

Online-Only Supplementary Materials

1. Supplementary materials for bile acids measurement.
2. Supplementary materials for statistic.
3. Supplementary Figure 1. Schematic diagram illustrates the study design adopted.
4. Supplementary Table 1. Association between quartile of serum bile acids and incident diabetes (Odds ratio, 95% confidence interval).
5. Supplementary Table 2. Association between bile acid per SD increment and risk of incident diabetes with further adjustment of HbA1c.
6. Supplementary Table 3. Baseline characteristics of the nested case and control in Ad-hoc Analysis.
7. Supplementary Table 4. Association between bile acid per SD increment and risk of incident diabetes in Ad-hoc Analysis.
8. Supplementary Table 5. Characteristics of studies investigating associations of bile acids and risk of type 2 diabetes.

1. Supplementary materials for bile acids measurement

Serum samples collected at baseline were analyzed for BAs using the UPLC-MS/MS with multiple reactions monitoring methods in an ACQUITY UPLC system (Waters, USA) coupled to a triple quadrupole mass spectrometer (Waters, USA). For liquid chromatography separation, an UPLC BEH C8 column (100 mm×2.1 mm i.d.; 1.7 µm; Waters, USA) was used at 40 °C and a flow rate of 0.2 mL/min. The mobile phases were 10mM ammonium bicarbonate solution (A) and 100% acetonitrile (B). The gradient elution was set at 25% (v/v) B for 0.5 min, linearly increased to 40% B during the next 12.5 min, linearly increased to 90% B during the next 1 min, linearly decreased to 25% B during the next 0.5 min, and finally maintained at this composition for an additional 2.5 min. The injection volume was 5 µl. The mass spectrometer was used in electrospray ionization (ESI) negative mode.

2. Supplementary Materials for statistic

2.1 Ad hoc analysis

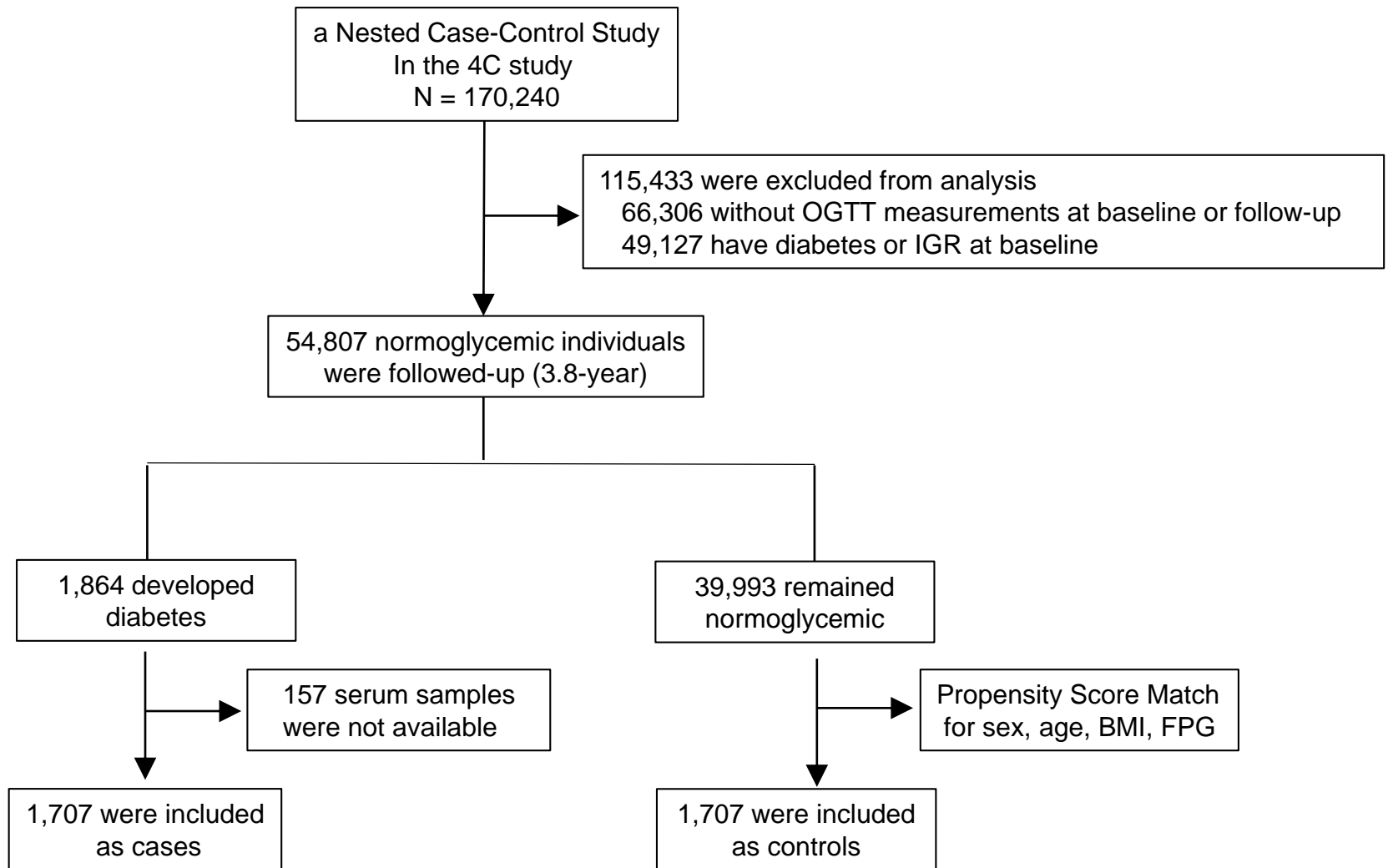
Although we did age, sex, BMI, and FPG propensity score matching for case and control groups, there are however notable metabolic differences between the groups. To better match the case and control, we carried out an ad-hoc analysis on those matched for these parameters as well. Frequency matching is a sampling design used in case–control studies to assure that cases and controls have the same distributions over strata defined by matching factors. First, we did frequency match of the two groups on these factors, including SBP, 2hr glucose, cholesterol, triglycerides, liver enzymes and HOMA-IR. The final dataset showed no significant differences in these factors between the two groups in 1326 pairs of cases and controls (Supplementary Table 3). We next conducted further analyses regarding the odds ratio of incident diabetes associated with serum BAs (Supplementary Table 4).

