

## **Prediagnostic plasma metabolite concentrations and liver cancer risk: a population-based study of Chinese men**

Zhuo-Ying Li<sup>a,b</sup>, Qiu-Ming Shen<sup>b</sup>, Jing Wang<sup>b</sup>, Jia-Yi Tuo<sup>b,c</sup>, Yu-Ting Tan<sup>b</sup>, Hong-Lan Li<sup>b</sup>, Yong-Bing Xiang<sup>b,a,c</sup>

### **Affiliations:**

<sup>a</sup> School of Public Health, Fudan University, Shanghai 200032, China

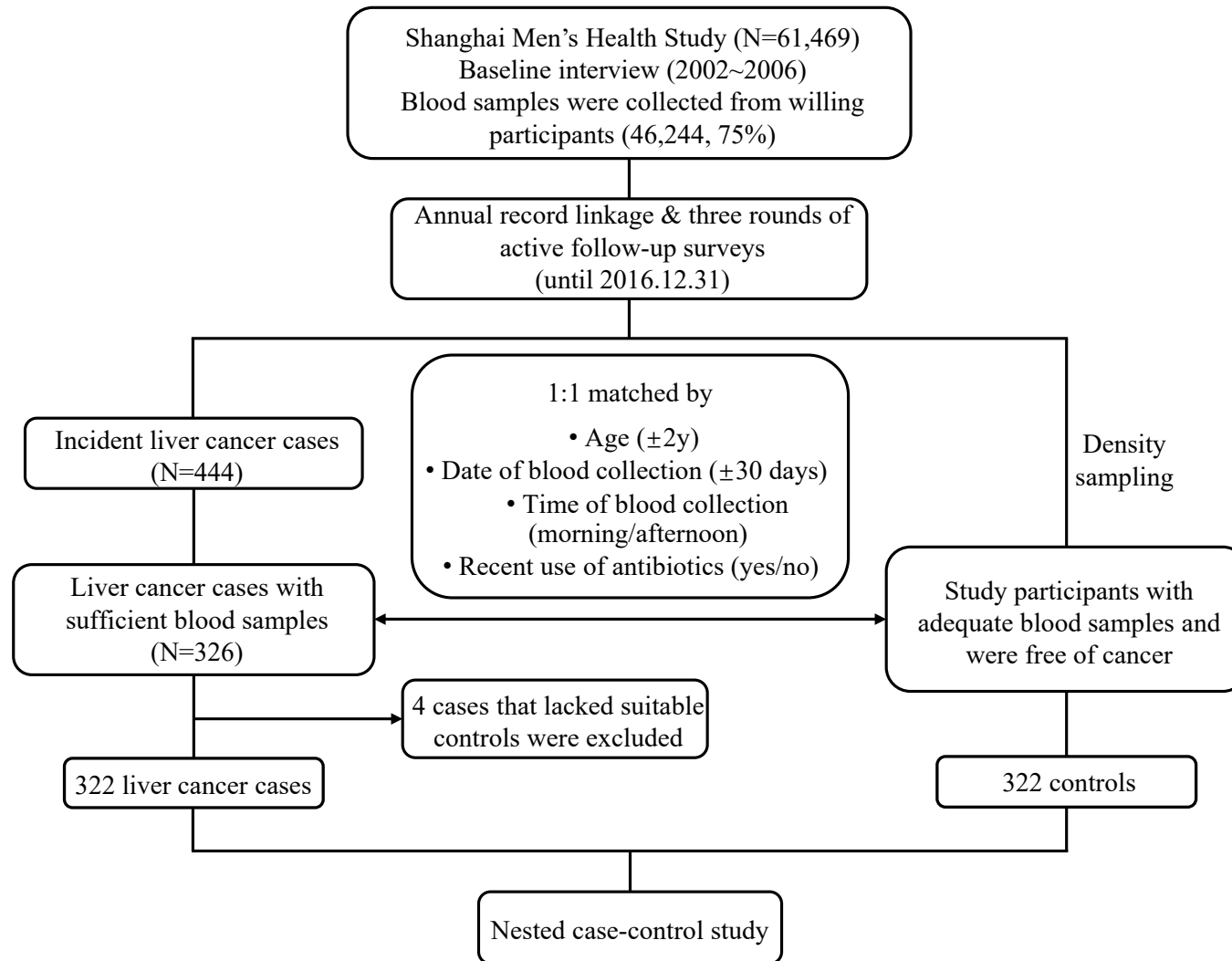
<sup>b</sup> State Key Laboratory of System Medicine for Cancer & Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200032, China

<sup>c</sup> School of Public Health, Shanghai Jiaotong University School of Medicine, Shanghai 200025, China

