

# **Stepien M et al: Pre-diagnostic alterations in circulating bile acid profiles in the development of hepatocellular carcinoma**

## **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 precipitate 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 DAsource (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]:

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

where TG refers to triglycerides (mg/dL), BMI to body mass index (kg/m<sup>2</sup>). 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
<b>Unconjugated primary bile acids</b>								
Cholic acid (CA)	Tertile range		0 - 1.8	>1.8 - 7.5	>=7.5			
	Crude model <sup>1</sup>		Ref.	<b>0.52</b> <b>(0.32, 0.85)</b>	<b>0.46</b> <b>(0.28, 0.76)</b>	0.010	<b>0.79</b> <b>(0.71, 0.88)</b>	<b>&lt;0.001</b>
	Multivariable adjusted model <sup>2</sup>		Ref.	0.65 (0.36, 1.18)	<b>0.48</b> <b>(0.26, 0.91)</b>	0.040	<b>0.81</b> <b>(0.71, 0.92)</b>	<b>0.002</b>
Chenodeoxycholic acid (CDCA)	Tertile range		0 - 5.1	>5.1 - 12.0	>=12.0			
	Crude model <sup>1</sup>		Ref.	<b>0.35</b> <b>(0.22, 0.57)</b>	<b>0.42</b> <b>(0.26, 0.69)</b>	0.002	<b>0.75</b> <b>(0.67, 0.85)</b>	<b>&lt;0.001</b>
	Multivariable adjusted model <sup>2</sup>		Ref.	<b>0.32</b> <b>(0.18, 0.59)</b>	<b>0.39</b> <b>(0.21, 0.72)</b>	0.008	<b>0.72</b> <b>(0.62, 0.84)</b>	<b>&lt;0.001</b>
Hyochoolic acid (HCA)	Tertile range		0 - 0.2	>0.2- 0.4	>=0.4			
	Crude model <sup>1</sup>		Ref.	<b>0.57</b> <b>(0.37, 0.88)</b>	<b>0.35</b> <b>(0.21, 0.57)</b>	<b>&lt;0.001</b>	<b>0.70</b> <b>(0.61, 0.81)</b>	<b>&lt;0.001</b>
	Multivariable adjusted model <sup>2</sup>		Ref.	0.58 (0.34, 1.00)	<b>0.47</b> <b>(0.26, 0.86)</b>	0.030	<b>0.75</b> <b>(0.64, 0.89)</b>	<b>0.001</b>
<b>Unconjugated secondary/tertiary bile acids</b>								
Deoxycholic acid (DCA)	Tertile range		0 - 7.4	>7.4 - 14.8	>=14.8			
	Crude model <sup>1</sup>		Ref.	<b>0.52</b> <b>(0.33, 0.82)</b>	<b>0.33</b> <b>(0.20, 0.55)</b>	<b>&lt;0.001</b>	<b>0.71</b> <b>(0.63, 0.81)</b>	<b>&lt;0.001</b>
	Multivariable adjusted model <sup>2</sup>		Ref.	<b>0.48</b> <b>(0.27, 0.85)</b>	<b>0.28</b> <b>(0.15, 0.54)</b>	<b>&lt;0.001</b>	<b>0.66</b> <b>(0.57, 0.77)</b>	<b>&lt;0.001</b>
Ursodeoxycholic acid (UDCA)	Tertile range		0 - 0.6	>0.6 - 1.6	>=1.6			
	Crude model <sup>1</sup>		Ref.	<b>0.55</b> <b>(0.35, 0.87)</b>	<b>0.45</b> <b>(0.27, 0.74)</b>	0.003	<b>0.81</b> <b>(0.73, 0.91)</b>	<b>&lt;0.001</b>
	Multivariable adjusted model <sup>2</sup>		Ref.	<b>0.50</b> <b>(0.29, 0.87)</b>	<b>0.49</b> <b>(0.26, 0.92)</b>	0.040	<b>0.78</b> <b>(0.67, 0.90)</b>	<b>0.001</b>

