatocellular carcinoma; hsCRP, high-sensitivity C-reactive protein; LCA, lithocholic acid; LOD, limit of detection; LOQ, limit of quantification; MS, mass spectrometry; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; OR, odds ratio; PLSDA, partial least square discriminant analyses; ROC, receiver operation characteristics; TaMCA, tauro-alfa-muricholic acid; TDCA, taurodeoxycholic acid; TG, triglycerides; THCA, Taurohyocholic acid; TUDCA, tauroursodeoxycholic acid.

Abstract

Bile acids (BA) play different roles in cancer development. Some are carcinogenic and BA signaling is also involved in various metabolic, inflammatory, and immune-related processes. The liver is the primary site of BA synthesis. Liver dysfunction and microbiome compositional changes, such as during hepatocellular carcinoma (HCC) development, may modulate BA metabolism increasing concentration of carcinogenic BAs. Observations from prospective cohorts are sparse. We conducted a study (233 HCC case-control pairs) nested within a large observational prospective cohort with blood samples taken at recruitment when healthy with follow-up over time for later cancer development. A targeted metabolomics method was used to quantify 17 BAs (primary/secondary/tertiary; conjugated/un-conjugated) in pre-diagnostic plasma. Odd ratios (OR) for HCC risk associations were calculated by multivariable conditional logistic regression models. Positive HCC risk associations were observed for the molar sum of all BAs ($OR_{doubling}$=2.30, 95%CI=1.76-3.00) and choline- and taurine-conjugated BAs. Relative concentrations of BAs showed positive HCC risk associations for glycoholic acid and most taurine-conjugated BAs. We observe an association between increased HCC risk and higher levels of major circulating BAs, from several years prior to tumor diagnosis and after multivariable adjustment for confounders and liver functionality. Increased in BA concentration is accompanied by a shift in BA profile towards higher proportions of taurine-conjugated BAs, indicating early alterations of BA metabolism with HCC development. Future studies are needed to assess BA profiles for improved stratification of patients at high HCC risk and to determine whether supplementation with certain BAs may ameliorate liver dysfunction.

Introduction

Hepatocellular carcinoma (HCC) is the most common type of primary liver cancer, is a leading cause of cancer-related mortality worldwide and has limited therapeutic options. There is considerable understanding about the roles of chronic hepatitis B/C infection, heavy alcohol drinking, smoking and dietary aflatoxin exposures in HCC development in different populations. However, obesity, diabetes and non-alcoholic fatty liver disease (NAFLD) are also emerging as important HCC risk factors, particularly as rates of obesity and diabetes increase concomitantly with decreasing rates of chronic hepatitis infections [1-3].

One of the major biological functions of the liver is bile acid (BA) biosynthesis, metabolism and excretion, meaning that the organ is exposed to BAs from both *de novo* synthesis within the liver itself and from intestinal/gut re-absorption of primary and secondary BAs [4]. BAs are essential for intestinal lipid absorption and have numerous important metabolic, regulatory, and signalling functions [4, 5]. But, they can also promote cell proliferation, inflammation and oxidative stress, potentially leading to DNA damage and tumour growth [6, 7]. Primary BA are synthesized in the liver, conjugated with taurine or glycine, and stored in the gall bladder as bile which is excreted into the intestinal tract with food consumption [8]. Excreted BA are largely deconjugated, the majority are re-absorbed via the entero-hepatic circulation and some reach the colon where they are converted to secondary BA by gut microbial action before also being largely re-absorbed [8].

BA metabolism is affected by the functional capacity of the liver, as well as by various dietary and lifestyle exposures [9, 10] and the composition of the gut microbiota [11]. These factors can each alter not only the total levels of BAs but also the overall profile of the body BA pool, for example via alterations in the conjugation profile of BAs and changes in the rate of conversion of primary-to-secondary BAs brought about by modifications in gut microbiome composition [9, 12, 13]. BA metabolism may also be altered by various disease states. For example, NAFLD is often accompanied by elevation of BA levels and a change in circulating BA profiles [14] while non-alcoholic steatohepatitis (NASH) and alcoholic hepatitis lead to accumulation of more hepato-toxic BAs and enhanced transformation of primary to secondary BAs [15, 16].

De-regulation of BA metabolism is a likely early event in HCC development [17, 18]. Increasing perturbations of BA metabolism and accumulation of toxic BAs have been observed in the progression of liver cirrhosis and the development of cirrhosis-derived HCC [19]. We have previously observed strong positive HCC risk associations with circulating levels of two secondary glycine-conjugated bile acids, glycocholic acid (GCA) and glycochenodeoxycholic acid (GCDCA), in the European Prospective Investigation into Cancer (EPIC), an observational cohort study [20]. Similar observations also exist in separate cohorts of male Finnish smokers [21], Taiwanese hepatitis B or C positive individuals [22] and in the Singapore Chinese Health Study [23]. De-regulation of BA metabolism and higher total BA levels in HCC have also been observed in other prospective [22-25] and retrospective [26] cohorts as well as in studies of HCC patients [27, 28]. To date, most of the publications on this topic have been based on Asian populations and there is a paucity of information from prospective cohort studies from European populations. In this detailed analysis, we build on our

previous observations [20] by conducting comprehensive profiling and quantification of individual BAs within the EPIC study, a large multi-centre European observational prospective cohort, using pre-diagnostically collected plasma samples taken from healthy participants at enrolment who were then followed-up until disease diagnosis.

Material and Methods

Study design

The rationale, study population and data collection methods of the EPIC cohort have been previously described [29]. Briefly, between 1992-2000 over 520,000 apparently healthy men/women were recruited from 10 European countries: Denmark, France, Germany, Greece, Italy, Norway, Spain, Sweden, the Netherlands, and United Kingdom. Detailed dietary/lifestyle/anthropometry data and blood samples were collected at recruitment. Participants were followed-up for determination of post-recruitment cancer diagnoses. All cohort participants provided written informed consent. Cancer diagnoses were determined through record linkage with regional population cancer registries or by active follow up via a combination of methods up to 2012. Cases were defined as C22.0 (with morphology codes 8170/3, 8171/3, 8180/3), according to the 10th revision of the International Statistical Classification of Diseases, Injury and Causes of Death (ICD10) and the 2nd edition of the International Classification of Diseases for Oncology (ICD-O-2).

Nested case-control study design

After a mean of 8.6 years (maximum 19 years) post-recruitment, 233 cohort participants developed first-incident, primary, histologically confirmed HCC and had available baseline blood samples for laboratory measurements. Each HCC case was matched to one healthy control participant using incidence density sampling from all cohort participants alive and free of cancer (except non-melanoma skin cancer). Matching criteria were: age at blood collection (±1 year), sex, study center, time of the day at blood collection (±3 hours), fasting status at blood collection (<3, 3-6, and >6 hours; to account for potential differences in BA concentrations by fasting); women were additionally matched by menopausal status (pre-/peri-/post-menopausal), and hormone replacement therapy use at time of blood collection (yes/no). Additional design details are provided in **Supplementary Methods.**

Laboratory measurement of circulating bile acids

A series of 17 plasma BAs were quantitatively measured using a targeted metabolomic profiling method [30]. To retain all participants in the analyses, the value of the relevant limit was assigned to those whose measured BA concentrations were either below the limit of detection (LOD) or below/above the limit of quantification (LOQ). THCA was excluded from the main analyses because >40% of total subjects (48% cases; 76% controls) had values below the LOD or LOQ. BAs were expressed as concentrations (nanomoles, nM) and in terms of relative contribution (calculated as the percentage of each BA to the molar sum of the 17 BAs and expressed as % of total BAs), groupings of BA families by species, BAs of hepatic or bacterial origin (i.e. primary, secondary), BA hydrophilicity/hydrophobicity and conjugation status. Ratios of specific BAs were also computed.

Assessment of liver/metabolic dysfunction, definition of "suspected" NAFLD

Existing measures of biomarkers of metabolic/liver function and hepatitis B/C infection (**Supplementary Methods**) were used to assess correlations with circulating BAs. The fatty liver index (FLI) [31] and a metabolic syndrome score [32] were also calculated using existing biomarkers/data (see **Supplementary Methods** and Table 1 footnotes). Individuals with "suspected" NAFLD were defined as cases/controls with no prevalent viral hepatitis B/C infection, moderate alcohol intake of <30g/d (men) or <20 g/d (women) and characterized with at least one of the following: ALT>55 U/L, GGT>64 U/L in men and GGT>36 U/L in women, FLI>60, or presence of the metabolic syndrome.

Statistical Analyses

Conditional logistic regression was used to compute odds ratios (OR) and 95% confidence intervals (95%CI) to assess the association between BAs and HCC risk. BAs were log2-transformed and assessed as both continuous exposures and by tertiles with cut-points based on the distribution of controls. OR from continuous analyses indicate HCC risk corresponding to a doubling of circulating BA concentration or relative contribution of each BA to the molar sum of BAs. A crude model, conditioned on the matching criteria only, was applied first, followed by a detailed multivariable model with additional *a priori*-defined adjustments for BMI (kg/m^2), waist circumference (cm), recreational/household physical activity (Met-

hours/week), baseline alcohol consumption (g/day), lifetime alcohol intake pattern, smoking status and highest level of education attainment (see **Table 1**). We also tested additional dietary variables and fasting status as potential confounders, but they did not appear as such (<10% change in estimate) in preliminary testing and were thus not included in the final multivariable model. In a second series of multivariable models, additional adjustments for circulating GGT concentrations or FLI (to account for differences in BA concentrations potentially due to disparities in liver functional capacity between cases and their matched controls), as well as hepatitis B/C positivity were implemented. Additional dietary variables and fasting status did not appear as confounders when tested and were not included in the multivariable models. Sensitivity analyses were also run limiting the analyses to case-control pairs where the case was:

(a) diagnosed >2 years post-enrolment (to assess reverse causality; n case-control pairs=209), or

(b) free of hepatitis B/C infection at baseline (to assess the role of BAs without viral etiology; n case-control pairs=114).

Two sub-group analyses were also conducted with the aim to assess the BA-HCC association under different severities of liver dysfunction in HCC cases and under conditions of severe metabolic dysfunction, respectively:

(a) stratification by number of abnormal liver function parameters in HCC cases (i.e. cases with none or 1 (n case-control pairs=123) vs cases with 2-5 abnormal cumulative liver function parameters (n case-control pairs=110); see Table 1 and 3 footnotes), and

(b) restriction to case-control pairs where both participants were “suspected” NAFLD” patients (n case-control pairs=27, crude models; defined above).

All statistical tests were two-sided and p values <0.05 were considered statistically significant. The p-values presented in the tables are the original p-values. The threshold p-value for Bonferroni correction for multiple testing considering 17 BAs is 0.003 for linear models. For categorical models in tertiles, the threshold p-value for Bonferroni correction was calculated to be 0.0015 (i.e. 0.05 divided by 34 individual tests). Statistical analyses were conducted using SAS version 9.2 (SAS Institute, NC).

Results

Baseline characteristics

Table 1 shows baseline characteristics of HCC cases and their matched controls. The median absolute concentration of total BAs was higher in HCC cases than controls (5,600 nM vs 2,311 nM; **Table 1**). Compared to controls, HCC cases had generally higher levels of individual BAs (ranging from 1.2 (DCA) to 4.5 (TCA) times), higher level of conjugated BAs (87.1% vs 61.5% in controls) and differing levels of glycine- (71% vs 89%) and taurine-conjugation (29% vs 11%). Correlations between circulating BAs and by liver function parameters are shown in **Supplementary Figure 1**.

Concentration of BAs and HCC risk

Associations between circulating BA concentrations and HCC risk are shown in **Table 2**. After Bonferroni p-value corrections for multiple comparisons, total circulating BAs (nM) were positively associated with HCC risk (multivariable adjusted OR per doubling of concentration=2.30; 95%CI:1.76-3.00). Positive HCC risk associations were observed for individual taurine- and glycine-conjugated BAs (**Table 2**). In both crude and multivariable adjusted continuous models, the findings were not altered by further adjustment for circulating GGT levels (**Table 2**) or FLI (**Supplementary Table 2a**). These additional adjustments were applied to control for the severity of liver dysfunction and potential disparity in liver functionality between cases and matched controls.

Associations for plasma concentrations of groupings of BAs are shown in **Supplementary Table 3a**. In multivariable models, higher concentration of groupings of primary (OR per doubling of concentration=2.20, 95%CI:1.72 - 2.82), secondary (OR per doubling of concentration=1.71, 95%CI:1.39 - 2.10) and total conjugated BAs (OR per doubling of concentration=2.31, 95%CI:1.77 - 3.00) were associated with increased HCC risk, whereas no risk associations were observed for groupings of unconjugated BAs. Further adjustments for FLI, hepatitis B/C status and sensitivity analyses restricted to cases diagnosed after 2 years of follow-up did not alter observed associations.

Loess curves showed a clear difference in the concentrations of total and conjugated BAs between HCC cases and controls up to 10 years prior to HCC diagnosis (**Supplementary Figure 2**).

Relative proportions of specific BAs and HCC risk

Supplementary Table 1 shows HCC risk estimates for relative proportions of BA (% concentration of each BA relative to total sum of all BAs). Glycine- and taurine-conjugated forms of CA, i.e. GCA (OR=2.13,

95%CI: 1.58-2.86) and TCA (OR=1.83, 95%CI: 1.50-2.22), showed positive HCC risk associations in continuous multivariable adjusted models (per doubling of relative proportion). Similar associations were observed for other taurine-conjugated BA, i.e. TaMCA (OR=1.47, 95%CI: 1.25-1.74), TCDCA (OR=1.74, 95%CI: 1.40-2.15), and TUDCA (OR=1.39, 95%CI: 1.13-1.72)(per doubling of relative proportion). However, inverse HCC risk associations were observed for unconjugated BAs and two glycine-conjugated BAs (GDCA and GLCA); while no association was observed for the remaining BAs. Most of these associations were maintained with further adjustment for FLI and in sensitivity analyses (**Supplementary Table 2b**). Analyses by groupings of BAs showed positive HCC risk associations conjugated BAs while unconjugated BAs were inversely associated (**Supplementary Table 3b**).

Sub-group analyses

Table 3 shows HCC risk associations for plasma BA concentrations stratified by case-control sets where the case has ≤1 abnormal liver function parameter compared to a group where the case has 2-5 abnormal parameters. The former group showed modest but significant positive HCC risk associations for all BAs (OR per doubling of concentration of total sum of BAs=1.55, 95%CI:1.22-1.97), while the group with higher severity of liver dysfunction demonstrated stronger HCC risk associations (OR=5.61, 95%CI:2.78-11.33). We also performed a second sub-analysis restricting to a sub-set of 27 case-control pairs with “suspected” NAFLD. Baseline characteristics for this sub-group are shown in **Supplementary Table 4** and the HCC risk associations in **Supplementary Table 5**. Overall findings in this sub-set of subjects were not remarkably different from those observed for the whole series of HCC case-control pairs.