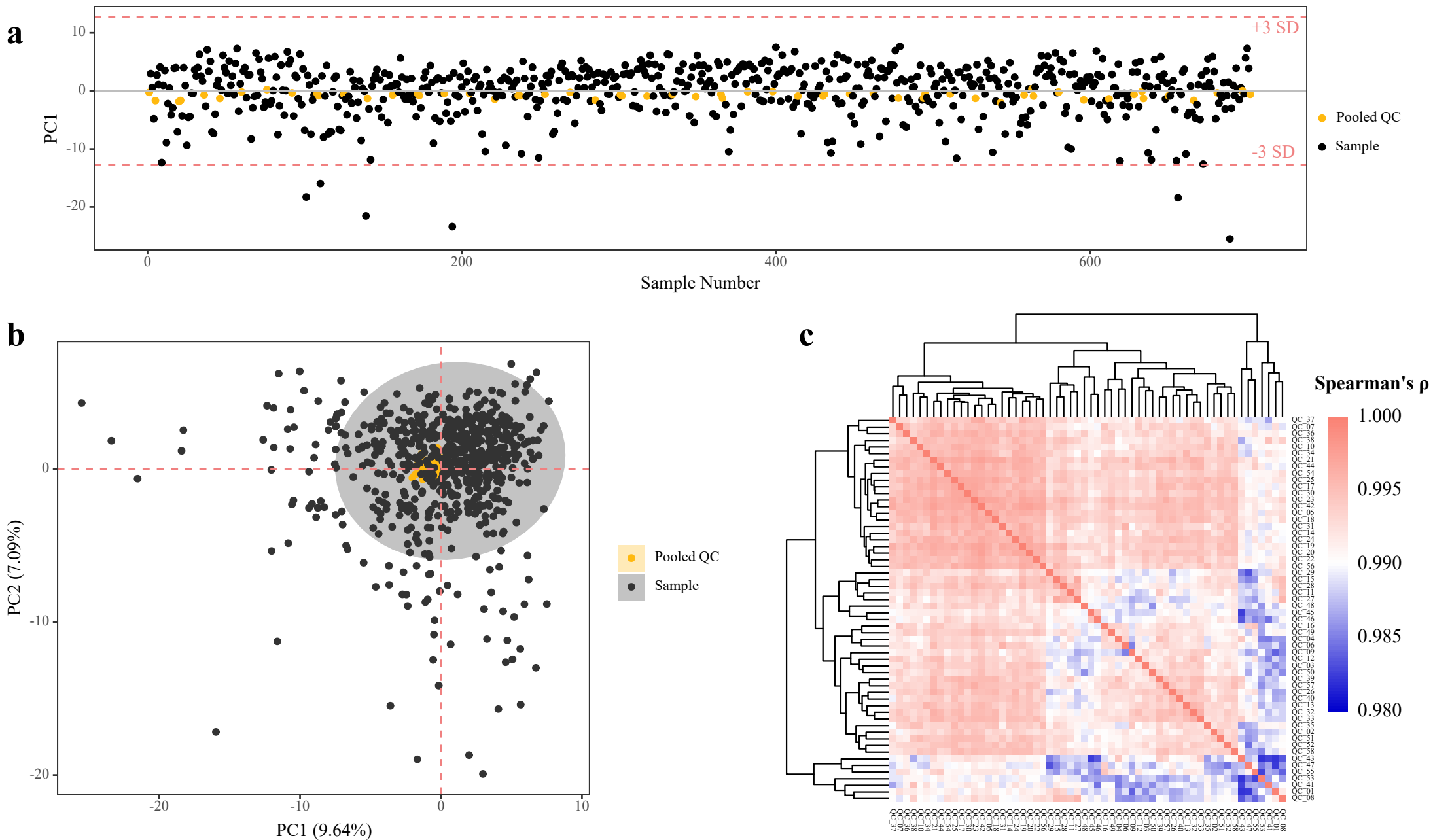
**ORCID:** Zhuo-Ying Li (0000-0003-4592-7136), Yong-Bing Xiang (0000-0002-3840-9915)

**\*Correspondence to:** Yong-Bing Xiang, State Key Laboratory of System Medicine for Cancer & Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiao Tong University School of Medicine, No. 25, Lane 2200, Xie Tu Road, Shanghai, 200032, P. R. China. E-mail: ybxiang@shsci.org.