Individual Bile Acids Expressed as Relative Proportions (% of total)			Tertile 1	Tertile 2	Tertile 3	p-trend	Continuous (per doubling of %)	p-value
<b>Glycine-conjugated bile acids</b>								
Glycocholic acid (GCA)	Tertile range		0.6 -5.6	>5.6 - 9.0	>=9.0			
	Crude model <sup>1</sup>		Ref.	0.98 (0.56, 1.73)	<b><u>3.35</u></b> <b><u>(1.99, 5.65)</u></b>	<b><u>&lt;0.001</u></b>	<b>1.96</b> <b>(1.54, 2.50)</b>	<b><u>&lt;0.001</u></b>
	Multivariable adjusted model <sup>2</sup>		Ref.	0.90 (0.43, 1.87)	<b><u>3.91</u></b> <b><u>(1.99, 7.70)</u></b>	<b><u>&lt;0.001</u></b>	<b>2.13</b> <b>(1.58, 2.86)</b>	<b><u>&lt;0.001</u></b>
Glycochenodeoxycholic acid (GCDCA)	Tertile range		2.6 - 26.9	>26.9 - 40.3	>40.3			
	Crude model <sup>1</sup>		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 <sup>2</sup>		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 <sup>1</sup>		Ref.	<b>0.59</b> <b>(0.39, 0.90)</b>	<b>0.54</b> <b>(0.34, 0.85)</b>	0.007	<b>0.83</b> <b>(0.73, 0.95)</b>	<b>0.007</b>
	Multivariable adjusted model <sup>2</sup>		Ref.	<b>0.43</b> <b>(0.24, 0.76)</b>	<b>0.44</b> <b>(0.24, 0.78)</b>	0.005	<b>0.76</b> <b>(0.64, 0.91)</b>	<b><u>0.002</u></b>
Glycohyocholic acid (GHCA)	Tertile range		0 - 0.3	>0.3- 0.6	>=0.6			
	Crude model <sup>1</sup>		Ref.	<b>0.50</b> <b>(0.31, 0.79)</b>	<b><u>0.44</u></b> <b><u>(0.27, 0.70)</u></b>	<b><u>0.001</u></b>	<b>0.74</b> <b>(0.62, 0.88)</b>	<b><u>&lt;0.001</u></b>
	Multivariable adjusted model <sup>2</sup>		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 <sup>1</sup>		Ref.	<b>0.62</b> <b>(0.40, 0.95)</b>	<b><u>0.33</u></b> <b><u>(0.20, 0.54)</u></b>	<b><u>&lt;0.001</u></b>	<b>0.73</b> <b>(0.64, 0.83)</b>	<b><u>&lt;0.001</u></b>
	Multivariable adjusted model <sup>2</sup>		Ref.	<b>0.56</b> <b>(0.33, 0.95)</b>	<b><u>0.33</u></b> <b><u>(0.18, 0.61)</u></b>	<b><u>0.001</u></b>	<b>0.73</b> <b>(0.62, 0.86)</b>	<b><u>&lt;0.001</u></b>
Glycoursodeoxycholic acid (GUDCA)	Tertile range		0.1 - 2.3	>2.3 - 3.9	>=3.9			
	Crude model <sup>1</sup>		Ref.	<b>0.51</b> <b>(0.32, 0.81)</b>	<b>0.58</b> <b>(0.36, 0.94)</b>	0.060	<b>0.80</b> <b>(0.68, 0.94)</b>	<b><u>0.006</u></b>