Discussion

This study was based on the EPIC observational prospective cohort composed of apparently healthy participants from whom blood samples were collected at baseline prior to diagnosis, with subsequent follow-up over time for cancer diagnoses, including HCC. We observed that participants who had higher baseline circulating concentration of total BAs had a greater HCC risk even after accounting for possible confounding by established HCC risk factors and additional adjustment for GGT levels or FLI as markers of potential liver dysfunction. Our data also showed changes in relative proportions of BAs in HCC development, with a BA profile composed of higher proportion of GCA and several taurine-conjugates, at

the expense of primary BAs, being more closely associated with HCC risk. We also observed clear increases in the concentration of conjugated BAs in cases compared to matched controls up to 10 years prior to HCC diagnosis. Collectively, our findings indicate that alteration of BA metabolism is an early event in HCC development – a key strength of our prospective design.

Our observations of increased circulating BA levels and changes in BA profiles in HCC development were not confounded by hepatitis infection status and are apparent even after adjustment for degree of liver dysfunction. Liver dysfunction may affect BA metabolism in several ways, such as impaired hepatic BA clearance, increased BA synthesis, leakage of BAs from injured hepatocytes, and alterations in the composition of the gut microbiome leading to changes in the production of secondary/tertiary BAs. We accounted for the influence of liver dysfunction on our findings in two separate ways. First, we modelled an additional adjustment for circulating GGT levels and secondly, we added FLI into our multivariable model. In both situations, our observation of a positive association between higher BA levels and HCC development were largely unchanged suggesting that our observations are only partially explained by severity of liver dysfunctionality.

We further explored the role of liver dysfunction by stratification into two different sub-groups, one with HCC case-control pairs where the case had no or low liver impairment and the second with those where the case had moderate to severe liver impairment. We show similar patterns of HCC risk association with BA profiles in the two sub-groups, although the magnitude of associations was lower in the sub-group with low liver impairment. This observation highlights the connection between BA metabolism and liver functionality. More importantly, it demonstrates that even in those with good liver functionality, alterations in BA metabolism may be indicative of HCC development. Coupled to our sensitivity analyses showing that exclusion of cases diagnosed within 2 years of baseline did not alter our observations, these results indicate alteration of BA metabolism as an early event in HCC development. This raises the possibility of additional studies to explore dysregulated BA metabolism as an additional tool for more refined stratification of patients who may be at higher risk for HCC development.

Unhealthy lifestyle exposures may, on the one hand, bring about a cancer promotive environment in the liver linked to metabolic dysfunction and inflammation, and on the other hand, also affect unfavourable changes in the gut microbiome, further impacting BA profile changes, possibly towards ones that contain more harmful/carcinogenic species, hence inducing a vicious cycle of further liver impairment, induction of malignant change and promotion of liver tumour growth. However, in our statistical modelling adjustment for smoking, alcohol consumption and physical activity did not meaningfully change our findings.

The comprehensive targeted, quantitative BA profiling method we applied allowed the differentiation of individual glycine- and taurine-conjugates, showing that HCC cases had higher relative proportions of taurine-conjugated BAs and GCA compared to matched controls, similar to observations from the Singapore Chinese Health Study cohort [23], which applied a comparable BA profiling method to that used in our study. Another recent study in different ethnic groups showed elevated TCA, TDCA and GCA proportions and a lower proportion of DCA in patients with broad hepatic impairment compared to healthy subjects [33], in line with our findings. Elevated glycine- and taurine-conjugated BAs have also been identified as main discriminants for HCC development in Chinese hepatitis patients [25], and higher TUDCA has been associated with HCC risk in a Korean cohort study [24]. Another recent study based on two cohorts of chronic hepatitis B and C patients of Chinese ancestry from Taiwan – using the Metabolon metabolomics platform - has also shown a positive HCC risk association with increased circulating levels of primary and taurine- or glycine-conjugated primary BA measured in pre-diagnostically collected blood [22]. Interestingly, we also observed a strong linear positive HCC risk association with levels of TaMCA, a murine taurine-conjugated BA not usually observed in humans [34]. The presence of TaMCA may be indicative of gut microbiome dysbiosis, i.e. alteration of the microbiome as a feature of poor dietary/lifestyle habits and processes of disease development, which affects BA metabolism, synthesis and composition [35]. In fact, we can speculate, based on our observations, that taurine-conjugation of BAs increases with impaired liver function in HCC development, possibly as an adaptive mechanism aimed to protect the liver from unfavourable metabolic effects of chronic exposure to more toxic BA. A potential shift towards taurine-conjugation of BAs in HCC development has also been observed in the recent analysis of the Singapore Chinese Health Study [23]. The authors speculate an increased production of these BA with higher fat consumption, and a role for them in promotion of liver cirrhosis and gut barrier dysfunction [23]. But other

studies suggest potentially protective roles for taurine-conjugated BA in the presence of cardiometabolic risk factors [36] – something that deserves further investigation in relation to liver diseases. BA deconjugation and production of secondary BAs is dependent on the gut bacterial microbiome [8]. Interestingly, Petrick et al observed an inverse HCC risk association between secondary BA and HCC risk in chronic hepatitis patients [22], also possibly implicating alterations in the bacterial microbiome composition, something that has previously been observed during hepatitis infections [37, 38]. In HCC, alteration of the composition of the gut microbiome towards more pathogenic bacteria has been observed, with a reduction of microbial diversity and an increase in bacterial genera that produce lipopolysaccharide (LPS) [8]. These observations are in line with our own previous findings of a strong, positive HCC risk association with increased LPS exposures [39], suggesting an alteration of the gut microbiome composition that may then modulate the body BA pool, alter circulating BA profiles and affect various complex microbiome-BA signalling pathways possibly creating a pro-inflammatory hepatic environment [8]. Following from this, it has been recently suggested that modulation of BA towards more favourable profiles through manipulation of gut microbiome composition may be a treatment strategy for liver cancer patients [40]. In addition, we also observe that higher relative concentration of hydrophilic BA (UDCA and CA) are associated with lower HCC risk. These BA are known to be liver protective, and it may thus be speculated that their supplementation may be effective in ameliorating liver dysfunction.

NAFLD is emerging as an important HCC risk factor [41] and has been associated with elevation of circulating BAs [14], in line with our present observations. Similar observations were made in the Singapore Chinese Health Study cohort [23]. We assessed the BA-HCC risk association in a restricted sub-set of subjects where we speculate that both the case and matched control pair had “suspected” NAFLD. We acknowledge that the lack of a clinical diagnosis of NAFLD in our cases is a limitation, but we surmise that subjects with a specific series of exposures – very low alcohol intake, high GGT and FLI, and metabolic syndrome – were likely to have NAFLD. Interestingly, this smaller sub-group showed the same pattern of observations seen in the full case series, i.e. association of HCC risk associated with elevated total and conjugated BAs, indicating that perturbations of BA metabolism are a factor in HCC development irrespective of the main etiology of the HCC. Our findings did not demonstrate differing BA profiles by hepatitis infection status or in the suspected NAFLD sub-group. However, a recent patient study has shown

distinct BA profiles in patients with different chronic liver diseases [42]. The challenge for future studies will be to determine whether BA profiling may be utilized as a tool for early diagnosis and differentiation of various chronic liver diseases, including HCC.

Our study has several key strengths, but foremost is its prospective design, nested within a large, multi-national observational cohort. We also collected detailed pre-diagnostic baseline and confounder information and biological samples on which our biomarker analyses were based. A key limitation pertains to the unavailability of clinical data in our HCC cases, particularly on fibrosis or cirrhosis. We addressed this by carefully adjusting for indices of liver functionality and steatosis. We also included fasting status at blood collection as a matching criterion to account for the effect of this variable on BA metabolism.

In conclusion, in this study based on cases and controls from a well characterized prospective observational cohort, we show that increased circulating BA levels, particularly GCA and taurine-conjugated BAs, were strongly associated with HCC development. It remains to be determined whether BA profiling can serve in better risk stratification of subjects who are at higher risk of HCC development and whether manipulation of BA profiles towards less toxic species may improve liver impairment in these patients.

Conflict of Interest: The authors disclose no conflicts.

Disclaimer: Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

Ethics Statement: The EPIC cohort was initially approved by the IARC Ethics Committee in January 1995 and again in May, 2017. The present study was approved by the IARC Ethics Committee (Project No. 16-06; February 2016; PI: M. Jenab) and by the relevant ethical review boards of the participating institutions. All EPIC participants provided written informed consent at enrollment.

Data Availability Statement: 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. Further information is available from the corresponding author upon request.

Funding: This work was supported in part by the French National Cancer Institute (L'Institut National du Cancer; INCa; grant numbers 2009-139 and 2014-1-RT-02-CIRC-1; PI: M. Jenab) and by internal funds of the IARC.

The coordination of EPIC is financially supported by the International Agency for Research on Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London which has additional infrastructure support provided by the NIHR Imperial Biomedical Research Centre (BRC).

The national cohorts are supported by: Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), Federal Ministry of Education and Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy, Compagnia di SanPaolo 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); Health Research Fund (FIS) - Instituto de Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, and the Catalan Institute of Oncology - ICO (Spain); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford). (United Kingdom). The funding sources had no influence on the design of the study; the collection, analysis, and interpretation of data; the writing of the report; or the decision to submit the paper for publication.

Acknowledgments: We would like to acknowledge Dr Krasimira Aleksandrova for input on the present manuscript, along with the National Institute for Public Health and the Environment (Bilthoven, the Netherlands), the Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands, Department of Public Health Aarhus University (Aarhus, Denmark), University of Cambridge (Cambridge, United Kingdom), for their contributions and ongoing support to the EPIC Study.

Author contributions: The authors' responsibilities were as follows. MS and MJ: conceptualized, designed, obtained funding for and carried out the present research; ML, AL, EJ: conducted laboratory analyses; MS, MJ: performed the statistical analysis; MS and MJ: contributed jointly to data interpretation and writing of the manuscript. VV: provided advice on statistical analyses; TK, GP, CV, CD, MBR, MJG: provided input and critical comment on data interpretation and manuscript writing. Contributing authors from each individual collaborating center provided the original data and biological samples, information on the respective populations, advice on study design/analysis, and interpretation of the results. All authors provided an approval of the final version of the manuscript for publication.

References

1. Petrick JL, McGlynn KA. The changing epidemiology of primary liver cancer. Current epidemiology reports 2019;6:104-11.
2. Sanyal AJ, Yoon SK, Lencioni R. The etiology of hepatocellular carcinoma and consequences for treatment. The oncologist 2010;15 Suppl 4:14-22.

3. Schlesinger S, Aleksandrova K, Pischon T, Fedirko V, Jenab M, Trepo E, Boffetta P, Dahm CC, Overvad K, Tjønneland A, Halkjær J, Fagherazzi G, et al. Abdominal obesity, weight gain during adulthood and risk of liver and biliary tract cancer in a European cohort. International journal of cancer 2013;132:645-57.

4. Chiang JYL, Ferrell JM. Bile Acid Metabolism in Liver Pathobiology. Gene expression 2018;18:71-87.

5. Li T, Chiang JY. Bile acids as metabolic regulators. Current opinion in gastroenterology 2015;31:159-65.

6. Bernstein H, Bernstein C, Payne CM, Dvorakova K, Garewal H. Bile acids as carcinogens in human gastrointestinal cancers. Mutation research 2005;589:47-65.

7. Phelan JP, Reen FJ, Caparros-Martin JA, O'Connor R, O'Gara F. Rethinking the bile acid/gut microbiome axis in cancer. Oncotarget 2017;8:115736-47.

8. Wu L, Feng J, Li J, Yu Q, Ji J, Wu J, Dai W, Guo C. The gut microbiome-bile acid axis in hepatocarcinogenesis. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie 2021;133:111036.

9. Ramírez-Pérez O, Cruz-Ramón V, Chinchilla-López P, Méndez-Sánchez N. The Role of the Gut Microbiota in Bile Acid Metabolism. Annals of hepatology 2017;16:s15-s20.

10. Trefflich I, Marschall HU, Giuseppe RD, Ståhlman M, Michalsen A, Lampen A, Abraham K, Weikert C. Associations between Dietary Patterns and Bile Acids-Results from a Cross-Sectional Study in Vegans and Omnivores. Nutrients 2019;12.

11. Chávez-Talavera O, Tailleux A, Lefebvre P, Staels B. Bile Acid Control of Metabolism and Inflammation in Obesity, Type 2 Diabetes, Dyslipidemia, and Nonalcoholic Fatty Liver Disease. Gastroenterology 2017;152:1679-94.e3.

12. Beyoglu D, Idle JR. The metabolomic window into hepatobiliary disease. Journal of hepatology 2013;59:842-58.

13. Jia W, Xie G, Jia W. Bile acid-microbiota crosstalk in gastrointestinal inflammation and carcinogenesis. Nature reviews. Gastroenterology & hepatology 2018;15:111-28.

14. Gottlieb A, Canbay A. Why Bile Acids Are So Important in Non-Alcoholic Fatty Liver Disease (NAFLD) Progression. Cells 2019;8.

15. Ciocan D, Voican CS, Wrzosek L, Hugot C, Rainteau D, Humbert L, Cassard AM, Perlemuter G. Bile acid homeostasis and intestinal dysbiosis in alcoholic hepatitis. Alimentary pharmacology & therapeutics 2018;48:961-74.

16. Puri P, Daita K, Joyce A, Mirshahi F, Santhekadur PK, Cazanave S, Luketic VA, Siddiqui MS, Boyett S, Min HK, Kumar DP, Kohli R, et al. The presence and severity of nonalcoholic steatohepatitis is associated with specific changes in circulating bile acids. Hepatology (Baltimore, Md.) 2018;67:534-48.

17. Di Ciaula A, Wang DQ, Molina-Molina E, Lunardi Baccetto R, Calamita G, Palmieri VO, Portincasa P. Bile Acids and Cancer: Direct and Environmental-Dependent Effects. Annals of hepatology 2017;16:s87-s105.

18. Kettner NM, Voicu H, Finegold MJ, Coarfa C, Sreekumar A, Putluri N, Katchy CA, Lee C, Moore DD, Fu L. Circadian Homeostasis of Liver Metabolism Suppresses Hepatocarcinogenesis. Cancer cell 2016;30:909-24.

19. Liu N, Feng J, Lv Y, Liu Q, Deng J, Xia Y, Guo C, Zhou Y. Role of bile acids in the diagnosis and progression of liver cirrhosis: A prospective observational study. Experimental and therapeutic medicine 2019;18:4058-66.

20. Stepien M, Keski-Rahkonen P, Kiss A, Robinot N, Duarte-Salles T, Murphy N, Perlemuter G, Viallon V, Tjønneland A, Rostgaard-Hansen AL, Dahm CC, Overvad K, et al. Metabolic perturbations prior to hepatocellular carcinoma diagnosis: Findings from a prospective observational cohort study. International journal of cancer 2020.

21. Loftfield E, Rothwell JA, Sinha R, Keski-Rahkonen P, Robinot N, Albanes D, Weinstein SJ, Derkach A, Sampson J, Scalbert A, Freedman ND. Prospective Investigation of Serum Metabolites, Coffee Drinking, Liver Cancer Incidence, and Liver Disease Mortality. Journal of the National Cancer Institute 2020;112:286-94.

22. Petrick JL, Florio AA, Koshiol J, Pfeiffer RM, Yang B, Yu K, Chen CJ, Yang HI, Lee MH, McGlynn KA. Prediagnostic concentrations of circulating bile acids and hepatocellular carcinoma risk: REVEAL-HBV and HCV studies. International journal of cancer 2020;147:2743-53.