Supplementary Figures



Supplementary Figure 1 Design of the nested case-control study

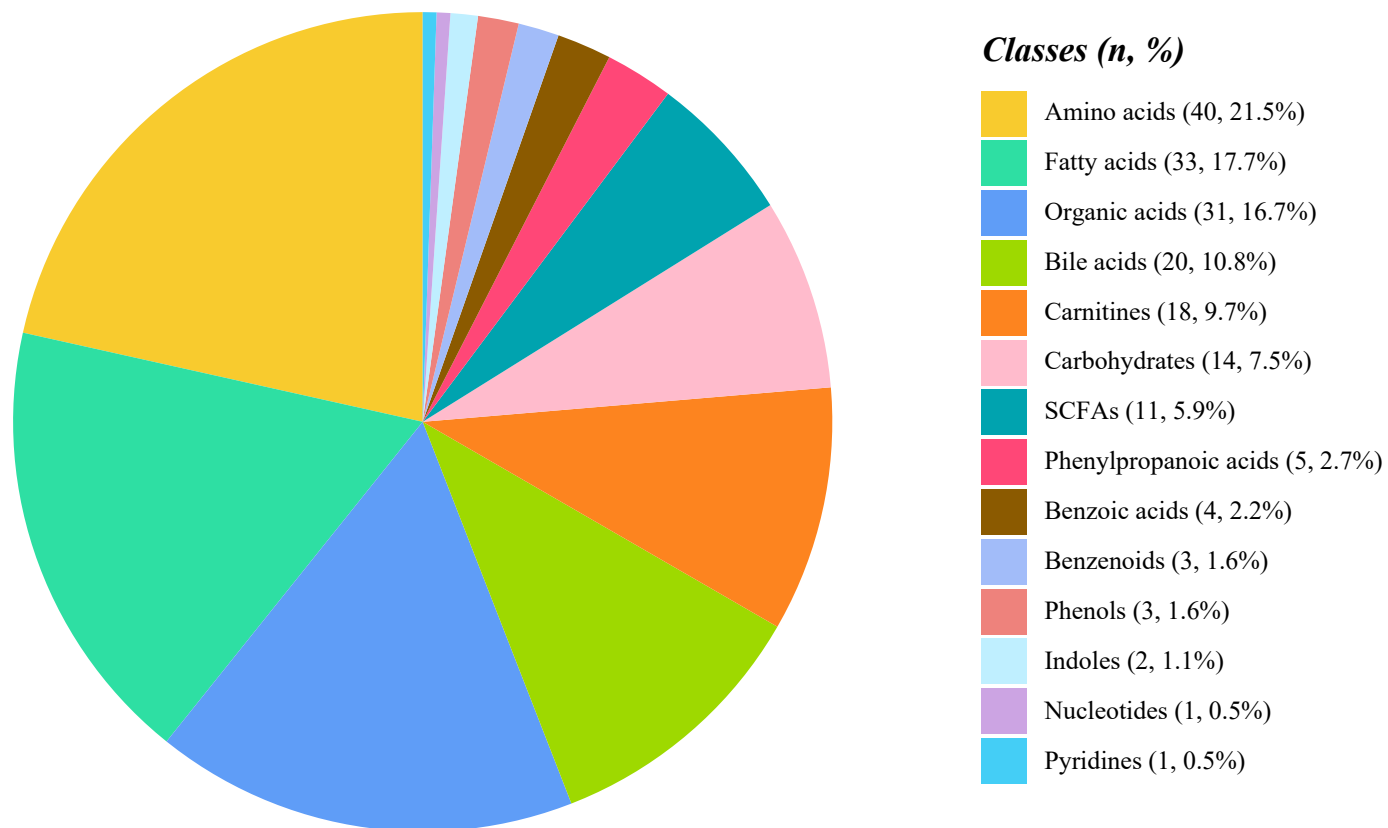


Supplementary Figure 2 Reproducibility and reliability of the metabolite quantitation