2.2 Multiscale embedded correlation network analysis using MEGENA

R package MEGENA was used to build correlation network from differentially correlated BA pairs. Differential correlation was calculated using R package DGCA. Only BA pairs with differential correlation ($P < 0.05$) were included for analyses. Colored lines represent the direction and strength of the correlation in incident diabetes and controls, and the number of lines that follows indicates the number of BA pairs in the global networks exhibiting this pattern of change. For instance, 1) the pink line ++/0/ 13 in the second column under the network indicates that correlation between two connected BA pairs was unrelated (0) in controls, and they became

positively correlated (++) in incident diabetes. A total of 13 BA pairs connected by pink lines in the network displayed this pattern of change. 2) the dark blue line +/++/ 3 in the third column under the network indicates that correlation between two connected BA pairs was (++) in controls, and the correlation became weaker (+) in incident diabetes. A total of 3 BA pairs connected by dark blue lines in the network displayed this pattern of change. + $P < 0.05$, ++ $P < 0.01$, both are positive correlated. – $P < 0.05$, -- $P < 0.01$, both are negative correlated. 0 $P > 0.05$.

Supplementary Figure 1. Schematic diagram illustrates the study design adopted.



Supplementary Table 1. Association between serum bile acids and incident diabetes (Odds ratio, 95% confidence interval)					
	Quartiles of BAs				
	Q1	Q2	Q3	Q4	<i>P</i> trend
Multi-variable adjusted model*					
Primary Unconjugated BA					
CA	1.00 (ref)	0.95 (0.78-1.16)	0.87 (0.72-1.06)	0.74 (0.60-0.90)	0.0024
CDCA	1.00 (ref)	0.82 (0.67-1.003)	0.90 (0.74-1.10)	0.76 (0.62- 0.93)	0.0216
Primary Conjugated BA					
GCA	1.00 (ref)	1.21 (0.99-1.49)	1.11 (0.90-1.36)	1.50 (1.23-1.84)	0.0004
GCDCA	1.00 (ref)	0.97 (0.79-1.19)	1.09 (0.89-1.34)	1.34 (1.09-1.64)	0.0018
TCA	1.00 (ref)	1.22 (0.995-1.53)	1.32 (1.07-1.62)	1.61 (1.31-1.97)	<0.0001
TCDCA	1.00 (ref)	1.15 (0.94-1.41)	1.27 (1.04-1.56)	1.51 (1.23-1.84)	<0.0001