23. Thomas CE, Luu HN, Wang R, Xie G, Adams-Haduch J, Jin A, Koh WP, Jia W, Behari J, Yuan JM. Association between Pre-Diagnostic Serum Bile Acids and Hepatocellular Carcinoma: The Singapore Chinese Health Study. Cancers 2021;13.

24. Jee SH, Kim M, Kim M, Yoo HJ, Kim H, Jung KJ, Hong S, Lee JH. Metabolomics Profiles of Hepatocellular Carcinoma in a Korean Prospective Cohort: The Korean Cancer Prevention Study-II. Cancer prevention research (Philadelphia, Pa.) 2018;11:303-12.

25. Zhang W, Zhou L, Yin P, Wang J, Lu X, Wang X, Chen J, Lin X, Xu G. A weighted relative difference accumulation algorithm for dynamic metabolomics data: long-term elevated bile acids are risk factors for hepatocellular carcinoma. Scientific reports 2015;5:8984.

26. Wang H, Shang X, Wan X, Xiang X, Mao Q, Deng G, Wu Y. Increased hepatocellular carcinoma risk in chronic hepatitis B patients with persistently elevated serum total bile acid: a retrospective cohort study. Scientific reports 2016;6:38180.

27. Chen T, Xie G, Wang X, Fan J, Qiu Y, Zheng X, Qi X, Cao Y, Su M, Wang X, Xu LX, Yen Y, et al. Serum and urine metabolite profiling reveals potential biomarkers of human hepatocellular carcinoma. Molecular & cellular proteomics : MCP 2011;10:M110.004945.

28. Tan Y, Yin P, Tang L, Xing W, Huang Q, Cao D, Zhao X, Wang W, Lu X, Xu Z, Wang H, Xu G. Metabolomics study of stepwise hepatocarcinogenesis from the model rats to patients: potential biomarkers effective for small hepatocellular carcinoma diagnosis. Molecular & cellular proteomics : MCP 2012;11:M111.010694.

29. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. International journal of epidemiology 1997;26 Suppl 1:S6-14.

30. García-Cañaveras JC, Donato MT, Castell JV, Lahoz A. Targeted profiling of circulating and hepatic bile acids in human, mouse, and rat using a UPLC-MRM-MS-validated method. Journal of lipid research 2012;53:2231-41.

31. Bedogni G, Bellentani S, Miglioli L, Masutti F, Passalacqua M, Castiglione A, Tiribelli C. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. BMC gastroenterology 2006;6:33.

32. Alberti KG, Zimmet P, Shaw J. The metabolic syndrome--a new worldwide definition. Lancet (London, England) 2005;366:1059-62.

33. Luo L, Aubrecht J, Li D, Warner RL, Johnson KJ, Kenny J, Colangelo JL. Assessment of serum bile acid profiles as biomarkers of liver injury and liver disease in humans. PloS one 2018;13:e0193824.

34. Wahlström A, Sayin SI, Marschall HU, Bäckhed F. Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. Cell metabolism 2016;24:41-50.

35. Ikegami T, Honda A. Reciprocal interactions between bile acids and gut microbiota in human liver diseases. Hepatology research : the official journal of the Japan Society of Hepatology 2018;48:15-27.

36. Guizoni DM, Vettorazzi JF, Carneiro EM, Davel AP. Modulation of endothelium-derived nitric oxide production and activity by taurine and taurine-conjugated bile acids. Nitric oxide : biology and chemistry 2020;94:48-53.

37. Wang J, Wang Y, Zhang X, Liu J, Zhang Q, Zhao Y, Peng J, Feng Q, Dai J, Sun S, Zhao Y, Zhao L, et al. Gut Microbial Dysbiosis Is Associated with Altered Hepatic Functions and Serum Metabolites in Chronic Hepatitis B Patients. Frontiers in microbiology 2017;8:2222.

38. Inoue T, Nakayama J, Moriya K, Kawaratani H, Momoda R, Ito K, Iio E, Nojiri S, Fujiwara K, Yoneda M, Yoshiji H, Tanaka Y. Gut Dysbiosis Associated With Hepatitis C Virus Infection. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 2018;67:869-77.

39. Fedirko V, Tran HQ, Gewirtz AT, Stepien M, Trichopoulou A, Aleksandrova K, Olsen A, Tjønneland A, Overvad K, Carbonnel F, Boutron-Ruault MC, Severi G, et al. Exposure to bacterial products lipopolysaccharide and flagellin and hepatocellular carcinoma: a nested case-control study. BMC medicine 2017;15:72.

40. Ma C, Han M, Heinrich B, Fu Q, Zhang Q, Sandhu M, Agdashian D, Terabe M, Berzofsky JA, Fako V, Ritz T, Longerich T, et al. Gut microbiome-mediated bile acid metabolism regulates liver cancer via NKT cells. Science (New York, N.Y.) 2018;360.

41. Marengo A, Rosso C, Bugianesi E. Liver Cancer: Connections with Obesity, Fatty Liver, and Cirrhosis. Annual review of medicine 2016;67:103-17.

42. Sang C, Wang X, Zhou K, Sun T, Bian H, Gao X, Wang Y, Zhang H, Jia W, Liu P, Xie G, Chen T. Bile Acid Profiles Are Distinct among Patients with Different Etiologies of Chronic Liver Disease. Journal of proteome research 2021;20:2340-51.

Table 1. Baseline demographic, anthropometric, lifestyle characteristics and biomarker measures of hepatocellular carcinoma cases (HCC; n= 233) and their matched controls (n=233) in the EPIC nested case-control study.

	Matched Controls	HCC Cases	*p value*
Age at blood collection [1]	59.1 (6.9)	59.1 (6.9)	
Sex [1]			
Men	153 (65.67)	153.0 (65.7)	
Women	80 (34.33)	80.0 (34.3)	
Body Mass Index (BMI, kg/m^2)	26.7 (3.8)	28.4 (4.9)	*<0.0001*
Fasting status [1]			
Not fasted (<3 h)	99 (42.5)	104 (44.6)	
In between (3-6 h)	53 (22.8)	50 (21.5)	
Fasted (>6 h)	77 (30.1)	74 (31.8)	
Waist circumference (cm)	91.3 (11.5)	96.7 (13.6)	*<0.0001*
Physical activity (Mets)	84.5 (52.5)	85.6 (54.8)	*0.81*
Baseline dietary intakes(g/d)			
Fat	85.9 (29.7)	84.8 (36.6)	*0.67*
Fibre	23.5 (8.1)	22.1 (9.4)	*0.05*
Sugar	100.6 (44.8)	107.4 (50.1)	*0.10*
Energy (kcal/d)	2208.4 (639.4)	2264.5 (940.6)	*0.37*
Alcohol (g/d)	16.0 (19.4)	23.9 (33.5)	*0.02*
Alcohol drinking pattern			
Never drinker	15 (6.44)	16 (6.87)	*0.02*
Former drinker	12 (5.15)	31 (13.3)	
Drinker at recruitment	27 (11.59)	29 (12.45)	
Lifetime drinker	179 (76.82)	157 (67.38)	
Highest education level			
Primary school completed	111 (47.64)	123 (52.79)	*0.23*
Technical/professional school	51 (21.89)	55 (23.61)	
Secondary school	31 (13.3)	16 (6.87)	
Longer education (incl. University)	35 (15.02)	34 (14.59)	
Not specified	5 (2.15)	5 (2.15)	
Smoking status			
Never	103 (44.21)	66 (28.33)	*<0.0001*
Former	78 (33.48)	71 (30.47)	
Smoker	51 (21.89)	94 (40.34)	
Unknown	1 (0.43)	2 (0.86)	
Bile Acid levels (nM)			
Cholic acid (CA)	75.4 (10 , 1557.4)	116.8 (13.2 , 1642.3)	*0.04*
Chenodeoxycholic acid (CDCA)	161.1 (21.3 , 2,083.8)	262.2 (17.7 , 2566.0)	*0.30*
Hyocholic acid (HCA)	7.3 (1.2 , 52.0)	10.6 (1.2 , 67.0)	*<0.0001*
Deoxycholic acid (DCA)	256.6 (14.7 , 1,171.9)	308.3 (16.6 , 1450.1)	*0.64*
Ursodeoxycholic acid (UDCA)	23 (2.4 , 199.5)	35.5 (2.4 , 478.8)	*0.003*
Glycocholic acid (GCA)	173.6 (29.9 , 855.7)	496.7 (69.5 , 9706)	*<0.0001*
Glycochenodeoxycholic acid (GCDCA)	804.0 (162.8 , 2,887.2)	1687.5 (211.9 , 13,566.0)	*<0.0001*
Glycodeoxycholic acid (GDCA)	194.7 (18.5 , 803.7)	315.7 (37.4 , 2,890.1)	*<0.0001*
Glycohyocholic acid (GHCA)	9.1 (5 , 34.8)	14.9 (5.0 , 205.6)	*<0.0001*
Glycolithocholic acid (GLCA)	17.4 (3 , 105.6)	27.8 (3.3 , 192.9)	*<0.0001*
Glycoursodeoxycholic acid (GUDCA)	74.3 (14.8 , 397.7)	147.3 (20.6 , 1,585.2)	*<0.0001*

Tauro-alfa-muricholic acid (TaMCA)	18.7 (3.2 , 168.1)	84.6 (8.1 , 6,630.0)	*<0.0001*
Taurocholic acid (TCA)	65.7 (13.0 , 439.9)	246.7 (20.4 , 8,536.0)	*<0.0001*
Taurochenodeoxycholic acid (TCDCA)	24.4 (2.5 , 175.3)	63.7 (4.3 , 900.7)	*<0.0001*
Taurodeoxycholic acid (TDCA)	1.2 (1.2 , 7.2)	2.8 (1.2 , 196.8)	*<0.0001*
Tauroursodeoxycholic acid (TUDCA)	3.9 (1.2 , 19.7)	14 (1.2 , 404.7)	*<0.0001*
Total BA sum	2,311 (662 , 8847)	5,600 (1,059, 78,310)	*<0.0001*
Biomarkers of Liver functionality			
Gamma-glutamyl transferase (GGT; IU/L)	25.5 (33.4)	125.7 (207.4)	*<0.0001*
For men only	29.5 (36.8)	166.1 (242.5)	*<0.0001*
For women only	18.0 (24.0)	49.3 (64.2)	*<0.0001*
Alanine aminotransferase (ALT; IU/L)	20.9 (15.3)	41.0 (36.7)	*<0.0001*
Alkaline Phosphatase (ALP; IU/L)	58.5 (20.7)	81.1 (61.9)	*<0.0001*
Total Bilirubin (µmol/L)	8.5 (4.3)	10.7 (8.6)	*<0.0001*
Aspartate aminotransferase (AST; IU/L)	23.4 (11.2)	47.1 (36.8)	*<0.0001*
Aspartate aminotransferase to Alanine aminotransferase ratio (AST/ALT)	1.5 (1.2)	1.6 (1.9)	*0.34*
Fatty liver index (FLI) [2]	38.5 (28.2)	59.7 (31.2)	*<0.0001*
BARD fibrosis score	1.65 (1.08)	1.96 (1.26)	*<0.0001*
Viral status			
Hepatitis B positive	13 (5.6)	37 (15.9)	*<0.0001*
Hepatitis C positive	5 (2.2)	54 (23.2)	*<0.0001*
Hepatitis B and/or C positive	17 (7.3)	82 (35.2)	*<0.0001*
Metabolic syndrome			
No	136 (58.4)	97 (41.6)	*0.0003*
Yes	87 (37.3)	126 (54.1)	
Diabetes status			
No	199 (85.4)	182 (78.1)	*0.05*
Yes	12 (5.2)	29 (12.4)	
Do not know	5 (2.1)	6 (2.6)	
Hypertension			
No	142 (61)	127 (54.5)	*0.08*
Yes	58 (24.9)	80 (34.3)	
Do not know	12 (5.2)	8 (3.4)	
Cardiovascular problem			
No	144 (61.8)	120 (51.5)	*0.01*
Yes	62 (26.6)	86 (36.9)	

[1] a case-control matching factor.

[2] please see supplementary methods for information on the calculation algorithm for the FLI. Continuous variables are presented as means (standard deviation) or median (5, 95%; bile acids). Categorical variables are presented as n (%). Paired sample t-test or non-parametric Wilcoxon's signed rank test was used to test the difference between cases and controls for continuous variables and chi-square test for categorical variables.

Metabolic syndrome (MetS) defined as any 3 of the following: abdominal obesity (waist circumference >=94 cm in men or >=80 cm or women, elevated triglycerides (>1.7 mmol/L), reduced HDL cholesterol (<1.03 mmol/L in men or <1.29 mmol/L in women), elevated BP (systolic >130, diastolic >=85, or previously diagnosed hypertension), abnormal glucose metabolism (HbA1c >=5.7% or self-reported diabetes at baseline). Clinical thresholds for liver function biomarkers: ALT>55 IU/L (n controls=12; n HCC cases=51), AST>34 IU/L (n controls=26; n HCC cases=126), GGT >64 IU/L for men (n controls=14; n HCC cases=88) and > 36 U/L for women (n controls=9; n HCC cases=27), ALP > 150 U/L (n controls=0; n HCC cases=13), total bilirubin > 20.5 µmol/L (n controls=5; n HCC cases=18); values were provided by the laboratory. Number of control participants with abnormal liver function parameters: no abnormal parameters=187, one abnormal parameter=32, two abnormal parameters=11, three abnormal parameters=3, four abnormal parameters=0, five abnormal parameters=0. Number of HCC cases with abnormal liver function parameters: no abnormal parameters=80,

one abnormal parameter=43, two abnormal parameters=61, three abnormal parameters=40, four abnormal parameters=7, five abnormal parameters=2. The number of matched case-control sets where: (a) both have a similar number of abnormal liver function parameters=69, (b) the HCC case has lesser number of abnormal liver function parameters than its matched control=23, (c) the HCC case has a greater number of abnormal liver function parameters than its matched control=141. The number of matched case-control sets where the HCC case has no or one abnormal liver function parameter=123. The number of matched case-control sets where the HCC case has between two to five abnormal liver function parameter=110. Missing values present for controls (fasting status n=4, MetS n=3, self-reported: diabetes status n=17, hypertension n=21, cardiovascular problem n=27, physical activity n=12) and HCC cases (fasting status n= 5, MetS n=1, self-reported: diabetes status n=16, hypertension n=18, cardiovascular problem n=27).

Table 2. Odds ratios (95 % confidence intervals) of HCC risk across tertiles of individual bile acid levels.