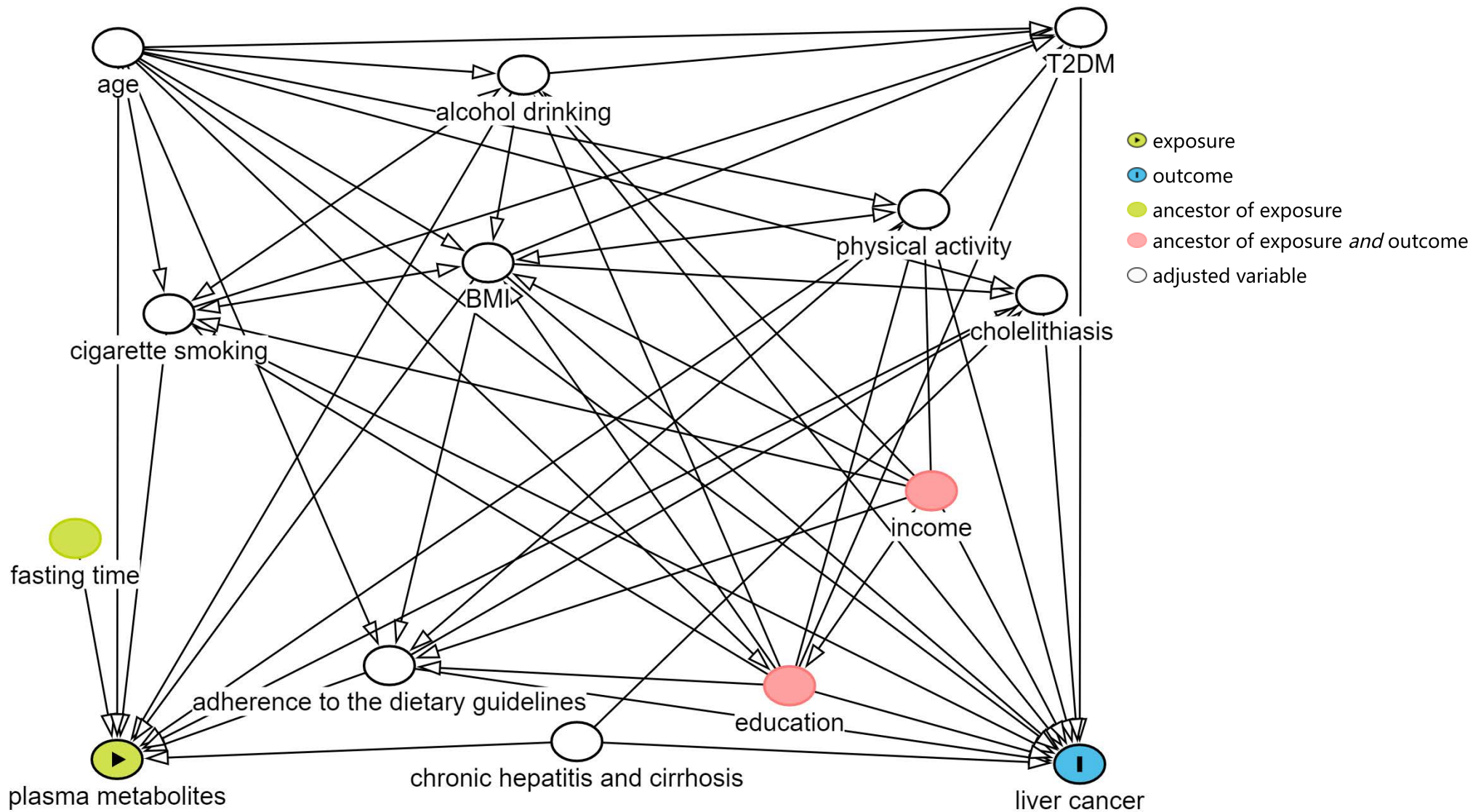
a: Multivariate Control Chart (MCC). MCC displays the data measured on the process over time and monitors variations produced by the laboratory procedures. Each dot in the MCC represents an individual sample. Most of the score values was within three standard deviation ( $\pm 3SD$ ).

b: PCA score plots. The pooled QC samples were clustered and close to the origin, indicating good reproducibility.