GDCS	1.00 (ref)	1.18 (0.96-1.44)	1.08 (0.88-1.33)	1.35 (1.10-1.67)	0.0132
TCDCS	1.00 (ref)	0.98 (0.80-1.20)	1.03 (0.85-1.26)	1.18 (0.96-1.44)	0.0943
GCDCA-glucuronide	1.00 (ref)	1.04 (0.85-1.28)	1.17 (0.95-1.44)	1.16 (0.94-1.43)	0.1117
Secondary Unconjugated BA					
DCA	1.00 (ref)	0.88 (0.73-1.07)	0.82(0.67-0.996)	0.83 (0.68-1.01)	0.0411
LCA	1.00 (ref)	0.96 (0.78-1.17)	0.96 (0.78-1.17)	1.01 (0.83-1.24)	0.8843
UDCA	1.00 (ref)	0.93 (0.76-1.14)	0.79 (0.65-0.97)	0.90 (0.74-1.09)	0.1275
Secondary Conjugated BA					
GDCS	1.00 (ref)	0.91 (0.75-1.11)	0.92 (0.75-1.12)	1.01 (0.83-1.23)	0.9474
TDCS	1.00 (ref)	0.89 (0.73-1.09)	0.95 (0.78-1.16)	1.03 (0.84-1.26)	0.6428
GDCA	1.00 (ref)	0.81 (0.66-0.99)	0.98 (0.81-1.20)	0.96 (0.79-1.17)	0.8550
GDCA-glucuronide	1.00 (ref)	1.05 (0.86-1.27)	0.94 (0.77-1.15)	0.97 (0.79-1.19)	0.5426
GLCA	1.00 (ref)	0.76 (0.62-0.92)	0.82 (0.67-1.001)	0.92 (0.76-1.12)	0.6582
GLCAS	1.00 (ref)	0.86 (0.70-1.05)	0.86 (0.70-1.06)	1.05 (0.86-1.29)	0.6358
GUDCA	1.00 (ref)	0.98 (0.80-1.20)	1.01 (0.82-1.23)	1.13 (0.92-1.38)	0.2186
TDCA	1.00 (ref)	0.94 (0.77-1.15)	0.97 (0.79-1.19)	1.17 (0.96-1.43)	0.0896
TLCA	1.00 (ref)	0.77 (0.62-0.94)	1.01 (0.82-1.23)	1.11 (0.91-1.36)	0.0660
TLCAS	1.00 (ref)	0.90 (0.74-1.10)	0.88 (0.72-1.08)	1.06 (0.86-1.30)	0.6526
TUDCA	1.00 (ref)	1.19 (0.98-1.45)	1.06 (0.86-1.31)	1.44 (1.18-1.77)	0.0016
Ratio					
TCA/CA	1.00 (ref)	0.94 (0.77-1.15)	1.32 (1.08-1.61)	1.57 (1.28-1.91)	<0.0001
GCA/CA	1.00 (ref)	1.09 (0.89-1.34)	1.44 (1.18-1.76)	1.60 (1.31-1.96)	<0.0001
TCDCA/CDCA	1.00 (ref)	1.28 (1.04-1.57)	1.38 (1.12-1.70)	1.73 (1.41-2.12)	<0.0001
GCDCA/CDCA	1.00 (ref)	1.09 (0.89-1.33)	1.34 (1.10-1.65)	1.51 (1.24-1. 85)	<0.0001
12 α OH/non12 α OH BA	1.00 (ref)	1.08(0.89 -1.31)	0.97 (0.80-1.18)	0.93 (0.77 -1.13)	0.3198
Fully adjusted model †					
Primary Unconjugated BA					
CA	1.00 (ref)	1.06 (0.85-1.33)	0.89 (0.71-1.11)	0.82 (0.65-1.04)	0.0392
CDCA	1.00 (ref)	0.85 (0.68-1.07)	0.91 (0.73-1.14)	0.80 (0.64-1.004)	0.0952
Primary Conjugated BA					
GCA	1.00 (ref)	1.25 (0.99-1.59)	1.05 (0.83-1.33)	1.38 (1.09-1.73)	0.0276