		Tertile 1	Tertile 2	Tertile 3	p trend	Continuous [3] (per doubling of concentration)	p-value	Continuous [4] (plus additional adjustment for GGT)	p-value
Unconjugated primary bile acids									
Cholic acid (CA)	Tertile range (nM)	4.79 -38.1	>38.1 -146.6	>=146.6					
	Crude model [1]	Ref.	1.30 (0.80, 2.13)	**1.68 (1.05, 2.67)**	0.050	**1.09 (1.01, 1.19)**	**0.040**	1.08 (0.97, 1.20)	0.165
	Multivariable adjusted model [2]	Ref.	1.24 (0.68, 2.27)	**1.86 (1.03, 3.36)**	0.040	**1.11 (1.00, 1.24)**	**0.040**	**1.15 (1.01, 1.31)**	**0.032**
Chenodeoxycholic acid (CDCA)	Tertile range (nM)	1.19 -85.9	>85.9 - 334.4	>=334.4					
	Crude model [1]	Ref.	1.03 (0.65, 1.64)	1.37 (0.88, 2.14)	0.130	**1.11 (1.01, 1.21)**	**0.030**	1.10 (0.98, 1.24)	0.100
	Multivariable adjusted model [2]	Ref.	0.78 (0.44, 1.39)	1.03 (0.59, 1.81)	0.550	1.07 (0.95, 1.19)	0.270	1.14 (0.99, 1.32)	0.067
Hyocholic acid (HCA)	Tertile range (nM)	1.19 - 4.1	>4.1 - 12.1	>=12.1					
	Crude model [1]	Ref.	1.19 (0.73, 1.94)	**1.75 (1.10, 2.76)**	0.010	**1.23 (1.09, 1.38)**	***<0.001***	**1.24 (1.06, 1.44)**	**0.007**
	Multivariable adjusted model [2]	Ref.	1.12 (0.59, 2.13)	**2.00 (1.11, 3.62)**	0.010	**1.29 (1.11, 1.50)**	***<0.001***	**1.32 (1.09, 1.59)**	**0.004**
Unconjugated secondary/tertiary bile acids									
Deoxycholic acid (DCA)	Tertile range (nM)	2.39 - 173.5	>173.5 - 384.2	>=384.2					
	Crude model [1]	Ref.	0.67 (0.42, 1.07)	0.98 (0.62, 1.54)	0.660	1.01 (0.92, 1.12)	0.800	0.99 (0.86, 1.13)	0.832
	Multivariable adjusted model [2]	Ref.	0.47 (0.24, 0.90)	0.71 (0.40, 1.27)	0.700	0.94 (0.83, 1.07)	0.360	1.01 (0.86, 1.20)	0.887
Ursodeoxycholic acid (UDCA)	Tertile range (nM)	2.39 - 13.5	>13.5 - 40.3	>=40.3					
	Crude model [1]	Ref.	1.32 (0.82, 2.12)	**1.80 (1.14, 2.85)**	0.010	**1.20 (1.08, 1.33)**	***<0.001***	**1.18 (1.04, 1.39)**	**0.012**
	Multivariable adjusted model [2]	Ref.	1.61 (0.88, 2.93)	1.70 (0.96, 3.00)	0.200	**1.17 (1.03, 1.33)**	**0.020**	**1.24 (1.05, 1.47)**	**0.010**
Glycine-conjugated bile acids									

Glycocholic acid (GCA)	Tertile range (nM)	5.09 -106.1	>106.1 - 237.1	>=237.1					
	Crude model [1]	Ref.	**1.97 (1.03, 3.79)**	***8.31 (4.27, 16.17)***	<0.001	**1.99 (1.65, 2.39)**	***<0.001***	**1.65 (1.35, 2.01)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	1.69 (0.81, 3.55)	***9.15 (4.25, 19.71)***	<0.001	**2.11 (1.67, 2.66)**	***<0.001***	**1.80 (1.39, 2.32)**	***<0.001***
Glycochenodeoxycholic acid (GCDCA)	Tertile range (nM)	40.29 -509	>509 - 1074.3	>=1074					
	Crude model [1]	Ref.	1.75 (0.92, 3.32)	***6.95 (3.58, 13.50)***	<0.001	**2.04 (1.68, 2.49)**	***<0.001***	**1.70 (1.35, 2.13)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	1.27 (0.60, 2.69)	***6.20 (2.87, 13.36)***	<0.001	**2.08 (1.63, 2.64)**	***<0.001***	**1.76 (1.32, 2.35)**	***<0.001***
Glycodeoxycholic acid (GDCA)	Tertile range (nM)	2.39 - 126.1	>126.1 - 307.6	>=307.6					
	Crude model [1]	Ref.	1.47 (0.87, 2.49)	***2.59 (1.54, 4.35)***	<0.001	**1.39 (1.23, 1.58)**	***<0.001***	**1.22 (1.04, 1.42)**	**0.015**
	Multivariable adjusted model [2]	Ref.	1.14 (0.60, 2.14)	**2.15 (1.14, 4.04)**	0.006	**1.32 (1.14, 1.53)**	***<0.001***	**1.22 (1.01, 1.48)**	**0.039**
Glycohyocholic acid (GHCA)	Tertile range (nM)	1.19 - 6.7	>6.7 - 13.3	>=13.3					
	Crude model [1]	Ref.	1.21 (0.74, 1.98)	***2.49 (1.57, 3.94)***	<0.001	**1.67 (1.40, 1.99)**	***<0.001***	**1.56 (1.23, 1.99)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	1.21 (0.63, 2.32)	***3.43 (1.84, 6.38)***	<0.001	**1.97 (1.54, 2.53)**	***<0.001***	**1.65 (1.22, 2.23)**	***<0.001***
Glycolithocholic acid (GLCA)	Tertile range (nM)	1.19 - 9.6	>9.6 -37.3	>=37.3					
	Crude model [1]	Ref.	**1.81 (1.12, 2.94)**	**2.08 (1.26, 3.42)**	0.030	**1.21 (1.08, 1.37)**	***<0.001***	1.12 (0.95, 1.32)	0.174
	Multivariable adjusted model [2]	Ref.	**1.89 (1.04, 3.44)**	**1.93 (1.07, 3.48)**	0.120	**1.23 (1.06, 1.41)**	**<0.006**	1.07 (0.89, 1.30)	0.460
Glycoursodeoxycholic acid (GUDCA)	Tertile range (nM)	2.39 - 46.7	>46.7 - 111.3	>=111.3					
	Crude model [1]	Ref.	1.11 (0.66, 1.87)	***2.50 (1.54, 4.04)***	<0.001	**1.48 (1.29, 1.69)**	***<0.001***	**1.33 (1.13, 1.57)**	***<0.001***

	Multivariable adjusted model [2]	Ref.	1.09 (0.57, 2.11)	**2.62 (1.43, 4.79)**	<0.001	**1.50 (1.26, 1.79)**	***<0.001***	**1.42 (1.15, 1.76)**	***<0.001***
Taurine-conjugated bile acids									
Tauro-alfa-muricholic acid (TaMCA)	Tertile range (nM)	1.19 - 2.6	>2.6 - 6.1	>=6.1					
	Crude model [1]	Ref.	1.43 (0.80, 2.54)	***5.34 (3.05, 9.36)***	<0.001	**1.85 (1.57, 2.19)**	***<0.001***	**1.76 (1.42, 2.19)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	1.54 (0.75, 3.16)	***6.77 (3.35, 13.68)***	<0.001	**1.94 (1.58, 2.37)**	***<0.001***	**1.94 (1.47, 2.56)**	***<0.001***
Taurocholic acid (TCA)	Tertile range (nM)	1.19 - 12	>12 - 27.5	>=27.5					
	Crude model [1]	Ref.	**2.30 (1.09, 4.87)**	***14.40 (6.58,31.52)***	<0.001	**1.84 (1.56, 2.16)**	***<0.001***	**1.60 (1.34, 1.91)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	**2.47 (1.03, 5.90)**	***19.53 (7.55,50.51)***	<0.001	**1.88 (1.55, 2.28)**	***<0.001***	**1.68 (1.35, 2.08)**	***<0.001***
Taurochenodeoxycholic acid (TCDCA)	Tertile range (nM)	4.99 - 41.6	>41.6 - 96.3	>=96.3					
	Crude model [1]	Ref.	1.60 (0.80, 3.22)	***9.17 (4.57, 18.42)***	<0.001	**1.88 (1.59, 2.23)**	***<0.001***	**1.67 (1.38, 2.03)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	0.93 (0.39, 2.23)	***8.65 (3.73, 20.07)***	<0.001	**1.91 (1.57, 2.33)**	***<0.001***	**1.69 (1.34, 2.14)**	***<0.001***
Taurodeoxycholic acid (TDCA)	Tertile range (nM)	1.19 - 14.4	>14.4 - 37.7	>=37.7					
	Crude model [1]	Ref.	1.17 (0.66, 2.07)	***5.96 (3.16, 11.24)***	<0.001	**1.61 (1.40, 1.84)**	***<0.001***	**1.41 (1.20, 1.66)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	1.36 (0.67, 2.76)	***6.95 (3.13, 15.44)***	<0.001	**1.60 (1.35, 1.89)**	***<0.001***	**1.40 (1.15, 1.70)**	***<0.001***
Tauroursodeoxycholic acid (TUDCA)	Tertile range (nM)	1.19 - 2.5	>2.5 - 4.7	>=4.7					
	Crude model [1]	Ref.	1.08 (0.51, 2.29)	***4.73 (2.30, 9.72)***	<0.001	**1.84 (1.56, 2.18)**	***<0.001***	**1.61 (1.33, 1.96)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	1.48 (0.60, 3.67)	***6.04 (2.47, 14.75)***	<0.001	**1.92 (1.56, 2.37)**	***<0.001***	**1.65 (1.30, 2.08)**	***<0.001***

Total sum of all BAs [5]	Tertile range (nM)	272.8 - 1704.5	>1704.5 - 3218.6	>=3218.6						
	Crude model [1]	Ref.	1.48 (0.78, 2.82)	***5.59 (3.09, 10.11)***	<0.001	**2.23 (1.79, 2.76)**	***<0.001***	**1.77 (1.40, 2.25)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	1.52 (0.72, 3.23)	***6.07 (2.98, 12.37)***	<0.001	**2.30 (1.76, 3.00)**	***<0.001***	**1.91 (1.41, 2.59)**	***<0.001***

Odd Ratios (OR) and 95% confidence intervals (95% CI) or p-values that are **bolded** indicate statistically significant values. OR (95%CI) or p-values that are **both bolded and underlined** indicate statistical significance with Bonferroni correction for multiple testing. In linear models, the threshold of the Bonferroni correction for multiple testing p-value was calculated to be 0.003 (i.e. 0.05/17). In categorical models, the threshold for Bonferroni correction was calculated to be 0.0015 (i.e. 0.05/34).

[1] OR (95% CI) conditioned on the matching factors. The number of matched case-control sets where the HCC case has no or one abnormal liver function parameter=123. The number of matched case-control sets where the HCC case has between two to five abnormal liver function parameters=110.

[2] OR (95% CI) calculated with multivariable adjusted conditional regression models (adjustment factors: matching factors + body mass index, waist circumference, alcohol intake at recruitment, physical activity, smoking status, alcohol intake pattern and attained education level).

[3] Continuous models with either crude or multivariable adjustment, where OR (95% CI) are indicative of the HCC risk associated with a doubling of the concentration (nM) of the respective BA.

[4] Continuous models with either crude or multivariable adjustment plus additional adjustment for circulating GGT concentration (IU/L) included in the model as a continuous variable, where OR (95% CI) are indicative of the HCC risk associated with a doubling of the concentration (nM) of the respective BA. Risk estimates adjusted for Fatty Liver Index (FLI), a composite score incorporating GGT, triglycerides, BMI, and waist circumference, are also provided in Supplementary Table 2.

[5] Total BA sum is the sum of nM concentrations of each individual BA.

Table 3. Odds ratios (95 % confidence intervals) of HCC risk in relation to the doubling of bile acid concentrations (nM) among sub-groups of HCC cases based on their degree of liver dysfunctionality.

	Cases with None or 1 abnormal liver function parameter[1] N case-control pairs=123			**Cases with 2 to 5 abnormal liver function parameters[1]** N case-control pairs=110		
Unconjugated primary bile acids	**FD**	**OR (95% CI)[2]**	**p**	**FD**	**OR (95% CI)[2]**	**p**
Cholic acid (CA)	1.0	1.01 (0.91, 1.12)	0.880	2.1	**1.22 (1.06, 1.41)**	**0.005**
Chenodeoxycholic acid (CDCA)	1.2	1.05 (0.93, 1.18)	0.430	2.0	**1.19 (1.03, 1.39)**	**0.020**
Hyocholic acid (HCA)	1.2	1.08 (0.92, 1.28)	0.350	1.8	**1.39 (1.16, 1.65)**	***0.0003***
Unconjugated secondary/tertiary bile acids						
Deoxycholic acid (DCA)	1.0	1.00 (0.87, 1.15)	0.970	1.3	1.03 (0.89, 1.19)	0.710
Ursodeoxycholic acid (UDCA)	1.3	1.12 (0.98, 1.28)	0.100	1.8	**1.31 (1.11, 1.55)**	**0.002**
Glycine-conjugated bile acids						
Glycocholic acid (GCA)	1.4	**1.47 (1.21, 1.79)**	***0.0001***	6.4	**5.41 (2.42, 12.08)**	***<0.0001***
Glycochenodeoxycholic acid (GCDCA)	1.4	**1.48 (1.18, 1.86)**	***0.0008***	3.6	**3.83 (2.27, 6.45)**	***<0.0001***
Glycodeoxycholic acid (GDCA)	1.2	**1.18 (1.01, 1.39)**	**0.040**	2.6	**1.70 (1.37, 2.10)**	***<0.0001***
Glycohyocholic acid (GHCA)	1.3	**1.31 (1.02, 1.68)**	**0.040**	3.2	**2.08 (1.55, 2.78)**	***<0.0001***
Glycolithocholic acid (GLCA)	1.1	1.06 (0.90, 1.25)	0.490	2.2	**1.38 (1.15, 1.64)**	***0.0004***
Glycoursodeoxycholic acid (GUDCA)	1.2	**1.27 (1.07, 1.52)**	**0.008**	2.7	**1.77 (1.40, 2.23)**	***<0.0001***
Taurine-conjugated bile acids						
Tauro-alfa-muricholic acid (TaMCA)	2.2	**1.60 (1.29, 1.98)**	***<0.0001***	8.7	**2.23 (1.65, 3.02)**	***<0.0001***
Taurocholic acid (TCA)	2.5	**1.50 (1.26, 1.80)**	***<0.0001***	16.5	**2.91 (1.81, 4.68)**	***<0.0001***
Taurochenodeoxycholic acid (TCDCA)	2.0	**1.54 (1.27, 1.87)**	***<0.0001***	7.2	**2.78 (1.85, 4.18)**	***<0.0001***
Taurodeoxycholic acid (TDCA)	2.1	**1.35 (1.15, 1.59)**	***0.0003***	3.8	**2.14 (1.61, 2.83)**	***<0.0001***
Tauroursodeoxycholic acid (TUDCA)	2.6	**1.57 (1.27, 1.94)**	***<0.0001***	5.8	**2.30 (1.68, 3.14)**	***<0.0001***
Total sum of all BAs	1.5	**1.55 (1.22, 1.97)**	***0.0003***	3.2	**5.61 (2.78, 11.33)**	***<0.0001***

Odd Ratios (OR) and 95% confidence intervals (95% CI) indicate HCC risk per doubling of the concentration. OR (95%CI) or p-values that are bolded indicate statistically significant values. OR (95%CI) or p-values that are both bolded and underlined

indicate statistical significance with Bonferroni correction for multiple testing. The threshold of the Bonferroni correction for multiple testing p-value was calculated to be 0.0015 (i.e. 0.05/34). FD: Fold difference (FD) between median levels of BA in HCC cases compared to matched controls. [1] abnormal liver parameters refer to ALT>55 U/L, AST>34 U/L, GGT >64 U/L for men and > 36 U/L for women, ALP > 150 U/L, total bilirubin > 20.5 μmol/L. [2] OR (95% CI) calculated with multivariable adjusted conditional regression models (adjustment factors: matching factors + body mass index, waist circumference, alcohol intake at recruitment, physical activity, smoking status, alcohol intake pattern and attained education level). The number of matched case-control sets where the HCC case has no or one abnormal liver function parameter=123. The number of matched case-control sets where the HCC case has between two to five abnormal liver function parameter=110.