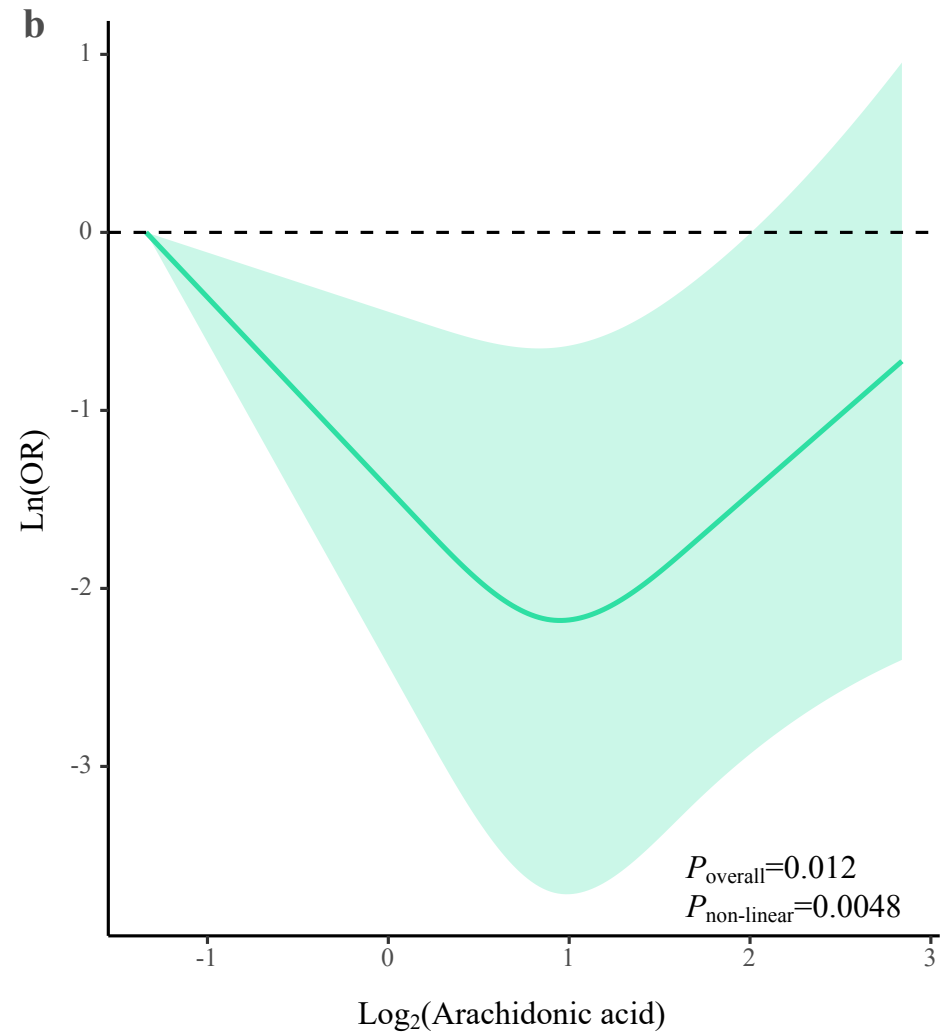
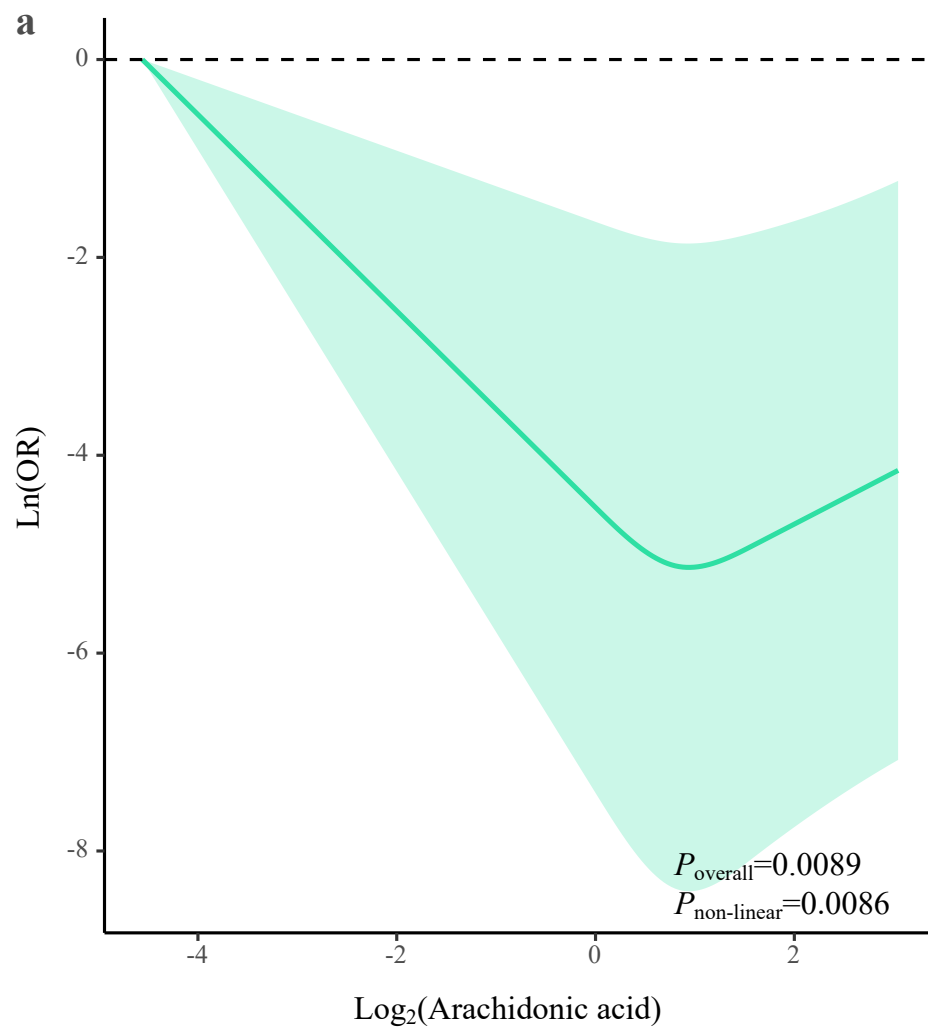
c: Spearman's rank correlation matrix for the pooled QC samples. The correlation coefficients of the pooled QC samples are all higher than 0.98, and most of them were higher than 0.99.