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 <sup>2</sup>		Ref.	0.63 (0.35, 1.14)	0.64 (0.35, 1.15)	0.190	<b>0.80</b> <b>(0.66, 0.98)</b>	<b>0.030</b>
<b>Taurine-conjugated bile acids</b>							
Tauro-alfa-muricholic acid (TaMCA)	Tertile range	0 - 0.1	>0.1 - 0.2	>=0.2			
	Crude model <sup>1</sup>	Ref.	0.79 (0.49, 1.28)	<b>1.93</b> <b>(1.20, 3.11)</b>	<b><u>0.001</u></b>	<b>1.31</b> <b>(1.15, 1.48)</b>	<b><u>&lt;0.001</u></b>
	Multivariable adjusted model <sup>2</sup>	Ref.	1.00 (0.55, 1.81)	<b>2.56</b> <b>(1.40, 4.68)</b>	<b><u>0.001</u></b>	<b>1.47</b> <b>(1.25, 1.74)</b>	<b><u>&lt;0.001</u></b>
Taurocholic acid (TCA)	Tertile range	0 - 0.5	>0.5 - 1.2	>=1.2			
	Crude model <sup>1</sup>	Ref.	1.62 (0.90, 2.90)	<b>6.56</b> <b>(3.58, 12.01)</b>	<b><u>&lt;0.001</u></b>	1.73 (1.48, 2.03)	<b><u>&lt;0.001</u></b>
	Multivariable adjusted model <sup>2</sup>	Ref.	1.70 (0.81, 3.57)	<b>9.79</b> <b>(4.48, 21.39)</b>	<b><u>&lt;0.001</u></b>	1.83 (1.50, 2.22)	<b><u>&lt;0.001</u></b>
Taurochenodeoxycholic acid (TCDCA)	Tertile range	0 - 2.0	>2 - 4.2	>=4.2			
	Crude model <sup>1</sup>	Ref.	1.39 (0.79, 2.42)	<b>3.16</b> <b>(1.90, 5.25)</b>	<b><u>&lt;0.001</u></b>	1.65 (1.40, 1.95)	<b><u>&lt;0.001</u></b>
	Multivariable adjusted model <sup>2</sup>	Ref.	1.50 (0.76, 2.98)	<b>3.76</b> <b>(1.92, 7.37)</b>	<b><u>&lt;0.001</u></b>	1.74 (1.40, 2.15)	<b><u>&lt;0.001</u></b>
Taurodeoxycholic acid (TDCA)	Tertile range	0 - 0.7	>0.7 - 1.6	>=1.6			
	Crude model <sup>1</sup>	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 <sup>2</sup>	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 <sup>1</sup>	Ref.	0.94 (0.61, 1.47)	<b>1.61</b> <b>(1.03, 2.51)</b>	0.020	<b>1.35</b> <b>(1.15, 1.59)</b>	<b><u>&lt;0.001</u></b>
	Multivariable adjusted model <sup>2</sup>	Ref.	1.03 (0.60, 1.79)	<b>1.83</b> <b>(1.03, 3.26)</b>	0.020	<b>1.39</b> <b>(1.13, 1.72)</b>	<b><u>0.002</u></b>

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.  $0.05/34$ ).

<sup>1</sup> OR (95% CI) conditioned on the matching factors.

<sup>2</sup> 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)	<b>3.61E-02</b>	1.13 (1.00, 1.26)	<b>4.20E-02</b>	1.13 (0.96, 1.34)	<b>1.51E-01</b>
	Chenodeoxycholic acid (CDCA)	1.09 (0.96, 1.23)	<b>1.73E-01</b>	1.09 (0.96, 1.23)	<b>1.87E-01</b>	1.13 (0.94, 1.36)	<b>1.83E-01</b>
	Hyocholic acid (HCA)	1.32 (1.13, 1.54)	<b>6.23E-04</b>	1.26 (1.07, 1.47)	<b>4.95E-03</b>	1.17 (0.93, 1.48)	<b>1.90E-01</b>
Unconjugated secondary / tertiary bile acids	Deoxycholic acid (DCA)	0.94 (0.81, 1.08)	<b>3.57E-01</b>	0.98 (0.86, 1.13)	<b>7.94E-01</b>	1.03 (0.83, 1.29)	<b>7.77E-01</b>
	Ursodeoxycholic acid (UDCA)	1.18 (1.03, 1.36)	<b>1.87E-02</b>	1.24 (1.07, 1.43)	<b>3.96E-03</b>	2.23 (1.53, 3.25)	<b>2.77E-05</b>
Glycine-conjugated bile acids	Glycocholic acid (GCA)	2.08 (1.63, 2.67)	<b>6.90E-09</b>	2.18 (1.64, 2.88)	<b>6.33E-08</b>	1.83 (1.35, 2.46)	<b>8.12E-05</b>
	Glycochenodeoxycholic acid (GCDCA)	2.09 (1.59, 2.74)	<b>9.63E-08</b>	1.53 (1.26, 1.85)	<b>1.81E-05</b>	1.88 (1.28, 2.74)	<b>1.16E-03</b>
	Glycodeoxycholic acid (GDCA)	1.30 (1.11, 1.54)	<b>1.46E-03</b>	2.22 (1.70, 2.89)	<b>4.30E-09</b>	2.74 (1.68, 4.46)	<b>5.42E-05</b>
	Glycohyocholic acid (GHCA)	1.92 (1.49, 2.49)	<b>6.73E-07</b>	1.32 (1.13, 1.54)	<b>6.08E-04</b>	2.32 (1.49, 3.61)	<b>1.85E-04</b>
	Glycolithocholic acid (GLCA)	1.18 (1.01, 1.38)	<b>3.38E-02</b>	1.91 (1.47, 2.48)	<b>1.51E-06</b>	1.48 (1.14, 1.92)	<b>3.30E-03</b>
	Glycoursodeoxycholic acid (GUDCA)	1.52 (1.25, 1.83)	<b>2.02E-05</b>	1.18 (1.01, 1.37)	<b>3.53E-02</b>	1.99 (1.32, 3.00)	<b>1.06E-03</b>