Supplementary Methods and Data

Pre-diagnostic alterations in circulating bile acid profiles in the development of hepatocellular carcinoma

Magdalena Stepien, Marina Lopez-Nogueroles, Agustin Lahoz, Tilman Kühn, Gabriel Perlemuter, Cosmin Voican, Dragos Ciocan, Marie-Christine Boutron-Ruault, Eugene Jansen, Vivian Viallon, Michael Leitzmann, Anne Tjønneland, Gianluca Severi, Francesca Romana Mancini, Catherine Dong, Rudolf Kaaks, Renee Turzanski Fortner, Manuela M. Bergmann, Heiner Boeing, Antonia Trichopoulou, Anna Karakatsani, Eleni Peppa, Domenico Palli, Vittorio Krogh, Rosario Tumino, Carlotta Sacerdote, Salvatore Panico, H. Bas Bueno-de-Mesquita, Guri Skeie, Susana Merino, Raul Zamora Ros, Maria Jose Sánchez, Pilar Amiano, Jose Mª Huerta, Aurelio Barricarte, Klas Sjöberg, Bodil Ohlsson, Hanna Nyström, Marten Werner, Aurora Perez-Cornago, Julie A. Schmidt, Heinz Freisling, Augustin Scalbert, Elisabete Weiderpass, Sofia Christakoudi, Marc J. Gunter, Mazda Jenab.

Table of Contents:

Supplementary methods

Exclusion criteria

We excluded: 25,184 subjects with prevalent cancer other than non-melanoma skin cancer, 4,148 with incomplete follow up data or missing information on the date of diagnosis, 4,982 with missing dietary information, 60 with missing lifestyle information, 1,217 with missing lifestyle and dietary information, 9,573 who were at the top or bottom 1% of the distribution of the ratio of reported energy intake to energy requirement, resulting in 476,160 participants in the analytic cohort.

For each identified case, the histology and the methods used to diagnose the cancer were reviewed to additionally exclude metastatic cases, those with ineligible histology codes or other types of hepatobiliary cancer (n=169).

Laboratory analyses

For bile acids (BA) quantification, 50 µL of plasma samples were spiked with deuterated internal standards stock solution. Methanol was added to precipate proteins which were removed by centrifugation. Supernatants were dried and reconstituted in 50 µL of methanol:water (50:50, V/V). Samples were analysed using an Acquity UPLC system (Waters, UK) equipped with an Acquity UPLC BEH C18 column (1.7µm, 2.1 x 100 mm; Waters). The mass spectrometry (**MS**) analysis was performed using a Waters Xevo TQ-S mass spectrometer (Waters) with an Electrospray ionization (**ESI**) source working in the negative-ion mode. Coefficients of variation for quality control samples for all batches ranged from 6.0 (glycochenodeoxycholic acid, GCDCA) to below 20.0%, except for taurohyocholic acid (THCA) which was 22.3% and was thus excluded from further statistical analyses. Samples were analysed in seven batches, each containing cases and their matched controls (Analytical Unit, Health Research Institute Hospital La Fe, Valencia, Spain).

Hepatitis B seropositivity was assessed using either the ARCHITECT HBsAg chemiluminescent micro-particle immunoassay (CMIA) from Abbott Diagnostics (France) or the HBS-Ag test from DIAsource (Belgium). Hepatitis C seropositivity was detected using either the ARCHITECT anti-HCV CMIA (Abbott Diagnostics, France) or by Elisa (HCV-Ab test, DRG International). Liver function biomarkers (gamma-glutamyltransferase, **GGT**; alanine aminotransferase, **ALT**; aspartate aminotransferase, **AST**; alkaline phosphatase, **ALP**; total bilirubin and albumin), high-

sensitivity C-reactive protein (**hsCRP**), alpha-fetoprotein (**AFP**), serum lipids and glycated haemoglobin (**HbA1c**) were measured using standard protocols on either the ARCHITECT c Systems™, the AEROSET System (Abbott Diagnostics), or a DxC800 auto-analyzer (Beckman-Coulter, USA)(Centre de Biologie République, Lyon France; National Institute for Health Protection, National Institute for Public Health and the Environment, Bilthoven, Netherlands). Clinical thresholds for liver function biomarkers (i.e. ALT>55 IU/L (n controls=12; n HCC cases=51), AST>34 IU/L (n controls=26; n HCC cases=126), GGT >64 IU/L for men (n controls=14; n HCC cases=88) and > 36 U/L for women (n controls=9; n HCC cases=27), ALP > 150 U/L (n controls=0; n HCC cases=13), total bilirubin > 20.5 µmol/L (n controls=5; n HCC cases=18) were provided by the laboratory and utilized to assess the total number of abnormal liver function parameters per subject.

Calculation of Fatty Liver Index and the Metabolic Syndrome Score

The fatty liver index (FLI) was computed according to the equation of Bedogni [24]:

$$\mathrm{FLI} = \frac{\exp^{(0.953 * \log(\mathrm{TG}) + 0.139 * \mathrm{BMI} + 0.718 * \log(\mathrm{GGT}) + 0.053 * \text{waist circumference} - 15.745)}}{1 + \exp^{(0.953 * \log(\mathrm{TG}) + 0.139 * \mathrm{BMI} + 0.718 * \log(\mathrm{GGT}) + 0.053 * \text{waist circumference} - 15.745)}} * 100,$$

where TG refers to triglycerides (mg/dL), BMI to body mass index (kg/m^2). Waist circumference was measured in cm and GGT in IU/L.

The metabolic syndrome score was computed based on the harmonized definition [25].

Statistical analyses

Missing values for waist circumference (n=24), TG (n=4) and HDL cholesterol (n=5) were replaced using the 'proc mi' command in SAS with 10 imputations in the model containing also case-control and smoking status, BMI and alcohol intake at baseline.

We calculated Spearman correlation coefficients (adjusted for sex and age) among the BAs (amongst matched controls only) and with liver function biomarkers and scores (comparing HCC cases and matched controls). A correlation heatmap was created using R studio (version 3.5.1) in order to illustrate the correlations. Loess curves were constructed to visualize levels of BAs by follow-up time, where different follow-up time of cases was assigned to their respective controls.

Supplementary Table 1. Odds ratios and 95 % confidence intervals of HCC risk across tertiles of individual bile acids expressed as relative proportions (% of total BA sum), i.e. a change in the level of each individual bile acid is assessed while the total bile acid concentration is held constant.

Individual Bile Acids Expressed as Relative Proportions (% of total)		**Tertile 1**	**Tertile 2**	**Tertile 3**	**p-trend**	**Continuous (per doubling of %)**	**p-value**
Unconjugated primary bile acids							
Cholic acid (CA)	Tertile range	0 - 1.8	>1.8 - 7.5	>=7.5			
	Crude model [1]	Ref.	**0.52 (0.32, 0.85)**	**0.46 (0.28, 0.76)**	0.010	**0.79 (0.71, 0.88)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	0.65 (0.36, 1.18)	**0.48 (0.26, 0.91)**	0.040	**0.81 (0.71, 0.92)**	***0.002***
Chenodeoxycholic acid (CDCA)	Tertile range	0 - 5.1	>5.1 - 12.0	>=12.0			
	Crude model [1]	Ref.	***0.35 (0.22, 0.57)***	***0.42 (0.26, 0.69)***	0.002	**0.75 (0.67, 0.85)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	***0.32 (0.18, 0.59)***	**0.39 (0.21, 0.72)**	0.008	**0.72 (0.62, 0.84)**	***<0.001***
Hyocholic acid (HCA)	Tertile range	0 - 0.2	>0.2- 0.4	>=0.4			
	Crude model [1]	Ref.	**0.57 (0.37, 0.88)**	***0.35 (0.21, 0.57)***	***<0.001***	**0.70 (0.61, 0.81)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	0.58 (0.34, 1.00)	**0.47 (0.26, 0.86)**	0.030	**0.75 (0.64, 0.89)**	***0.001***
Unconjugated secondary/tertiary bile acids							
Deoxycholic acid (DCA)	Tertile range	0 - 7.4	>7.4 - 14.8	>=14.8			
	Crude model [1]	Ref.	**0.52 (0.33, 0.82)**	***0.33 (0.20, 0.55)***	***<0.001***	**0.71 (0.63, 0.81)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	**0.48 (0.27, 0.85)**	***0.28 (0.15, 0.54)***	***<0.001***	**0.66 (0.57, 0.77)**	***<0.001***
Ursodeoxycholic acid (UDCA)	Tertile range	0 - 0.6	>0.6 - 1.6	>=1.6			
	Crude model [1]	Ref.	**0.55 (0.35, 0.87)**	**0.45 (0.27, 0.74)**	0.003	**0.81 (0.73, 0.91)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	**0.50 (0.29, 0.87)**	**0.49 (0.26, 0.92)**	0.040	**0.78 (0.67, 0.90)**	***0.001***

Individual Bile Acids Expressed as Relative Proportions (% of total)		**Tertile 1**	**Tertile 2**	**Tertile 3**	**p-trend**	**Continuous (per doubling of %)**	**p-value**
Glycine-conjugated bile acids							
Glycocholic acid (GCA)	Tertile range	0.6 -5.6	>5.6 - 9.0	>=9.0			
	Crude model [1]	Ref.	0.98 (0.56, 1.73)	***3.35 (1.99, 5.65)***	***<0.001***	**1.96 (1.54, 2.50)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	0.90 (0.43, 1.87)	***3.91 (1.99, 7.70)***	***<0.001***	**2.13 (1.58, 2.86)**	***<0.001***
Glycochenodeoxycholic acid (GCDCA)	Tertile range	2.6 - 26.9	>26.9 - 40.3	>40.3			
	Crude model [1]	Ref.	1.34 (0.83, 2.14)	0.99 (0.59, 1.66)	0.920	0.97 (0.74, 1.27)	0.800
	Multivariable adjusted model [2]	Ref.	1.46 (0.83, 2.57)	1.05 (0.55, 1.97)	0.880	0.99 (0.71, 1.39)	0.960
Glycodeoxycholic acid (GDCA)	Tertile range	0 - 6.4	>6.4 - 11.9	>=11.9			
	Crude model [1]	Ref.	**0.59 (0.39, 0.90)**	**0.54 (0.34, 0.85)**	0.007	**0.83 (0.73, 0.95)**	**0.007**
	Multivariable adjusted model [2]	Ref.	**0.43 (0.24, 0.76)**	**0.44 (0.24, 0.78)**	0.005	**0.76 (0.64, 0.91)**	***0.002***
Glycohyocholic acid (GHCA)	Tertile range	0 - 0.3	>0.3- 0.6	>=0.6			
	Crude model [1]	Ref.	**0.50 (0.31, 0.79)**	***0.44 (0.27, 0.70)***	***0.001***	**0.74 (0.62, 0.88)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	0.64 (0.36, 1.12)	0.60 (0.33, 1.06)	0.110	0.87 (0.70, 1.08)	0.200
Glycolithocholic acid (GLCA)	Tertile range	0 - 0.5	>0.5 -1.2	>=1.2			
	Crude model [1]	Ref.	**0.62 (0.40, 0.95)**	***0.33 (0.20, 0.54)***	***<0.001***	**0.73 (0.64, 0.83)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	**0.56 (0.33, 0.95)**	***0.33 (0.18, 0.61)***	***0.001***	**0.73 (0.62, 0.86)**	***<0.001***
Glycoursodeoxycholic acid (GUDCA)	Tertile range	0.1 - 2.3	>2.3 - 3.9	>=3.9			
	Crude model [1]	Ref.	**0.51 (0.32, 0.81)**	**0.58 (0.36, 0.94)**	0.060	**0.80 (0.68, 0.94)**	***0.006***

Individual Bile Acids Expressed as Relative Proportions (% of total)		**Tertile 1**	**Tertile 2**	**Tertile 3**	**p-trend**	**Continuous (per doubling of %)**	**p-value**
	Multivariable adjusted model [2]	Ref.	0.63 (0.35, 1.14)	0.64 (0.35, 1.15)	0.190	**0.80 (0.66, 0.98)**	**0.030**
Taurine-conjugated bile acids							
Tauro-alfa-muricholic acid (TaMCA)	Tertile range	0 - 0.1	>0.1 - 0.2	>=0.2			
	Crude model [1]	Ref.	0.79 (0.49, 1.28)	**1.93 (1.20, 3.11)**	***0.001***	**1.31 (1.15, 1.48)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	1.00 (0.55, 1.81)	**2.56 (1.40, 4.68)**	***0.001***	**1.47 (1.25, 1.74)**	***<0.001***
Taurocholic acid (TCA)	Tertile range	0 - 0.5	>0.5 - 1.2	>=1.2			
	Crude model [1]	Ref.	1.62 (0.90, 2.90)	***6.56 (3.58, 12.01)***	***<0.001***	1.73 (1.48, 2.03)	***<0.001***
	Multivariable adjusted model [2]	Ref.	1.70 (0.81, 3.57)	***9.79 (4.48, 21.39)***	***<0.001***	1.83 (1.50, 2.22)	***<0.001***
Taurochenodeoxycholic acid (TCDCA)	Tertile range	0 - 2.0	>2 - 4.2	>=4.2			
	Crude model [1]	Ref.	1.39 (0.79, 2.42)	***3.16 (1.90, 5.25)***	***<0.001***	1.65 (1.40, 1.95)	***<0.001***
	Multivariable adjusted model [2]	Ref.	1.50 (0.76, 2.98)	***3.76 (1.92, 7.37)***	***<0.001***	1.74 (1.40, 2.15)	***<0.001***
Taurodeoxycholic acid (TDCA)	Tertile range	0 - 0.7	>0.7 - 1.6	>=1.6			
	Crude model [1]	Ref.	0.90 (0.55, 1.45)	1.00 (0.92, 2.38)	0.070	1.09 (0.97, 1.23)	0.140
	Multivariable adjusted model [2]	Ref.	0.87 (0.48, 1.57)	1.11 (0.59, 2.07)	0.680	1.05 (0.89, 1.22)	0.590
Tauroursodeoxycholic acid (TUDCA)	Tertile range	0 - 0.1	>0.1 - 0.2	>=0.2			
	Crude model [1]	Ref.	0.94 (0.61, 1.47)	**1.61 (1.03, 2.51)**	0.020	**1.35 (1.15, 1.59)**	***<0.001***
	Multivariable adjusted model [2]	Ref.	1.03 (0.60, 1.79)	**1.83 (1.03, 3.26)**	0.020	**1.39 (1.13, 1.72)**	***0.002***

Odd Ratios (OR), 95% confidence intervals (95% CI) or p-values that are **bolded** indicate statistically significant values. OR (95%CI) or p-values that are **both bolded and underlined** indicate statistical significance with Bonferroni correction for multiple testing. In linear models, the threshold of

the Bonferroni correction for multiple testing p-value was calculated to be 0.003 (i.e. 0.05/17). In categorical models, the threshold for Bonferroni correction was calculated to be 0.0015 (i.e. 0.05/34).