GCDCA	1.00 (ref)	1.02 (0.81-1.29)	1.07 (0.85-1.35)	1.23 (0.98-1.55)	0.0636
TCA	1.00 (ref)	1.21 (0.96-1.53)	1.31 (1.04-1.66)	1.41 (1.12-1.79)	0.0037
TCDCA	1.00 (ref)	1.19 (0.94-1.49)	1.26 (0.99-1.60)	1.33 (1.06-1.68)	0.0160
GDCS	1.00 (ref)	1.10 (0.87-1.38)	1.00 (0.79-1.27)	1.27 (0.995-1.61)	0.1058
TCDCS	1.00 (ref)	0.93 (0.74-1.17)	1.04 (0.83-1.30)	1.11 (0.88-1.41)	0.2439
GCDCA-glucuronide	1.00 (ref)	1.06 (0.84-1.34)	1.12 (0.89-1.42)	1.07 (0.84-1.37)	0.5073
Secondary Unconjugated BA					
DCA	1.00 (ref)	1.04 (0.83-1.30)	0.91 (0.73-1.14)	0.87 (0.70-1.10)	0.1520
LCA	1.00 (ref)	0.89 (0.71-1.12)	0.94 (0.75-1.18)	0.99 (0.79-1.24)	0.9775
UDCA	1.00 (ref)	0.92 (0.74-1.15)	0.81 (0.64-1.01)	0.90 (0.72-1.12)	0.2072
Secondary Conjugated BA					
GDCS	1.00 (ref)	0.90 (0.72-1.13)	0.96 (0.77-1.20)	1.00 (0.80-1.25)	0.8717
TDCS	1.00 (ref)	0.91 (0.72-1.14)	0.98 (0.78-1.22)	0.97 (0.78-1.22)	0.9790
GDCA	1.00 (ref)	0.85 (0.68-1.07)	1.03 (0.82-1.29)	0.95 (0.76-1.19)	0.9554
GDCA-glucuronide	1.00 (ref)	1.09 (0.88-1.36)	0.99 (0.79-1.24)	0.91 (0.72-1.14)	0.2977
GLCA	1.00 (ref)	0.86 (0.69-1.08)	0.90 (0.71-1.13)	0.98 (0.78-1.21)	0.9461
GLCAS	1.00 (ref)	0.92 (0.74-1.16)	0.94 (0.75-1.19)	1.11 (0.88-1.40)	0.3881
GUDCA	1.00 (ref)	1.12 (0.89-1.40)	1.05 (0.84-1.32)	1.11 (0.88-1.39)	0.5078
TDCA	1.00 (ref)	0.97 (0.78-1.22)	0.99 (0.79-1.25)	1.09 (0.88-1.37)	0.3936
TLCA	1.00 (ref)	0.78 (0.62-0.98)	1.00 (0.80-1.25)	1.08 (0.86-1.35)	0.1997
TLCAS	1.00 (ref)	1.02 (0.81-1.28)	0.87 (0.69-1.10)	1.09 (0.87-1.36)	0.7658
TUDCA	1.00 (ref)	1.26 (1.001-1.58)	1.08 (0.85-1.37)	1.34 (1.07-1.69)	0.0450
Ratio					
TCA/CA	1.00 (ref)	0.91 (0.72-1.14)	1.18 (0.94-1.48)	1.41 (1.12-1.77)	0.0005
GCA/CA	1.00 (ref)	1.08 (0.85-1.36)	1.35 (1.08-1.70)	1.43 (1.14-1.80)	0.0005
TCDCA/CDCA	1.00 (ref)	1.21 (0.96-1.53)	1.38 (1.09-1.74)	1.53 (1.21-1.93)	0.0003
GCDCA/CDCA	1.00 (ref)	1.00 (0.79-1.25)	1.17 (0.93-1.47)	1.40 (1.11-1.75)	0.0012
12 α OH/non12 α OH BA	1.00 (ref)	0.94 (0.76-1.18)	0.90 (0.71-1.13)	0.89 (0.71-1.12)	0.2917
*Adjusted for age, sex, BMI, smoking status, alcohol intake, physical activity, education status, diabetes family history, SBP, fasting blood glucose, triglycerides and total cholesterol; †Adjustment of multivariable-adjusted model plus diet pattern, HDL, ALT, AST, 2h-PG, HOMA-IR and waist circumference.					