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

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
Unconjugated primary bile acids	Cholic acid (CA)	0.85 (0.74, 0.97)	<b><i>1.64E-02</i></b>	0.84 (0.73, 0.96)	<b><i>1.31E-02</i></b>	0.85 (0.71, 1.03)	<i>1.01E-01</i>
	Chenodeoxycholic acid (CDCA)	0.76 (0.65, 0.89)	<b><i>8.60E-04</i></b>	0.76 (0.64, 0.89)	<b><i>6.19E-04</i></b>	0.79 (0.63, 1.00)	<b><i>4.57E-02</i></b>
	Hyocholic acid (HCA)	0.81 (0.67, 0.97)	<b><i>1.88E-02</i></b>	0.76 (0.64, 0.91)	<b><i>2.68E-03</i></b>	0.70 (0.53, 0.91)	<b><i>9.25E-03</i></b>
Unconjugated secondary / tertiary bile acids	Deoxycholic acid (DCA)	0.67 (0.57, 0.79)	<b><i>2.29E-06</i></b>	0.68 (0.57, 0.80)	<b><i>5.35E-06</i></b>	0.65 (0.49, 0.85)	<b><i>1.78E-03</i></b>
	Ursodeoxycholic acid (UDCA)	0.82 (0.70, 0.95)	<b><i>8.08E-03</i></b>	0.83 (0.71, 0.97)	<b><i>1.97E-02</i></b>	0.91 (0.73, 1.12)	<i>3.74E-01</i>
Glycine-conjugated bile acids	Glycocholic acid (GCA)	2.04 (1.47, 2.83)	<b><i>1.98E-05</i></b>	2.10 (1.50, 2.93)	<b><i>1.46E-05</i></b>	1.76 (1.15, 2.69)	<b><i>8.68E-03</i></b>
	Glycochenodeoxycholic acid (GCDCA)	0.96 (0.67, 1.38)	<i>8.17E-01</i>	0.89 (0.62, 1.29)	<i>5.50E-01</i>	0.78 (0.43, 1.41)	<i>4.09E-01</i>
	Glycodeoxycholic acid (GDCA)	0.75 (0.62, 0.91)	<b><i>3.59E-03</i></b>	0.78 (0.64, 0.95)	<b><i>1.26E-02</i></b>	0.82 (0.62, 1.09)	<i>1.74E-01</i>
	Glycohyocholic acid (GHCA)	0.91 (0.72, 1.15)	<i>4.31E-01</i>	0.77 (0.60, 0.98)	<b><i>3.50E-02</i></b>	0.72 (0.51, 1.03)	<i>7.05E-02</i>
	Glycolithocholic acid (GLCA)	0.70 (0.58, 0.85)	<b><i>2.60E-04</i></b>	0.70 (0.58, 0.85)	<b><i>2.67E-04</i></b>	0.68 (0.51, 0.90)	<b><i>6.41E-03</i></b>
	Glycoursodeoxycholic acid (GUDCA)	0.82 (0.67, 1.01)	<i>6.15E-02</i>	0.84 (0.68, 1.05)	<i>1.19E-01</i>	1.07 (0.78, 1.46)	<i>6.87E-01</i>
Taurine-conjugated bile acids	Tauro-alfa-muricholic acid (TaMCA)	1.45 (1.22, 1.72)	<b><i>2.66E-05</i></b>	1.45 (1.22, 1.73)	<b><i>4.05E-05</i></b>	1.21 (0.95, 1.55)	<i>1.30E-01</i>