[1] OR (95% CI) conditioned on the matching factors.

[2] OR (95% CI) calculated with multivariable adjusted conditional regression models (adjustment factors: matching factors + body mass index, waist circumference, alcohol intake at recruitment, physical activity, smoking status, alcohol intake pattern and attained education level).

Supplementary Table 2a: Odds ratios (95 % confidence intervals) of HCC risk with individual bile acids (BA). Values are per doubling of BA concentration in multivariable adjusted models and in sensitivity analyses.

Individual Bile Acids Plasma concentrations (nM)		Multivariable Adjusted Model with Further Adjustment for FLI		Sensitivity Analyses based on Multivariable Adjusted Models			
				Cases Diagnosed >2 years Post-Recruitment (n=209 case-control pairs)		Cases Without Hepatitis B/C Infection at Recruitment (n=114 case-control pairs)	
		OR (95% CI)*	*p*	OR (95% CI)*	*p*	OR (95% CI)*	*p*
Unconjugated primary bile acids	Cholic acid (CA)	1.13 (1.01, 1.26)	*3.61E-02*	1.13 (1.00, 1.26)	*4.20E-02*	1.13 (0.96, 1.34)	*1.51E-01*
	Chenodeoxycholic acid (CDCA)	1.09 (0.96, 1.23)	*1.73E-01*	1.09 (0.96, 1.23)	*1.87E-01*	1.13 (0.94, 1.36)	*1.83E-01*
	Hyocholic acid (HCA)	1.32 (1.13, 1.54)	***6.23E-04***	1.26 (1.07, 1.47)	*4.95E-03*	1.17 (0.93, 1.48)	*1.90E-01*
Unconjugated secondary / tertiary bile acids	Deoxycholic acid (DCA)	0.94 (0.81, 1.08)	*3.57E-01*	0.98 (0.86, 1.13)	*7.94E-01*	1.03 (0.83, 1.29)	*7.77E-01*
	Ursodeoxycholic acid (UDCA)	1.18 (1.03, 1.36)	*1.87E-02*	1.24 (1.07, 1.43)	*3.96E-03*	2.23 (1.53, 3.25)	***2.77E-05***
Glycine-conjugated bile acids	Glycocholic acid (GCA)	2.08 (1.63, 2.67)	***6.90E-09***	2.18 (1.64, 2.88)	***6.33E-08***	1.83 (1.35, 2.46)	***8.12E-05***
	Glycochenodeoxycholic acid (GCDCA)	2.09 (1.59, 2.74)	***9.63E-08***	1.53 (1.26, 1.85)	***1.81E-05***	1.88 (1.28, 2.74)	***1.16E-03***
	Glycodeoxycholic acid (GDCA)	1.30 (1.11, 1.54)	***1.46E-03***	2.22 (1.70, 2.89)	***4.30E-09***	2.74 (1.68, 4.46)	***5.42E-05***
	Glycohyocholic acid (GHCA)	1.92 (1.49, 2.49)	***6.73E-07***	1.32 (1.13, 1.54)	***6.08E-04***	2.32 (1.49, 3.61)	***1.85E-04***
	Glycolithocholic acid (GLCA)	1.18 (1.01, 1.38)	*3.38E-02*	1.91 (1.47, 2.48)	***1.51E-06***	1.48 (1.14, 1.92)	*3.30E-03*
	Glycoursodeoxycholic acid (GUDCA)	1.52 (1.25, 1.83)	***2.02E-05***	1.18 (1.01, 1.37)	*3.53E-02*	1.99 (1.32, 3.00)	***1.06E-03***

Taurine-conjugated bile acids	Tauro-alfa-muricholic acid (TaMCA)	1.89 (1.53, 2.34)	***4.49E-09***	1.96 (1.57, 2.44)	***3.40E-09***	1.92 (1.36, 2.71)	***2.04E-04***
	Taurocholic acid (TCA)	1.84 (1.50, 2.25)	***4.10E-09***	1.57 (1.31, 1.87)	***6.53E-07***	1.22 (0.96, 1.56)	*1.07E-01*
	Taurochenodeoxycholic acid (TCDCA)	1.87 (1.51, 2.31)	***7.20E-09***	2.01 (1.58, 2.57)	***1.98E-08***	2.14 (1.51, 3.04)	***2.13E-05***
	Taurodeoxycholic acid (TDCA)	1.55 (1.30, 1.84)	***1.04E-06***	1.92 (1.55, 2.38)	***2.20E-09***	2.66 (1.71, 4.13)	***1.35E-05***
	Tauroursodeoxycholic acid (TUDCA)	1.89 (1.52, 2.34)	***9.60E-09***	2.03 (1.61, 2.57)	***2.70E-09***	1.34 (1.07, 1.67)	*1.04E-02*
	Total sum of BAs	2.27 (1.71, 3.01)	***1.55E-08***	2.59 (1.87, 3.57)	***7.30E-09***	3.15 (1.77, 5.60)	***9.48E-05***

Results are for (a) multivariable adjusted models with additional adjustment for fatty liver index (FLI) or liver function score and (b) sensitivity analyses based on (i) cases diagnosed after two years from recruitment, and (ii) cases without hepatitis B/C infection at recruitment into the cohort.
Numbers in red indicate p<0.05. P-values that are **both bolded and underlined** indicate statistical significance with Bonferroni correction for multiple testing. Bonferroni correction for multiple testing p value threshold = 0.003.
* Odds ratios (OR, 95% CI) calculated with multivariable adjusted conditional regression models (adjustment factors: matching factors + body mass index, waist circumference, alcohol intake at recruitment, physical activity, smoking status, alcohol intake pattern and attained education level). OR represent doubling of concentration/percent contribution. Relative BA proportions calculated as relative percent of the sum of all BA.

Supplementary Table 2b: Odds ratios (95 % confidence intervals) of HCC risk with individual bile acids (BA). Values are per doubling of relative proportions of plasma BA in the total plasma BA pool, assessed in multivariable adjusted models and in sensitivity analyses.

Individual Bile Acids Expressed as Relative Proportions (% of total)		**Multivariable Adjusted Model with Further Adjustment for FLI**		**Sensitivity Analyses based on Multivariable Adjusted Models**			
				Cases Diagnosed >2 years Post-Recruitment (n=209 case-control pairs)		**Cases Without Hepatitis B/C Infection at Recruitment (n=114 case-control pairs)**	
		OR (95% CI)*	***p***	**OR (95% CI)***	***p***	**OR (95% CI)***	***p***
Unconjugated primary bile acids	Cholic acid (CA)	0.85 (0.74, 0.97)	*1.64E-02*	0.84 (0.73, 0.96)	*1.31E-02*	0.85 (0.71, 1.03)	*1.01E-01*
	Chenodeoxycholic acid (CDCA)	0.76 (0.65, 0.89)	***8.60E-04***	0.76 (0.64, 0.89)	***6.19E-04***	0.79 (0.63, 1.00)	*4.57E-02*
	Hyocholic acid (HCA)	0.81 (0.67, 0.97)	*1.88E-02*	0.76 (0.64, 0.91)	***2.68E-03***	0.70 (0.53, 0.91)	*9.25E-03*
Unconjugated secondary / tertiary bile acids	Deoxycholic acid (DCA)	0.67 (0.57, 0.79)	***2.29E-06***	0.68 (0.57, 0.80)	***5.35E-06***	0.65 (0.49, 0.85)	***1.78E-03***
	Ursodeoxycholic acid (UDCA)	0.82 (0.70, 0.95)	*8.08E-03*	0.83 (0.71, 0.97)	*1.97E-02*	0.91 (0.73, 1.12)	*3.74E-01*
Glycine-conjugated bile acids	Glycocholic acid (GCA)	2.04 (1.47, 2.83)	***1.98E-05***	2.10 (1.50, 2.93)	***1.46E-05***	1.76 (1.15, 2.69)	*8.68E-03*
	Glycochenodeoxycholic acid (GCDCA)	0.96 (0.67, 1.38)	*8.17E-01*	0.89 (0.62, 1.29)	*5.50E-01*	0.78 (0.43, 1.41)	*4.09E-01*
	Glycodeoxycholic acid (GDCA)	0.75 (0.62, 0.91)	*3.59E-03*	0.78 (0.64, 0.95)	*1.26E-02*	0.82 (0.62, 1.09)	*1.74E-01*
	Glycohyocholic acid (GHCA)	0.91 (0.72, 1.15)	*4.31E-01*	0.77 (0.60, 0.98)	*3.50E-02*	0.72 (0.51, 1.03)	*7.05E-02*
	Glycolithocholic acid (GLCA)	0.70 (0.58, 0.85)	***2.60E-04***	0.70 (0.58, 0.85)	***2.67E-04***	0.68 (0.51, 0.90)	*6.41E-03*
	Glycoursodeoxycholic acid (GUDCA)	0.82 (0.67, 1.01)	*6.15E-02*	0.84 (0.68, 1.05)	*1.19E-01*	1.07 (0.78, 1.46)	*6.87E-01*
Taurine-conjugated bile acids	Tauro-alfa-muricholic acid (TaMCA)	1.45 (1.22, 1.72)	***2.66E-05***	1.45 (1.22, 1.73)	***4.05E-05***	1.21 (0.95, 1.55)	*1.30E-01*

Taurocholic acid (TCA)	1.76 (1.43, 2.16)	***8.26E-08***	1.82 (1.46, 2.27)	***7.74E-08***	1.67 (1.26, 2.20)	***3.17E-04***
Taurochenodeoxycholic acid (TCDCA)	1.64 (1.32, 2.04)	***1.11E-05***	1.69 (1.34, 2.14)	***1.22E-05***	1.58 (1.17, 2.14)	***2.65E-03***
Taurodeoxycholic acid (TDCA)	1.02 (0.86, 1.20)	*8.46E-01*	1.06 (0.89, 1.26)	*5.08E-01*	1.11 (0.86, 1.42)	*4.34E-01*
Tauroursodeoxycholic acid (TUDCA)	1.37 (1.09, 1.71)	*6.06E-03*	1.42 (1.13, 1.79)	***2.89E-03***	1.91 (1.30, 2.81)	***9.53E-04***

Results are for (a) multivariable adjusted models with additional adjustment for fatty liver index (FLI) or liver function score and (b) sensitivity analyses based on (i) cases diagnosed after two years from recruitment, and (ii) cases without hepatitis B/C infection at recruitment into the cohort.
Numbers in red indicate p<0.05. P-values that are **both bolded and underlined** indicate statistical significance with Bonferroni correction for multiple testing.
Bonferroni correction for multiple testing p value threshold = 0.003.
** Odds ratios (OR, 95% CI) calculated with multivariable adjusted conditional regression models (adjustment factors: matching factors + body mass index, waist circumference, alcohol intake at recruitment, physical activity, smoking status, alcohol intake pattern and attained education level). OR represent doubling of concentration/percent contribution. Relative BA proportions calculated as relative percent of the sum of all BA.*

Supplementary Table 3a: Odds ratios (95 % confidence intervals) of HCC risk with groupings and ratios of bile acids (BA). Values are per doubling of plasma BA concentration (nM), in the crude model, multivariate adjusted model, and multivariate adjusted models with further adjustment for fatty liver index or hepatitis B/C infection status.

Groupings and Ratios of Bile Acids Plasma Concentrations (nM)	**Median Values (nM)**		**Crude Model [1] (Matching Factors)**		**Multivariable Adjusted Model [2]**		**Multivariable Adjusted Model [3] + Fatty Liver Index**		**Multivariable Adjusted Model [4] + Hepatitis B/C Status**	
	Case	**Control**	**OR (95% CI)**	***p***	**OR (95% CI)**	***p***	**OR (95% CI)**	***p***	**OR (95% CI)**	***p***
Bile Acid Groupings										
Unconjugated BA (CA , CDCA , DCA , UDCA)	847.4	599.9	1.18 (1.04, 1.33)	*8.61E-03*	1.10 (0.95, 1.28)	*1.87E-01*	1.11 (0.95, 1.30)	*1.91E-01*	1.18 (0.99, 1.40)	*6.85E-02*
Primary unconjugated BA (CA, CDCA)	394.8	245.6	1.11 (1.01, 1.22)	*2.76E-02*	1.09 (0.97, 1.22)	*1.44E-01*	1.11 (0.98, 1.25)	*1.02E-01*	1.14 (1.00, 1.31)	*5.68E-02*
Secondary unconjugated BA (DCA, UDCA)	360.6	302.0	1.11 (0.99, 1.24)	*7.02E-02*	1.00 (0.87, 1.16)	*9.54E-01*	1.01 (0.87, 1.18)	*8.98E-01*	1.09 (0.91, 1.30)	*3.42E-01*
Conjugated BA (GCA , GDCA , GDCA , GHCA , GLCA , GUDCA, TCA , TCDCA , TDCA , THCA , TUDCA , TaMCA)	3538.6	1560.6	2.24 (1.80, 2.78)	***0.00E+00***	2.31 (1.77, 3.00)	***4.00E-10***	2.32 (1.74, 3.11)	***1.31E-08***	2.23 (1.64, 3.05)	***4.26E-07***
Primary conjugated (GCA, GCDCA, TCA, TCDCA)	2584.1	1092.8	2.13 (1.73, 2.60)	***0.00E+00***	2.20 (1.72, 2.82)	***4.00E-10***	2.22 (1.69, 2.92)	***1.37E-08***	2.16 (1.60, 2.91)	***4.30E-07***
Secondary conjugated (GDCA, GUDCA, TDCA, TUDCA)	638.4	325.8	1.79 (1.51, 2.13)	***0.00E+00***	1.71 (1.39, 2.10)	***3.05E-07***	1.73 (1.37, 2.17)	***3.07E-06***	1.70 (1.34, 2.16)	***1.49E-05***
Taurine-conjugated (TCA , TCDCA , TDCA , THCA , TUDCA , TaMCA)	493.0	122.8	2.05 (1.69, 2.49)	***0.00E+00***	2.09 (1.67, 2.62)	***2.00E-10***	2.05 (1.61, 2.62)	***7.00E-09***	2.06 (1.57, 2.69)	***1.32E-07***
Glycine-conjugated (GCA , GCDCA , GDCA , GHCA , GLCA , GUDCA)	2902.6	1413.0	2.18 (1.76, 2.69)	***0.00E+00***	2.23 (1.72, 2.89)	***1.10E-09***	2.24 (1.69, 2.97)	***2.73E-08***	2.12 (1.57, 2.87)	***9.00E-07***