Supplementary Table 2. Association between bile acid per SD increment and risk of incident diabetes (n=3414).

	OR (95%CI)	P value
Primary Unconjugated BA		
CA	0.94 (0.87-1.02)	0.1589
CDCA	0.94 (0.86-1.02)	0.1316
Primary Conjugated BA		
GCA	1.16 (1.06-1.27)	0.0012**
GCDCA	1.13 (1.04-1.24)	0.0062*
TCA	1.16 (1.07-1.27)	0.0006**
TCDCA	1.17 (1.07-1.27)	0.0006**
GDCS	1.12 (1.02-1.23)	0.0161
TCDCS	1.05 (0.96-1.15)	0.2675
GCDCA-glucuronide	1.06 (0.97-1.16)	0.2000
Secondary Unconjugated BA		
DCA	0.89 (0.82-0.97)	0.0106*
LCA	1.02 (0.94-1.12)	0.6122
UDCA	1.00 (0.92-1.09)	0.9537
Secondary Conjugated BA		
GDCS	0.98 (0.90-1.07)	0.6435
TDCS	0.97 (0.89-1.06)	0.5556
GDCA	0.98 (0.90-1.07)	0.6344
GDCA-glucuronide	0.96 (0.88-1.04)	0.3128
GLCA	0.99 (0.91-1.08)	0.7954
GLCAS	0.99 (0.91-1.08)	0.8410
GUDCA	1.09 (1.003-1.19)	0.0418
TDCA	1.02 (0.94-1.11)	0.5864
TLCA	1.05 (0.96-1.14)	0.3115
TLCAS	0.98 (0.90-1.07)	0.6371
TUDCA	1.14 (1.05-1.25)	0.0026*
Subgroup		
Total BAs	1.08 (0.99-1.18)	0.0786
Total primary BAs	1.08 (0.99-1.18)	0.0741
Total primary unconjugated BAs	0.94 (0.86-1.02)	0.1231
Total primary conjugated BAs	1.15 (1.05-1.26)	0.0021*
Total secondary BAs	1.01 (0.92-1.10)	0.8778
Total secondary unconjugated BAs	0.95 (0.87-1.03)	0.2233
Total secondary conjugated BAs	1.05 (0.97-1.14)	0.2571
Total taurine-conjugated BAs	1.14 (1.04-1.24)	0.0041*
Total glycine-conjugated BAs	1.14 (1.05-1.25)	0.0026*
12 α OH BAs	1.03 (0.95-1.13)	0.4509
HI	0.98 (0.88-1.11)	0.7829
Ratio		

TCA/CA	1.09 (0.99-1.20)	0.0753
GCA/CA	1.06 (0.97-1.15)	0.2249
TCDCA/CDCA	1.18 (1.08-1.28)	0.0002**
GCDCA/CDCA	1.14 (1.05-1.24)	0.0017*
12 α OH/non12 α OH BA	0.97 (0.89-1.06)	0.4547
Adjusted for age, sex, BMI, smoking status, alcohol intake, physical activity, diet, education status, diabetes family history, SBP, fasting blood glucose, triglycerides, total cholesterol, HDL, ALT, AST, 2h-PG, HOMA-IR, HbA1c and waist circumference. *FDR<0.05, **FDR<0.01		