Taurocholic acid (TCA)	1.76 (1.43, 2.16)	<b><u>8.26E-08</u></b>	1.82 (1.46, 2.27)	<b><u>7.74E-08</u></b>	1.67 (1.26, 2.20)	<b><u>3.17E-04</u></b>
Taurochenodeoxycholic acid (TCDCA)	1.64 (1.32, 2.04)	<b><u>1.11E-05</u></b>	1.69 (1.34, 2.14)	<b><u>1.22E-05</u></b>	1.58 (1.17, 2.14)	<b><u>2.65E-03</u></b>
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)	<b><u>6.06E-03</u></b>	1.42 (1.13, 1.79)	<b><u>2.89E-03</u></b>	1.91 (1.30, 2.81)	<b><u>9.53E-04</u></b>

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

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

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)	<b><u>5.19E-03</u></b>	0.97 (0.95, 0.99)	<b><u>4.58E-03</u></b>	0.97 (0.95, 1.00)	<b><u>2.41E-02</u></b>

<sup>1</sup> Odds Ratios (OR) and 95% confidence intervals (95% CI) conditioned on the matching factors.

<sup>2</sup> 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).

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

<sup>4</sup> OR (95% CI) calculated with multivariable adjusted conditional regression models as in <sup>2</sup> 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 <sup>1</sup> (Matching Factors)		Multivariable Adjusted Model <sup>2</sup>		Multivariable Adjusted Model <sup>3</sup> + Fatty Liver Index		Multivariable Adjusted Model <sup>4</sup> + Hepatitis B/C Status	
	Case	Control	OR (95% CI)	<i>p</i>	OR (95% CI)	<i>p</i>	OR (95% CI)	<i>p</i>	OR (95% CI)	<i>p</i>
<b>Bile Acid Groupings</b>										
Unconjugated BA (CA , CDCA , DCA , UDCA )	17.2	29.9	0.61 (0.52, 0.73)	<b><u>2.45E-08</u></b>	0.54 (0.43, 0.68)	<b><u>3.41E-07</u></b>	0.58 (0.45, 0.74)	<b><u>1.26E-05</u></b>	0.63 (0.49, 0.82)	<b><u>4.56E-04</u></b>
Primary unconjugated BA (CA, CDCA)	6.7	12.4	0.75 (0.66, 0.84)	<b><u>1.84E-06</u></b>	0.73 (0.62, 0.85)	<b><u>6.96E-05</u></b>	0.77 (0.66, 0.91)	<b><u>1.52E-03</u></b>	0.80 (0.67, 0.95)	<b><u>1.30E-02</u></b>
Secondary unconjugated BA (DCA, UDCA)	8.1	12.7	0.70 (0.61, 0.80)	<b><u>1.67E-07</u></b>	0.62 (0.52, 0.74)	<b><u>1.14E-07</u></b>	0.64 (0.53, 0.77)	<b><u>2.81E-06</u></b>	0.68 (0.55, 0.83)	<b><u>1.47E-04</u></b>
Conjugated BA (GCA , GDCA , GDCA , GHCA , GLCA , GUDCA, TCA , TCDCA , TDCA , THCA , TUDCA , TaMCA)	82.8	70.1	2.05 (1.41, 2.97)	<b><u>1.69E-04</u></b>	2.12 (1.34, 3.34)	<b><u>1.33E-03</u></b>	1.91 (1.17, 3.13)	<b><u>9.84E-03</u></b>	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)	<b><u>5.69E-05</u></b>	2.09 (1.41, 3.09)	<b><u>2.28E-04</u></b>	1.91 (1.25, 2.91)	<b><u>2.67E-03</u></b>	1.67 (1.09, 2.58)	<b><u>1.97E-02</u></b>