Hydrophylic (CA, UDCA)	181.4	107.0	1.15 (1.05, 1.27)	***2.64E-03***	1.16 (1.03, 1.30)	*1.38E-02*	1.17 (1.04, 1.33)	*1.27E-02*	1.18 (1.03, 1.36)	*1.63E-02*
Hydrophobic (DCA)	308.3	256.6	1.01 (0.92, 1.12)	*8.04E-01*	0.94 (0.83, 1.07)	*3.60E-01*	0.94 (0.81, 1.08)	*3.57E-01*	1.00 (0.86, 1.17)	*9.99E-01*
Glycine-conjugated hydrophylic (GCA, GUDCA)	739.8	265.9	1.99 (1.66, 2.40)	***2.20E-13***	2.11 (1.67, 2.65)	***2.41E-10***	2.10 (1.64, 2.68)	***3.70E-09***	2.02 (1.55, 2.64)	***2.60E-07***
Glycine-conjugated hydrophobic (GDCA, GLCA)	367.9	218.0	1.41 (1.24, 1.61)	***2.71E-07***	1.34 (1.15, 1.56)	***2.02E-04***	1.32 (1.11, 1.57)	***1.36E-03***	1.31 (1.09, 1.56)	***3.11E-03***
Taurine-conjugated hydrophylic (TCA, TUDCA)	94.2	22.9	1.89 (1.59, 2.24)	***2.70E-13***	1.93 (1.58, 2.36)	***2.01E-10***	1.90 (1.54, 2.36)	***4.40E-09***	1.91 (1.50, 2.43)	***1.35E-07***
Taurine-conjugated hydrophobic (TDCA)	63.7	24.4	1.61 (1.40, 1.84)	***2.65E-11***	1.60 (1.35, 1.89)	***4.13E-08***	1.55 (1.30, 1.84)	***1.04E-06***	1.55 (1.28, 1.88)	***5.34E-06***
Total CA (CA , GCA , TCA)	1051.7	332.1	1.70 (1.47, 1.97)	***0.00E+00***	1.80 (1.50, 2.17)	***6.00E-10***	1.78 (1.46, 2.16)	***7.70E-09***	1.76 (1.41, 2.20)	***5.68E-07***
Total CDCA (CDCA , GCDCA , TCDCA)	2811.7	1141.3	2.03 (1.67, 2.45)	***0.00E+00***	2.07 (1.63, 2.62)	***1.90E-09***	2.07 (1.60, 2.67)	***3.11E-08***	2.10 (1.56, 2.82)	***8.73E-07***
Total DCA (DCA , GDCA , TDCA)	759.4	524.9	1.33 (1.17, 1.52)	***1.33E-05***	1.24 (1.07, 1.45)	*5.48E-03*	1.23 (1.04, 1.47)	*1.74E-02*	1.28 (1.06, 1.54)	*9.98E-03*
Total HCA (GHCA , HCA)	36.3	21.9	1.66 (1.40, 1.97)	***7.25E-09***	1.90 (1.51, 2.40)	***5.22E-08***	1.89 (1.48, 2.42)	***3.09E-07***	1.78 (1.37, 2.30)	***1.52E-05***
Total UDCA (GUDCA ,TUDCA , UDCA)	215.6	117.2	1.52 (1.32, 1.75)	***1.10E-08***	1.53 (1.28, 1.83)	***2.40E-06***	1.55 (1.28, 1.88)	***7.06E-06***	1.61 (1.29, 2.00)	***1.95E-05***
Bile Acid Ratios										
Primary to secondary unconjugated BA	1.34	0.95	1.02 (0.98, 1.06)	*3.68E-01*	1.05 (1.00, 1.10)	*7.07E-02*	1.04 (0.99, 1.10)	*1.26E-01*	1.04 (0.98, 1.10)	*2.45E-01*
Glycine-conjugated/Taurine-conjugated BA	5.53	9.46	0.92 (0.89, 0.95)	***1.19E-06***	0.92 (0.88, 0.96)	***3.28E-05***	0.93 (0.90, 0.97)	***8.31E-04***	0.94 (0.91, 0.98)	*6.92E-03*

UDCA/CDCA	0.15	0.13	1.62 (0.88, 2.96)	*1.20E-01*	1.54 (0.66, 3.59)	*3.16E-01*	1.42 (0.60, 3.34)	*4.28E-01*	1.28 (0.48, 3.43)	*6.19E-01*
DCA/CA	1.74	2.64	0.99 (0.97, 1.01)	*1.80E-01*	0.97 (0.95, 0.99)	***5.19E-03***	0.97 (0.95, 0.99)	*4.58E-03*	0.97 (0.95, 1.00)	*2.41E-02*

[1] Odds Ratios (OR) and 95% confidence intervals (95% CI) conditioned on the matching factors.

[2] OR (95% CI) calculated with multivariable adjusted conditional regression models (adjustment factors: matching factors + body mass index, waist circumference, alcohol intake at recruitment, physical activity, smoking status, alcohol intake pattern and attained education level).

[3] OR (95% CI) calculated with multivariable adjusted conditional regression models as in [2] above plus additional adjustment for FLI, please see Table 1 and Supplementary Methods for additional details.

[4] OR (95% CI) calculated with multivariable adjusted conditional regression models as in [2] above plus additional adjustment for hepatitis B/C positivity, please see Table 1 for additional details.

OR represent doubling of concentration contribution. Groupings are calculated as a sum of individual BA listed in brackets.

Numbers in red indicate p<0.05. P-values that are **both bolded and underlined** indicate statistical significance with Bonferroni correction for multiple testing.

Bonferroni correction for multiple testing p value threshold = 0.003.

Supplementary Table 3b: Odds ratios (95 % confidence intervals) of HCC risk with groupings of bile acids (BA). Values are per doubling of relative proportions of plasma BA as a percentage of the total BA pool, in the crude model, multivariate adjusted model, and multivariate adjusted models with further adjustment for fatty liver index or hepatitis B/C infection status.

Groupings and Ratios of Bile Acids Expressed as Relative Proportions (% of total)	**Median Values (%)**		**Crude Model [1] (Matching Factors)**		**Multivariable Adjusted Model [2]**		**Multivariable Adjusted Model [3] + Fatty Liver Index**		**Multivariable Adjusted Model [4] + Hepatitis B/C Status**	
	Case	**Control**	**OR (95% CI)**	***p***	**OR (95% CI)**	***p***	**OR (95% CI)**	***p***	**OR (95% CI)**	***p***
Bile Acid Groupings										
Unconjugated BA (CA , CDCA , DCA , UDCA)	17.2	29.9	0.61 (0.52, 0.73)	***2.45E-08***	0.54 (0.43, 0.68)	***3.41E-07***	0.58 (0.45, 0.74)	***1.26E-05***	0.63 (0.49, 0.82)	***4.56E-04***
Primary unconjugated BA (CA, CDCA)	6.7	12.4	0.75 (0.66, 0.84)	***1.84E-06***	0.73 (0.62, 0.85)	***6.96E-05***	0.77 (0.66, 0.91)	***1.52E-03***	0.80 (0.67, 0.95)	*1.30E-02*
Secondary unconjugated BA (DCA, UDCA)	8.1	12.7	0.70 (0.61, 0.80)	***1.67E-07***	0.62 (0.52, 0.74)	***1.14E-07***	0.64 (0.53, 0.77)	***2.81E-06***	0.68 (0.55, 0.83)	***1.47E-04***
Conjugated BA (GCA , GDCA , GDCA , GHCA , GLCA , GUDCA, TCA , TCDCA , TDCA , THCA , TUDCA , TaMCA)	82.8	70.1	2.05 (1.41, 2.97)	***1.69E-04***	2.12 (1.34, 3.34)	***1.33E-03***	1.91 (1.17, 3.13)	*9.84E-03*	1.62 (0.98, 2.68)	*5.93E-02*
Primary conjugated (GCA, GCDCA, TCA, TCDCA)	59.0	48.7	1.89 (1.39, 2.58)	***5.69E-05***	2.09 (1.41, 3.09)	***2.28E-04***	1.91 (1.25, 2.91)	***2.67E-03***	1.67 (1.09, 2.58)	*1.97E-02*
Secondary conjugated (GDCA, GUDCA, TDCA, TUDCA)	14.1	14.7	0.88 (0.74, 1.06)	*1.86E-01*	0.76 (0.60, 0.96)	*2.36E-02*	0.76 (0.59, 0.99)	*4.06E-02*	0.84 (0.64, 1.11)	*2.19E-01*
Taurine-conjugated (TCA , TCDCA , TDCA , THCA , TUDCA , TaMCA)	12.0	6.3	1.86 (1.54, 2.25)	***2.00E-10***	1.97 (1.55, 2.51)	***3.49E-08***	1.84 (1.44, 2.36)	***1.58E-06***	1.73 (1.33, 2.26)	***4.94E-05***
Glycine-conjugated (GCA , GCDCA , GDCA , GHCA , GLCA , GUDCA)	62.9	61.9	1.23 (0.88, 1.72)	*2.25E-01*	1.25 (0.82, 1.89)	*2.98E-01*	1.14 (0.73, 1.77)	*5.72E-01*	1.04 (0.64, 1.68)	*8.86E-01*

Hydrophylic (CA, UDCA)	3.5	5.7	0.77 (0.69, 0.87)	***2.83E-05***	0.78 (0.67, 0.91)	***1.24E-03***	0.82 (0.70, 0.96)	1.51E-02	0.84 (0.71, 1.00)	5.04E-02
Hydrophobic (DCA)	7.0	11.1	0.71 (0.63, 0.81)	***4.45E-08***	0.66 (0.57, 0.77)	***2.01E-07***	0.67 (0.57, 0.79)	**2.29E-06**	0.71 (0.60, 0.84)	**1.01E-04**
Glycine-conjugated hydrophylic (GCA, GUDCA)	15.3	11.6	2.02 (1.52, 2.67)	***9.45E-07***	2.16 (1.53, 3.05)	***1.05E-05***	2.14 (1.46, 3.13)	**9.87E-05**	1.92 (1.30, 2.84)	**1.11E-03**
Glycine-conjugated hydrophobic (GDCA, GLCA)	7.8	10.1	0.80 (0.70, 0.92)	***2.15E-03***	0.74 (0.61, 0.89)	***1.12E-03***	0.73 (0.60, 0.89)	**1.86E-03**	0.76 (0.62, 0.94)	1.10E-02
Taurine-conjugated hydrophylic (TCA, TUDCA)	2.6	1.1	1.84 (1.55, 2.19)	***3.14E-12***	1.93 (1.56, 2.39)	***1.10E-09***	1.85 (1.48, 2.31)	**5.74E-08**	1.76 (1.40, 2.23)	**2.04E-06**
Taurine-conjugated hydrophobic (TDCA)	1.4	1.2	1.09 (0.97, 1.23)	*1.38E-01*	1.05 (0.89, 1.22)	*5.88E-01*	1.02 (0.86, 1.20)	8.46E-01	1.04 (0.87, 1.25)	6.73E-01
Total CA (CA , GCA , TCA)	19.9	15.4	1.82 (1.41, 2.35)	***5.17E-06***	2.06 (1.52, 2.78)	***2.77E-06***	1.98 (1.44, 2.73)	**3.00E-05**	1.92 (1.35, 2.73)	**2.83E-04**
Total CDCA (CDCA , GCDCA , TCDCA)	50.3	49.9	1.05 (0.67, 1.65)	*8.20E-01*	1.15 (0.66, 2.00)	*6.23E-01*	1.24 (0.69, 2.23)	4.66E-01	1.29 (0.68, 2.43)	4.36E-01
Total DCA (DCA , GDCA , TDCA)	18.1	22.6	0.73 (0.62, 0.84)	***3.28E-05***	0.65 (0.53, 0.80)	***2.69E-05***	0.66 (0.53, 0.81)	**1.06E-04**	0.71 (0.57, 0.88)	**1.96E-03**
Total HCA (GHCA , HCA)	0.7	1.0	0.68 (0.56, 0.83)	***1.79E-04***	0.86 (0.68, 1.09)	*2.18E-01*	0.94 (0.72, 1.22)	6.34E-01	0.80 (0.60, 1.07)	1.32E-01
Total UDCA (GUDCA ,TUDCA , UDCA)	3.8	4.7	0.80 (0.68, 0.95)	*8.83E-03*	0.80 (0.66, 0.98)	*3.32E-02*	0.84 (0.67, 1.04)	1.01E-01	0.92 (0.72, 1.17)	4.88E-01

[1] Odds Ratios (OR) and 95% confidence intervals (95% CI) conditioned on the matching factors.

[2] OR (95% CI) calculated with multivariable adjusted conditional regression models (adjustment factors: matching factors + body mass index, waist circumference, alcohol intake at recruitment, physical activity, smoking status, alcohol intake pattern and attained education level).

[3] OR (95% CI) calculated with multivariable adjusted conditional regression models as in [2] above plus additional adjustment for FLI, please see Table 1 and Supplementary Methods for additional details.

[4] OR (95% CI) calculated with multivariable adjusted conditional regression models as in [2] above plus additional adjustment for hepatitis B/C positivity, please see Table 1 for additional details.

OR represent doubling of relative percent contribution. BA proportions are calculated as relative percent of the sum of all plasma BAs. Groupings are calculated as a sum of individual BA listed in brackets. Numbers in red indicate p<0.05. P-values that are **both bolded and underlined** indicate statistical significance with Bonferroni correction for multiple testing. Bonferroni correction for multiple testing p value threshold = 0.003.

Supplementary Table 4: Baseline characteristics of the subset of case-control pairs with "suspected" non-alcoholic fatty liver disease (NAFLD).

Subject characteristics	HCC cases					Controls					*p*
	n	**%**				**n**	**%**				
Male	16	59.3				16	59.3				
Female	11	40.1				11	40.1				
Smoking status											
Never	9	33.3				12	44.4				*0.67*
Former	11	40.7				10	37.0				
Smoker	6	22.2				5	18.5				
Unknown	1	3.7				0	0.0				
Alcohol drinking pattern											
Never drinker	0	0.0				1	3.7				*0.11*
Former drinker	6	22.2				1	3.7				
Drinker at recruitment	7	25.9				5	18.5				
Lifetime drinker	14	51.9				20	74.1				
Highest school level											
Primary school completed	15	55.6				12	44.4				*0.41*
Technical/professional school	5	18.5				11	40.7				
Secondary school	4	14.8				1	3.7				
Longer education (incl. University deg.)	2	7.4				2	7.4				
Not specified	1	3.7				1	3.7				
Subjects with MetS (n, %)	23	85.2				16	59.3				*0.03*
Baseline characteristics	**n**	**Mean**	**Median**	**Min.**	**Max.**	**n**	**Mean**	**Median**	**Min.**	**Max.**	
Age (years)	27	61.9	61.5	50.0	77.2	27	62.0	61.2	50.1	77.4	*0.07*
Physical activity (Mets)	27	83.2	72.5	14.3	198.6	27	96.6	94.9	0.0	179.3	*0.26*
Body mass index (BMI, kg/m2)	27	31.2	30.0	23.0	41.8	27	28.2	28.3	23.8	38.3	*0.03*
Waist circumference (cm)	27	102.4	105.0	73.0	124.0	27	96.2	98.0	78.4	110.0	*0.03*
Alcohol intake at recruitment (g/d)	27	4.5	1.8	0.0	16.8	27	8.7	5.4	0.0	29.5	*0.03*
Biomarkers at recruitment											

Total bile acid levels (nM)	27	23737.5	5715.9	1033.0	217431.2	27	3084.83	2293.7	786.1	10398.6	*<0.0001*
Gamma-glutamyl transferase (GGT, U/L)	27	139.1	74.9	3.0	783.1	27	42.3	27.0	3.0	289.6	*0.008*
Aspartate aminotransferese (AST, U/L)	27	42.4	29.0	13.2	194.9	27	23.6	19.0	13.0	59.9	*0.02*
Alanine aminotransferase (ALT, U/L)	27	36.4	21.8	4.1	187.8	27	22.8	20.0	3.3	72.1	*0.08*
AST/ALT ratio	27	2.2	1.1	0.6	25.3	27	1.3	1.0	0.6	5.6	*0.23*
Fatty liver index (FLI)	27	74.2	79.2	25.1	99.2	27	56.6	62.5	10.3	84.3	*0.001*
Triglicerides corrected for fasting status (mmol/L)	27	1.9	1.9	0.8	4.7	27	2.0	1.9	0.4	6.0	*0.63*
High-density lipoprotein (mmol/L)	27	1.3	1.3	0.6	2.2	27	1.5	1.4	1.0	2.3	*0.04*

Suspected NAFLD defined as: at least one of the following: FLI>60, presence of the metabolic syndrome, GGT >64 in men or GGT >36 in women, or ALT>55. Baseline subject characteristics were compared between cases and matched controls for continuous variables using paired sample t-test or Wilcoxon test and Fisher's exact test for categorical variables.