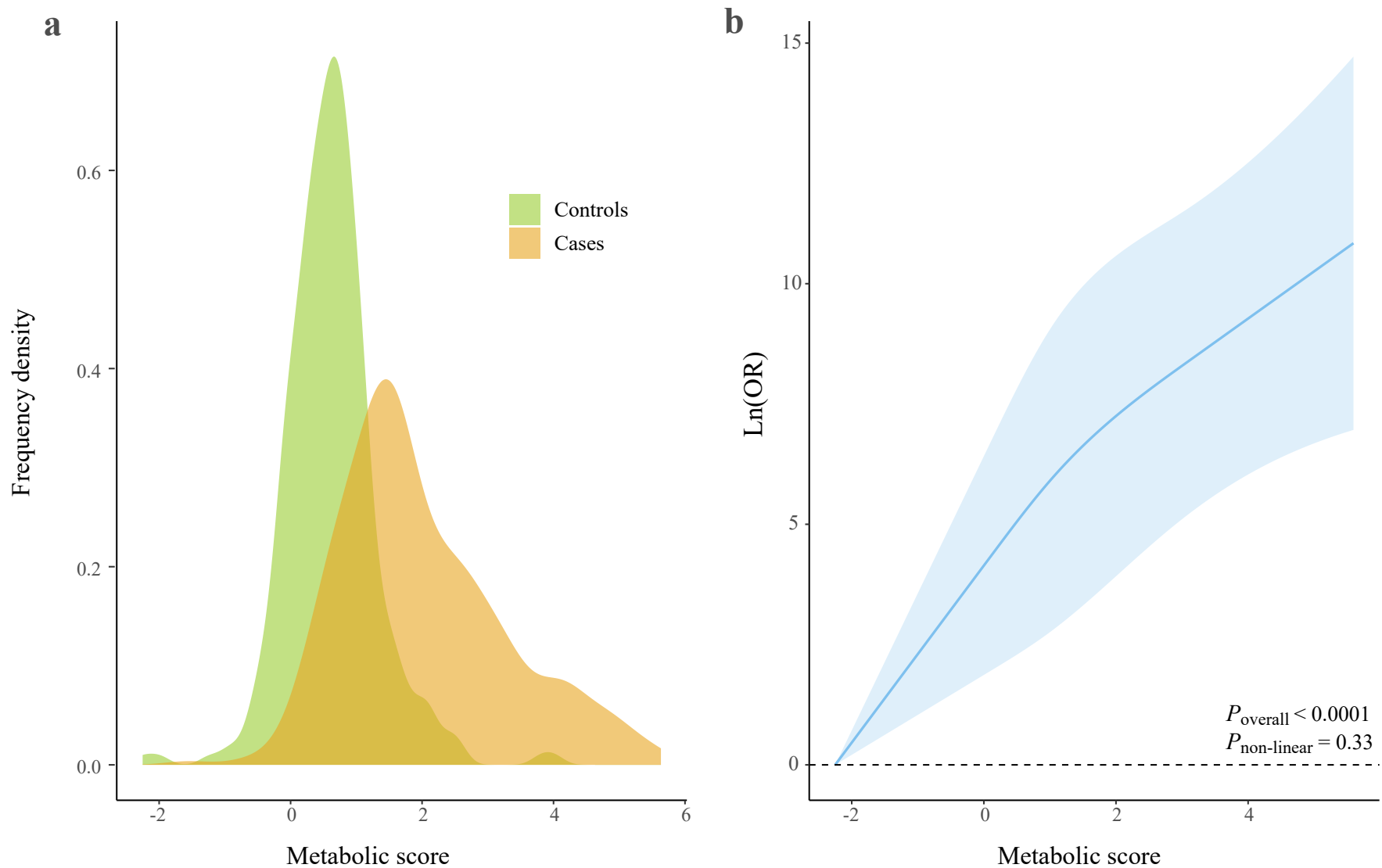
Supplementary Figure 3 Count and proportions of different classes of the quantitated metabolites