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)	<b><u>2.36E-02</u></b>	0.76 (0.59, 0.99)	<b><u>4.06E-02</u></b>	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)	<b><u>2.00E-10</u></b>	1.97 (1.55, 2.51)	<b><u>3.49E-08</u></b>	1.84 (1.44, 2.36)	<b><u>1.58E-06</u></b>	1.73 (1.33, 2.26)	<b><u>4.94E-05</u></b>
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)	<u><b>2.83E-05</b></u>	0.78 (0.67, 0.91)	<u><b>1.24E-03</b></u>	0.82 (0.70, 0.96)	<u><b>1.51E-02</b></u>	0.84 (0.71, 1.00)	5.04E-02
Hydrophobic (DCA)	7.0	11.1	0.71 (0.63, 0.81)	<u><b>4.45E-08</b></u>	0.66 (0.57, 0.77)	<u><b>2.01E-07</b></u>	0.67 (0.57, 0.79)	<u><b>2.29E-06</b></u>	0.71 (0.60, 0.84)	<u><b>1.01E-04</b></u>
Glycine-conjugated hydrophylic (GCA, GUDCA)	15.3	11.6	2.02 (1.52, 2.67)	<u><b>9.45E-07</b></u>	2.16 (1.53, 3.05)	<u><b>1.05E-05</b></u>	2.14 (1.46, 3.13)	<u><b>9.87E-05</b></u>	1.92 (1.30, 2.84)	<u><b>1.11E-03</b></u>
Glycine-conjugated hydrophobic (GDCA, GLCA)	7.8	10.1	0.80 (0.70, 0.92)	<u><b>2.15E-03</b></u>	0.74 (0.61, 0.89)	<u><b>1.12E-03</b></u>	0.73 (0.60, 0.89)	<u><b>1.86E-03</b></u>	0.76 (0.62, 0.94)	<u><b>1.10E-02</b></u>
Taurine-conjugated hydrophylic (TCA, TUDCA)	2.6	1.1	1.84 (1.55, 2.19)	<u><b>3.14E-12</b></u>	1.93 (1.56, 2.39)	<u><b>1.10E-09</b></u>	1.85 (1.48, 2.31)	<u><b>5.74E-08</b></u>	1.76 (1.40, 2.23)	<u><b>2.04E-06</b></u>
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)	<u><b>5.17E-06</b></u>	2.06 (1.52, 2.78)	<u><b>2.77E-06</b></u>	1.98 (1.44, 2.73)	<u><b>3.00E-05</b></u>	1.92 (1.35, 2.73)	<u><b>2.83E-04</b></u>
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)	<u><b>3.28E-05</b></u>	0.65 (0.53, 0.80)	<u><b>2.69E-05</b></u>	0.66 (0.53, 0.81)	<u><b>1.06E-04</b></u>	0.71 (0.57, 0.88)	<u><b>1.96E-03</b></u>
Total HCA (GHCA , HCA)	0.7	1.0	0.68 (0.56, 0.83)	<u><b>1.79E-04</b></u>	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)	<u><b>8.83E-03</b></u>	0.80 (0.66, 0.98)	<u><b>3.32E-02</b></u>	0.84 (0.67, 1.04)	1.01E-01	0.92 (0.72, 1.17)	4.88E-01

<sup>1</sup> Odds Ratios (OR) and 95% confidence intervals (95% CI) conditioned on the matching factors.

<sup>2</sup> 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).

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

<sup>4</sup> OR (95% CI) calculated with multivariable adjusted conditional regression models as in <sup>2</sup> 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).