Supplementary Table 3. Baseline Characteristics of the Nested Case and Control in Ad-hoc Analysis			
	Case	Control	<i>P</i>
N	1326	1326	
Age, years	57.59±8.89	57.69±8.69	0.7820
Male, n (%)	528 (39.8)	515 (38.8)	0.6053
Body Mass Index, kg/m ²	24.73±3.65	24.80±3.50	0.6557
Waist circumference	84.33±10.13	84.08±9.91	0.5045
High school or above education, n (%)	372 (28.1)	370 (27.9)	0.9311
Moderate and vigorous physical activity, n (%)	188 (14.2)	166 (12.5)	0.2091
Healthy diet, n (%)	610 (46.0)	561 (42.3)	0.0553
Family history of diabetes, n (%)	152 (11.5)	94 (7.1)	0.0001
Current smoker, n (%)	228 (17.2)	243 (18.3)	0.4460
Current drinker, n (%)	146 (11.0)	162 (12.2)	0.3322
Systolic Blood Pressure, mmHg	135.09±20.32	134.06±20.06	0.1890
Fasting plasma glucose, mmol/L	5.46 (5.15-5.74)	5.44 (5.16-5.70)	0.5063
2h-plasma glucose, mmol/L	6.40 (5.53-7.10)	6.27 (5.48-6.97)	0.0840
Fasting triglycerides, mmol/L	1.27 (0.91-1.76)	1.24 (0.90-1.68)	0.2662
Fasting HDL cholesterol, mmol/L	1.33±0.36	1.35±0.35	0.0653
Fasting LDL cholesterol, mmol/L	2.86±0.86	2.88±0.81	0.4418
Fasting cholesterol, mmol/L	4.89±1.10	4.93±1.04	0.3459
AST, IU/L	21.36±8.06	21.44±7.87	0.8122
ALT, IU/L	16.38±8.54	15.73±8.04	0.0511
HOMA-IR	1.56 (1.07-2.30)	1.55 (1.13-2.12)	0.8714
Values are number (percent), mean ± standard deviation, or median (inter-quartile range). AST, aspartate aminotransferase; ALT, alanine aminotransferase; HOMA-IR, HOMA of insulin resistance.			

Supplementary Table 4. Association between bile acid per SD increment and risk of incident diabetes (n=2652).

	OR (95% CI)	P value
Primary Unconjugated BA		
CA	0.90 (0.83-0.98)	0.0127*
CDCA	0.91 (0.84-0.99)	0.0281
Primary Conjugated BA		
GCA	1.14 (1.05-1.24)	0.0029*
GCDCA	1.12 (1.03-1.22)	0.0080*
TCA	1.13 (1.04-1.23)	0.0038*
TCDCA	1.12 (1.03-1.22)	0.0061*
GDCS	1.11 (1.02-1.22)	0.0192
TCDCS	1.03 (0.95-1.12)	0.4815
GCDCA-glucuronide	1.03 (0.94-1.12)	0.5810
Secondary Unconjugated BA		
DCA	0.89 (0.82-0.97)	0.0071*
LCA	1.01 (0.93-1.10)	0.7594
UDCA	0.97 (0.89-1.05)	0.4476
Secondary Conjugated BA		
GDCS	0.97 (0.90-1.06)	0.5034
TDCS	0.96 (0.88-1.04)	0.3354
GDCA	0.97 (0.89-1.05)	0.4532
GDCA-glucuronide	0.95 (0.87-1.03)	0.1822
GLCA	0.99 (0.91-1.07)	0.7727
GLCAS	0.98 (0.90-1.06)	0.6210
GUDCA	1.08 (0.99-1.17)	0.0861
TDCA	0.996 (0.92-1.08)	0.9224
TLCA	1.03 (0.95-1.12)	0.4493
TLCAS	0.97 (0.89-1.06)	0.4834
TUDCA	1.10 (1.01-1.19)	0.0248
Subgroup		
Total BAs	1.04 (0.96-1.13)	0.3296
Total PBAs	1.05 (0.96-1.14)	0.3066
Total primary unconjugated	0.90 (0.83-0.98)	0.0160
Total primary conjugated	1.13 (1.04-1.23)	0.0060*
Total SBAs	0.97 (0.89-1.05)	0.4252
Total secondary unconjugated	0.91 (0.84-0.99)	0.0260
Total secondary conjugated	1.03 (0.95-1.11)	0.5504
Total taurine-conjugated BAs	1.10 (1.01-1.20)	0.0221
Total glycine-conjugated BAs	1.12 (1.03-1.22)	0.0072*
12 α OH BAs	0.99 (0.91-1.08)	0.8619
HI	1.02 (0.92-1.13)	0.7246