Supplementary Table 5. Odds ratios (95 % confidence intervals) of individual BA concentrations and relative proportions in association with HCC risk in a sub-group of case-control pairs with "suspected" NAFLD.

Individual Bile Acids, continuous linear models		**Plasma concentrations (nM)**	**Relative proportions of the total BA pool (% total BA)[3]**
		OR (95% CI) per doubling of concentration	**OR (95%CI)** per doubling of % proportion
Unconjugated primary bile acids			
Cholic acid (CA)	Crude model [1]	1.09 (0.85, 1.39)	0.78 (0.61, 1.02)
	Multivariable adjusted model [2]	1.06 (0.79, 1.41)	0.83 (0.61, 1.12)
Chenodeoxycholic acid (CDCA)	Crude model [1]	1.09 (0.83, 1.43)	0.72 (0.52, 1.01)
	Multivariable adjusted model [2]	1.13 (0.81, 1.57)	0.83 (0.59, 1.17)
Hyocholic acid (HCA)	Crude model [1]	1.42 (0.89, 2.25)	0.71 (0.50, 0.99)
	Multivariable adjusted model [2]	1.53 (0.85, 2.77)	0.80 (0.53, 1.21)
Unconjugated secondary/tertiary bile acids			
Deoxycholic acid (DCA)	Crude model [1]	0.82 (0.57, 1.17)	0.60 (0.40, 0.90)
	Multivariable adjusted model [2]	1.09 (0.68, 1.74)	0.71 (0.47, 1.09)
Ursodeoxycholic acid (UDCA)	Crude model [1]	1.51 (0.99, 2.29)	0.74 (0.51, 1.06)
	Multivariable adjusted model [2]	1.58 (0.97, 2.57)	0.91 (0.59, 1.42)
Glycine-conjugated bile acids			
Glycocholic acid (GCA)	Crude model [1]	1.76 (1.14, 2.72)	1.80 (1.02, 3.17)
	Multivariable adjusted model [2]	1.50 (0.94, 2.38)	1.01 (0.48, 2.12)
Glycochenodeoxycholic acid (GCDCA)	Crude model [1]	2.26 (1.21, 4.24)	0.80 (0.41, 1.54)
	Multivariable adjusted model [2]	1.76 (0.93, 3.32)	0.34 (0.09, 1.27)

Glycodeoxycholic acid (GDCA)	Crude model [1]	1.65 (1.02, 2.66)	0.46 (0.23, 0.92)
	Multivariable adjusted model [2]	1.57 (0.90, 2.77)	0.60 (0.29, 1.25)
Glycohyocholic acid (GHCA)	Crude model [1]	2.53 (1.21, 5.31)	0.58 (0.30, 1.13)
	Multivariable adjusted model [2]	1.80 (0.87, 3.73)	0.47 (0.17, 1.35)
Glycolithocholic acid (GLCA)	Crude model [1]	1.81 (1.04, 3.16)	0.72 (0.49, 1.05)
	Multivariable adjusted model [2]	1.67 (0.92, 3.01)	0.94 (0.62, 1.43)
Glycoursodeoxycholic acid (GUDCA)	Crude model [1]	1.91 (1.15, 3.17)	0.97 (0.62, 1.52)
	Multivariable adjusted model [2]	1.50 (0.90, 2.52)	0.91 (0.53, 1.57)
Taurine-conjugated bile acids			
Tauro-alfa-muricholic acid (TaMCA)	Crude model [1]	2.82 (1.10, 7.20)	0.90 (0.58, 1.40)
	Multivariable adjusted model [2]	2.89 (0.79, 10.60)	0.97 (0.56, 1.71)
Taurocholic acid (TCA)	Crude model [1]	1.55 (1.12, 2.14)	1.63 (1.09, 2.43)
	Multivariable adjusted model [2]	1.37 (0.97, 1.94)	1.28 (0.84, 1.96)
Taurochenodeoxycholic acid (TCDCA)	Crude model [1]	1.61 (1.13, 2.28)	1.48 (0.96, 2.28)
	Multivariable adjusted model [2]	1.36 (0.95, 1.96)	1.16 (0.68, 1.97)
Taurodeoxycholic acid (TDCA)	Crude model [1]	1.69 (1.10, 2.60)	0.89 (0.57, 1.40)
	Multivariable adjusted model [2]	1.46 (0.93, 2.30)	1.01 (0.56, 1.82)
Tauroursodeoxycholic acid (TUDCA)	Crude model [1]	1.68 (1.12, 2.51)	1.46 (0.98, 2.18)
	Multivariable adjusted model [2]	1.38 (0.93, 2.06)	1.20 (0.82, 1.76)

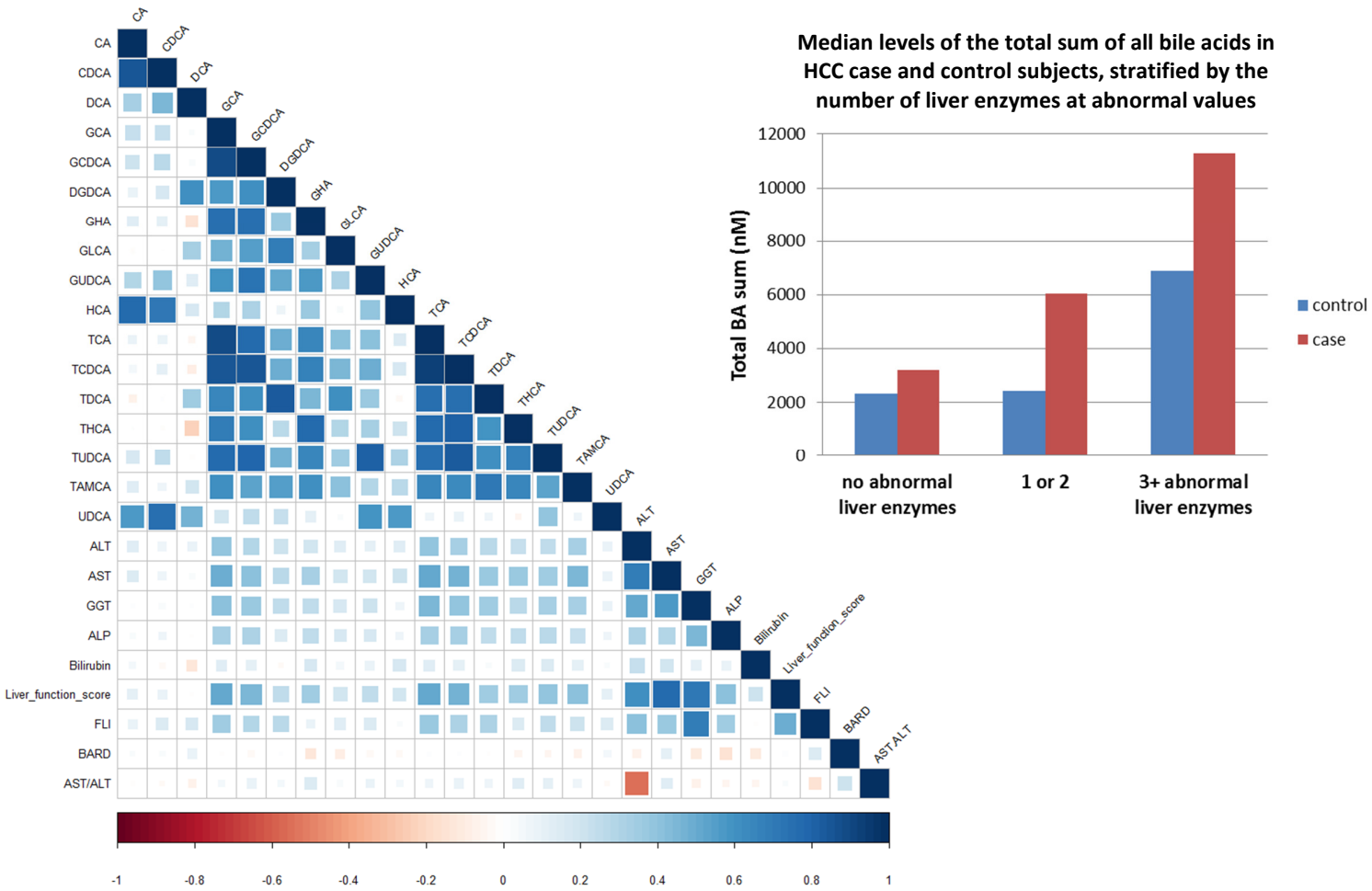
[1] Odds Ratios (OR) and 95% confidence intervals (95% CI) conditioned on the matching factors.

[2] OR (95% CI) calculated with multivariable adjusted conditional regression models (adjustment factors: matching factors + body mass index, waist circumference, alcohol intake at recruitment, physical activity, smoking status, alcohol intake pattern and attained education level).

OR (95%CI) represent the HCC risk per doubling of the BA concentration or percent contribution.

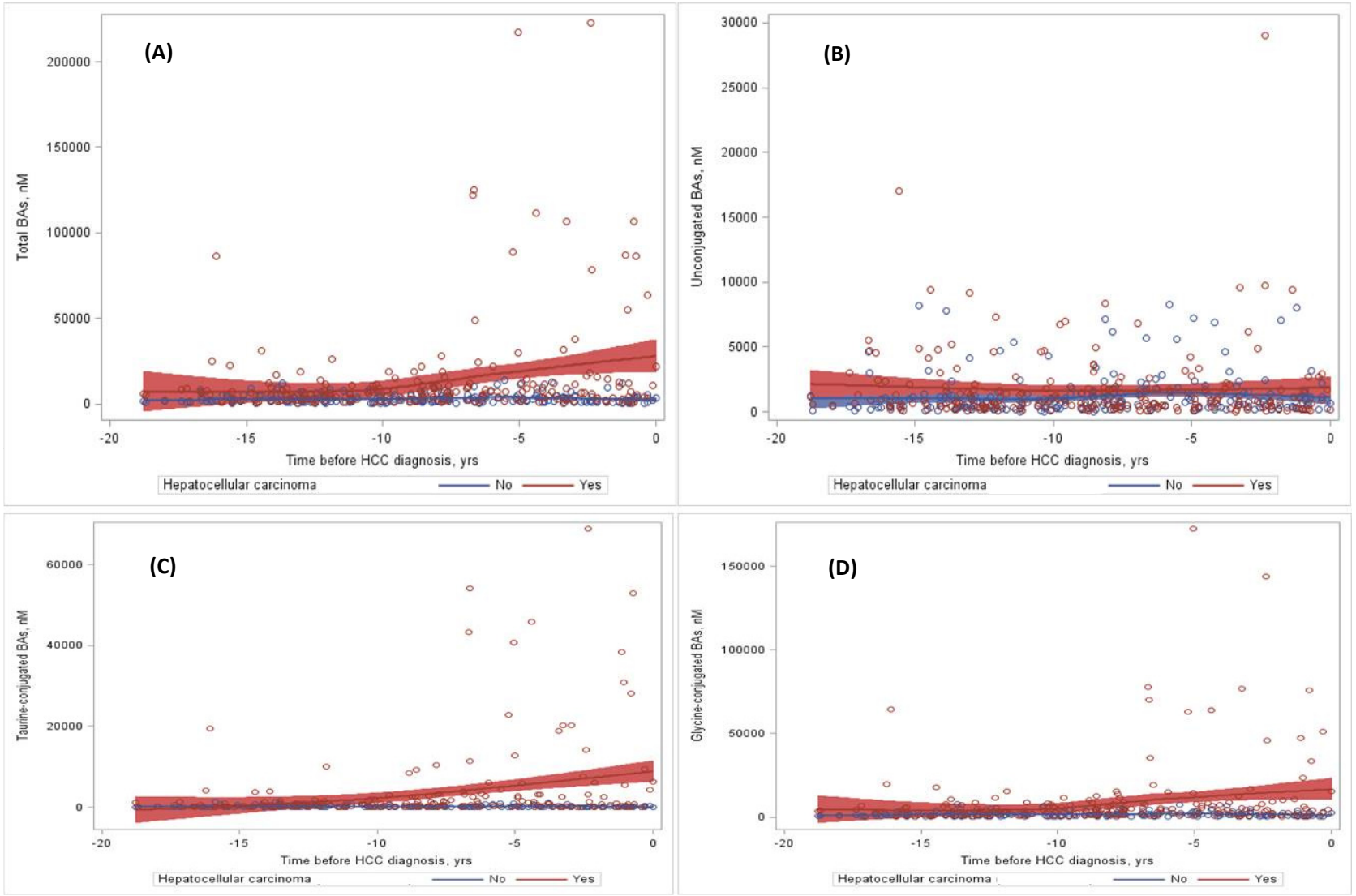
[3] Relative BA proportions are calculated as relative percent of the sum of all BA.

Supplementary Figure 1. Correlations among bile acids (BA) in control subjects and comparison of BA levels between HCC cases and matched controls by the number of abnormal liver function markers.



Correlation coefficients are based on Spearman partial correlation (adjusted for age and sex). * p-value calculated using Kruskall-Wallis non-parametric test.

Supplementary Figure 2. Differences between levels of (a) total bile acids (BA), (b) unconjugated BAs, (c) taurine-conjugated BAs and (d) glycine-conjugated BAs across years of follow up time in the EPIC cohort, from enrolment to HCC diagnosis.



The x-axis represents time period from enrolment to HCC diagnosis (Time=0) for each HCC case and its respective matched control throughout the follow-up period of the EPIC cohort. The y-axis represents concentration of **(A)** total sum of all BAs and main groupings of **(B)** unconjugated, **(C)** glycine-conjugated and (D) taurine-conjugated BAs, calculated as the sum of individual BAs in each particular grouping. Cases are denoted in red colour (i.e. HCC Yes), matched control subjects in blue (i.e. HCC No).