[illegible]

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

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

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

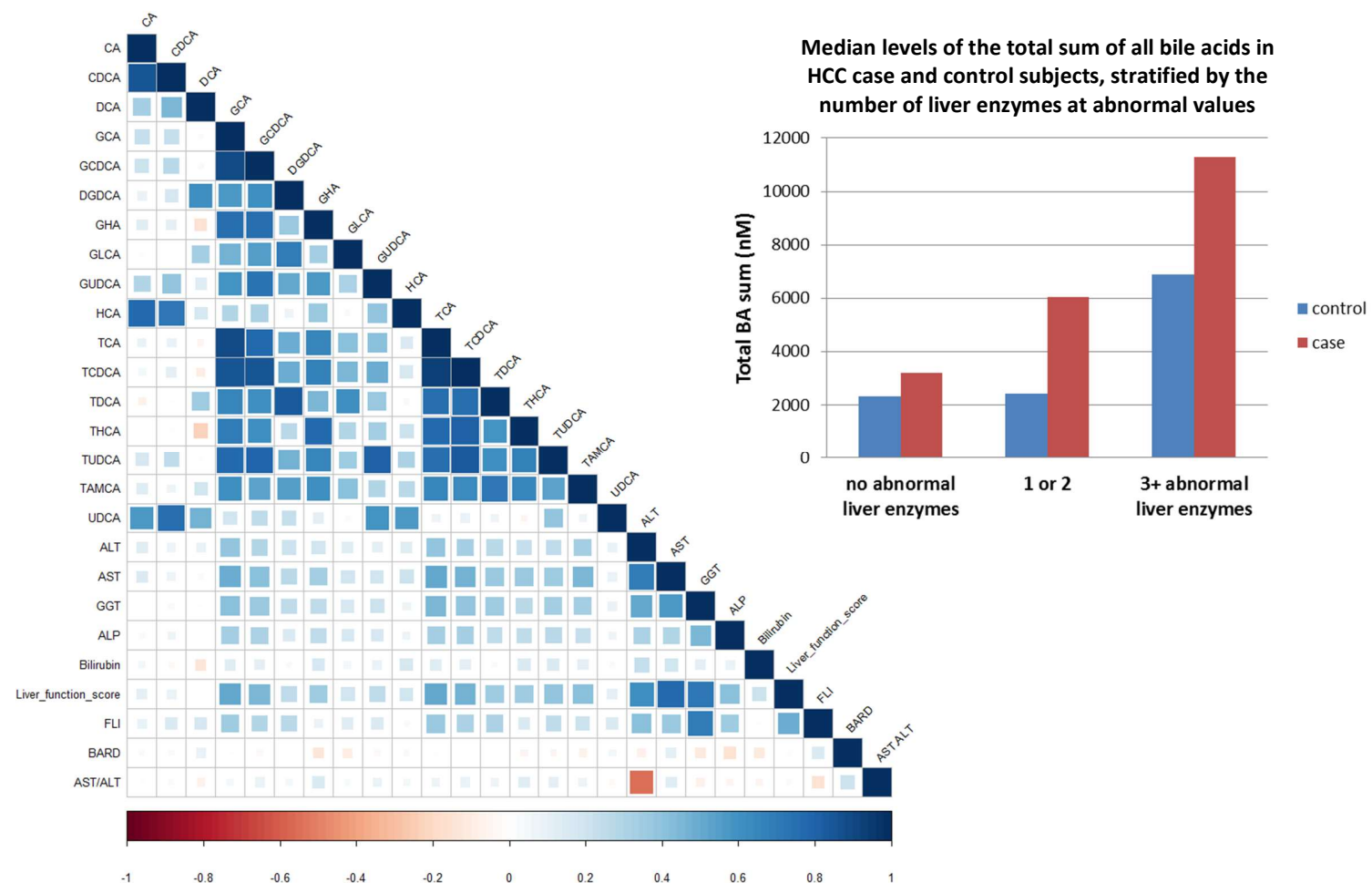
<sup>1</sup> Odds Ratios (OR) and 95% confidence intervals (95% CI) conditioned on the matching factors.