Ratio		
TCA/CA	1.12 (0.999-1.25)	0.0519
GCA/CA	1.09 (0.97-1.24)	0.1463
TCDCA/CDCA	1.16 (1.07-1.26)	0.0004**
GCDCA/CDCA	1.17 (1.08-1.27)	0.0002**
12 α OH/non12 α OH BA	0.95 (0.88-1.03)	0.2314

Adjusted for age, sex, BMI, smoking status, alcohol intake, physical activity, diet, education status, diabetes family history, SBP, fasting blood glucose, triglycerides, total cholesterol, HDL, ALT, AST, 2h-PG, HOMA-IR, and waist circumference.

Supplementary Table 5. Characteristics of studies investigating associations of bile acids and risk of type 2 diabetes

Reference	Study population	Study design	N, follow-up time	Methods/ tested BAs	Key findings
Marlene Wewalka et al. 2014, J Clin Endocrinol Metab	Volunteers U.S	Cross-sectional	62 NGR 25 IGT 12 T2DM	HPLC tandem mass spectrometry Fasting and postload serum BAs, including CA, CDCA, DCA, UDCA, muricholic acid (MCA), and their respective taurine (T) and glycine (G) conjugates	Fasting taurine-conjugated BA are higher in T2DM, compared with NGR subjects.
Christoph Nowak et al. 2016, Plos Genetics	ULSAM Sweden	Cohort, population based cross-sectional	952 non-diabetes	Untargeted plasma metabolomics by liquid chromatography/mass spectrometry (LC/MS) 3 α ,6 β ,7 β -trihydroxy-5 β -cholanoic acid, CDCA, GCDCA, CA, DCA, GCDCA, DCA-related metabolites	Positively associated with IR: deoxycholic acid-glycine conjugate, GDCA, GCA; Negatively associated with IR: 3 α ,6 β ,7 β -trihydroxy-5 β -cholanoic acid, 3 α ,6 β ,7 β -trihydroxy-5 β -cholanoic acid-related, CA;
Tove Fall et al. 2016, Diabetologia	ULSAM, Swedish; PIVUS, Swedish; TwinGene, Swedish; KORA S4, German.	Cohort, prospective, population based	ULSAM: 1060 non-DM/78 T2DM, 21 yrs; PIVUS: 900 non-DM/70 T2DM, 5 yrs; TwinGene: 1508 non-DM/122 T2DM, 6 yrs; KORA S4: 767 non-DM/88 T2DM, 7 yrs.	Untargeted plasma metabolomics by LC/MS 5,961 metabolic features	Associated with high risk of prevalent DM: deoxycholic acid glycine conjugate, GCA, DCA and GCDCA.
Vanessa D. de Mello et al. 2017, Scientific Reports	DPS Finland	Nested case-control	96 T2D, 5 yrs/104 non-DM, 15 yrs	Untargeted plasma metabolomics by LC/MS 8,607 metabolic features	Increased risk with DM: GCA, TCDCA, GCDCA, GDCA, DCA, CA
Rebecca A. Haeusler et al. 2013 Diabetes	Nondiabetes: RISC study; USA; Diabetes:recruited from Clinical Research Center; Italy.	Cross-sectional	200 non-DM/35 T2DM	Targeted plasma metabolomics by LC/MS CDCA, GCDCA, TCDCA, UDCA, HDCA, LCA, TLCA, CA, GCA, CA, DCA, GDCA, TDCA	In healthy subjects, insulin resistance was associated with increased 12 α -hydroxylated BAs