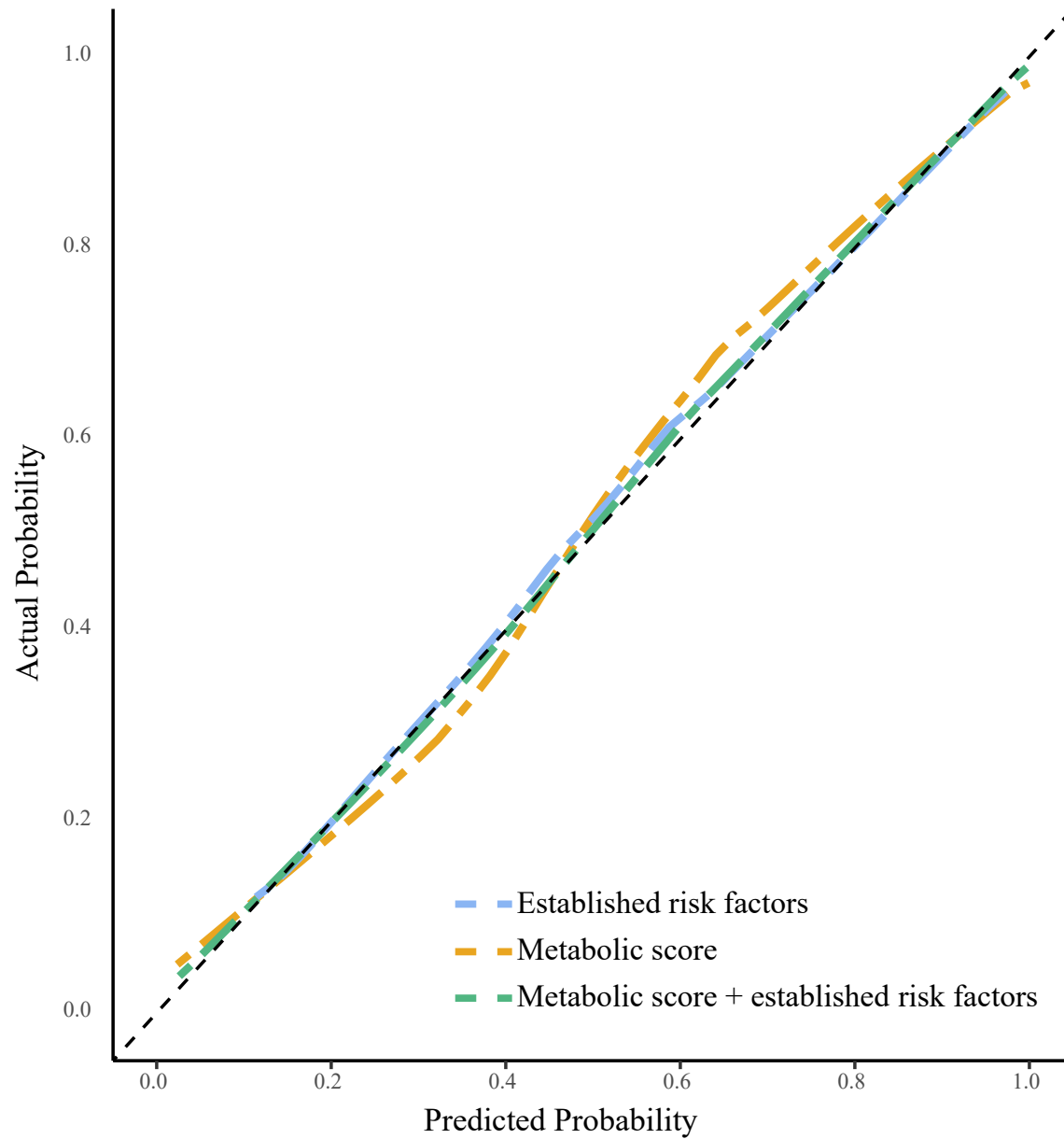
Supplementary Figure 4 Assumed directed acyclic graph on potentially causal pathways in the association between plasma metabolites and liver cancer risk



Supplementary Figure 5 Dose-response relationships between arachidonic acid and liver cancer risk among participants (a) whose follow-up time  $\geq 2$  years; (b) without history of chronic hepatitis, cirrhosis, or cholelithiasis or seropositive for HBsAg  
Log<sub>2</sub>-transformed concentrations of arachidonic acid was fitted with a 3-knots restricted cubic spline (10%, 50% and 90%). The ORs were derived from an unconditional logistic regression model. The model was adjusted for age, cigarette smoking, alcohol drinking, BMI, physical activity, CHFP score, medical history of hepatitis and cirrhosis (for a), medical history of cholelithiasis (for a), and medical history of T2DM. The shaded area refers to 95% CIs.