<sup>2</sup> 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.

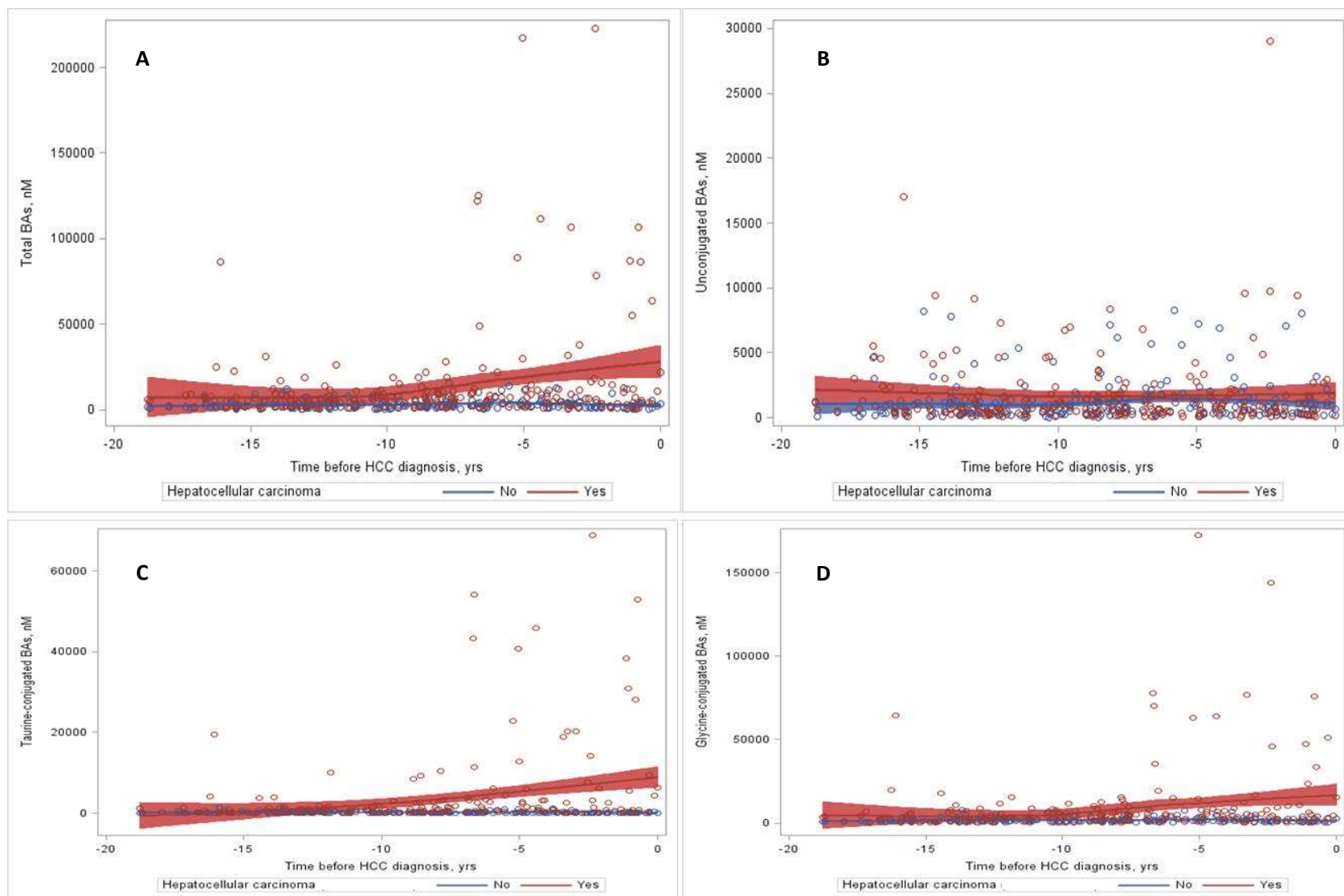
<sup>3</sup> Relative BA proportions are calculated as relative percent of the sum of all BA.

**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 Kruskal-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).