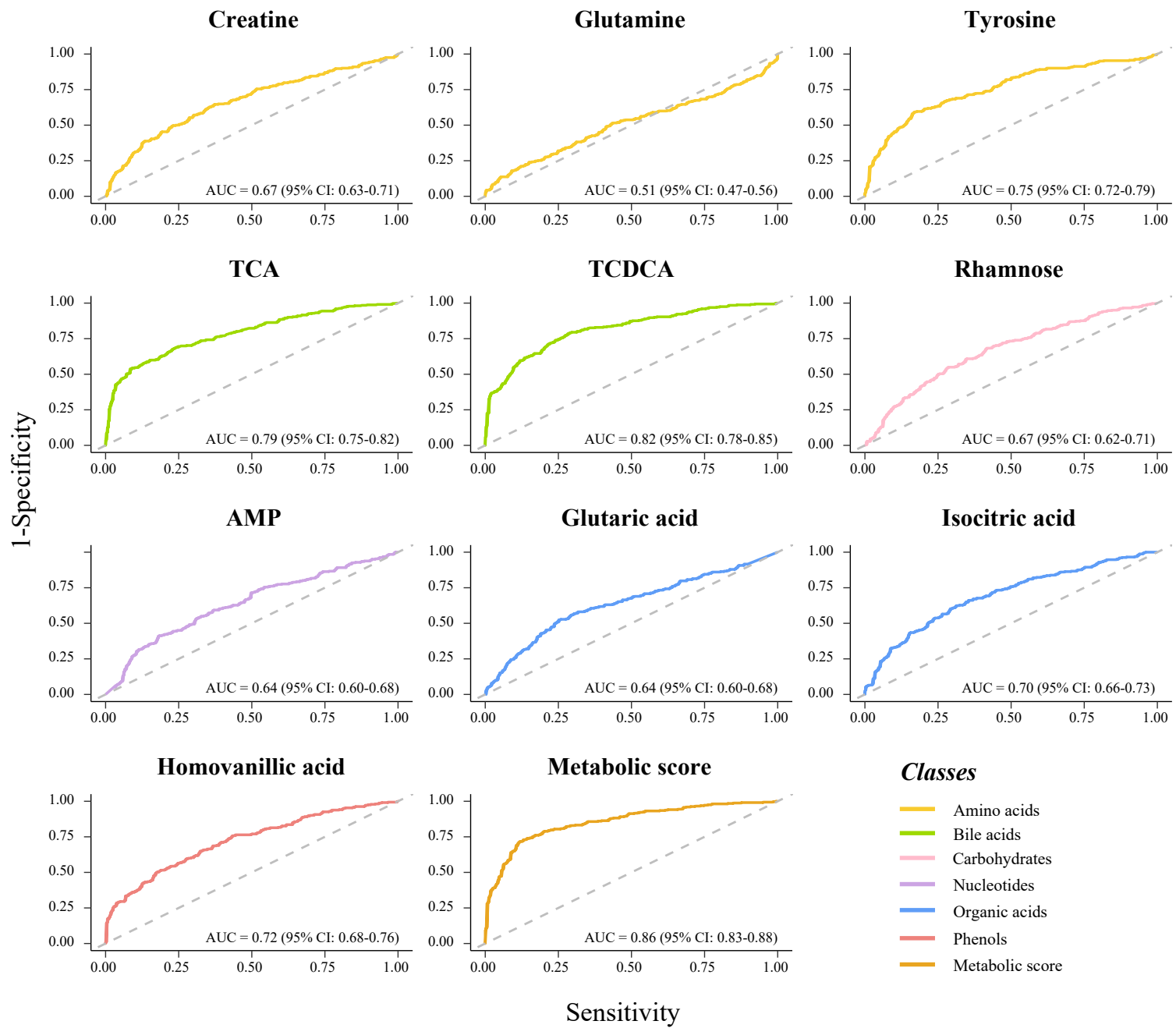


Supplementary Figure 6 Distributions of LASSO-derived metabolic score and the association between metabolic score and liver cancer risk  
a: The frequency density of metabolic score for liver cancer cases and controls;  
b: Dose-response association between metabolic score and liver cancer risk. The metabolic score was fitted with a 3-knots restricted cubic spline (10%, 50% and 90%). The ORs were derived from a conditional logistic regression model. The model was conditioned on matching factors and adjusted for age, cigarette smoking, alcohol drinking, BMI, physical activity, CHFP score, medical history of hepatitis and cirrhosis, medical history of cholelithiasis and medical history of T2DM. The shaded area refers to 95% CIs.



Supplementary Figure 7 Bias-corrected calibration plot for three prediction models





Supplementary Figure 8 Discrimination of the LASSO-derived metabolic score and its components