

Supplemental Online Appendix

Comprehensive clinical and genetic analyses of circulating bile acids and their associations with diabetes and its indices

Ibrahim Choucair^{1,2,*,\$}, Deepthi P. Mallela^{1,2,*}, James R. Hilser^{3,4}, Jaana A. Hartiala³, Ina Nemet^{1,2}, Valentin Gogonea^{1,2,5}, Lin Li^{1,2}, Aldons J. Lusis^{6,7}, Michael A. Fischbach⁸, W. H. Wilson Tang^{1,2,9}, Hooman Allayee, PhD^{3,4,‡}, Stanley L. Hazen^{1,2,9,‡}

¹ Department of Cardiovascular & Metabolic Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44106, USA;

² Center for Microbiome and Human Health, Cleveland Clinic, Cleveland, OH 44106, USA;

³ Department of Population & Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, 90033, USA;

⁴ Department of Biochemistry & Molecular Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, 90033, USA;

⁵ Department of Chemistry, Cleveland State University, Cleveland, OH 44115, USA;

⁶ Department of Medicine/Division of Cardiology, University of California, Los Angeles, CA, USA;

⁷ Department of Microbiology, Medicine and Human Genetics, University of California, Los Angeles, CA, USA;

⁸ Department of Bioengineering and ChEM-H, Stanford University, Stanford, CA, 94305, USA;

⁹ Department of Cardiovascular Medicine, Heart Vascular and Thoracic Institute, Cleveland Clinic, Cleveland, OH 44195, USA;

* contributed equally to this work

\$ current address: Department of Pathology and Laboratory Medicine, University of Kentucky College of Medicine, 800 Rose Street, MS117, Lexington, KY, 40536, USA.

‡ CORRESPONDENCE:

Stanley L. Hazen, MD, Ph.D., Department of Cardiovascular and Metabolic Sciences, Lerner Research Institute, Cleveland Clinic, 9500 Euclid Ave., NC1-127, Cleveland, OH, 44195. Email: Hazens@ccf.org

or

Hooman Allayee, Ph.D., Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, 2250 Alcazar Street, CSC202, Los Angeles, CA 90033. Email: hallayee@usc.edu

SUPPLEMENTAL ONLINE APPENDIX

S. No.	Content	Pages
1.	Supplemental Methods	3-8
1.1.	Mass spectrometry quantification of plasma bile acids (BAs)	3-4
1.2.	Statistical analysis	4-6
1.3.	Genotyping and imputation	6
1.4.	Genetic and Mendelian randomization analyses	7-8
2.	Supplemental Figures	9-22
3.	Supplemental Tables	23-39
4.	References	40

1. Supplemental Methods

1.1. Mass spectrometry quantification of plasma bile acids (BAs)

BAs analyzed in this study were quantified using previously published methods [1]. Briefly, we used stable isotope dilution high-performance liquid chromatography with tandem mass spectrometry (LC-MS/MS) implemented on an AB SCIEX 4000 triple quadrupole mass spectrometer (AB SCIEX; Framingham MA), with online electrospray ionization operating in negative ion mode. Plasma samples stored at -80°C were thawed at 4°C, and a 20 µL aliquot was mixed with 80 µL of an ice-cold methanolic working solution of internal standard (IS). Samples were vortexed, centrifuged, and the supernatants were transferred to glass HPLC vials with micro inserts for ultra-high performance liquid chromatography (UHPLC)-MS/MS analysis. Chromatographic separation was performed with four Shimadzu UHPLC systems (Shimadzu Scientific Instruments, Inc., MD, USA) and a dual column switching valve Rheodyne system (IDEX Health & Science, MA, USA) equipped with reverse phase chromatography columns (Kinetix C18, 2.6 µm, 150 mm x 4.6 mm ID; catalog # 00F-4462-E0; Phenomenex, Torrance, CA). Sixty BA molecular species were quantified in biological matrices from both human and mouse plasma using this method [1]. These BAs included the muricholic acids that are predominant in mice [1]. After excluding these mouse predominant BAs, 45 BAs were remaining. The LOD, LOQ, accuracy, (intra-day and inter-day) coefficients of variation (CVs) for all quantified BAs (n=45) from fasting plasma samples of 2,145 subjects are reported in **Tables S8 and S9**. Several of these BAs that are found in humans are also found in mice [1]. Therefore, this panel of BAs was designed for either human or animal (rodent) model studies. The limit of detection

(LOD) was defined as the lowest concentration of analyte in sample matrix (e.g. serum) that generated a signal-to-noise ratio of ≥ 3 [1]. The limit of quantification (LOQ) was defined as the lowest concentration of analyte in a sample matrix that generated a signal-to-noise ratio of ≥ 10 [1]. Only BAs that were >LOD in plasma samples of a third or more of the subjects were included in our analyses, leading to inclusion of 36 BAs in the present analyses.

1.2. Statistical analysis

For statistical analysis, subjects whose BA levels were not detectable (below LOD) were assigned a value of $\frac{1}{2}$ of the lowest concentration of the analyte (minimum value) quantified. In **Figure 4**, the grouped BA variables (total BAs, free, conjugated, glycine conjugated, taurine conjugated, total 12α -OH and total 6α -OH) were calculated by calculating the sum of the concentrations of individual BAs under the respective BA subcategory as reported in **Table S2**. For Spearman correlation (**Figure S3**) and the partial Spearman correlation heatmaps (**Figure S4**), the HOMA- β variable was calculated using the formula: $(360 \times \text{fasting insulin } (\mu\text{U/ml})) / (\text{fasting glucose (mg/dl)} - 63)$. Smoking in the present study was defined as documentation of active smoking at the time of enrollment and includes all tobacco products that are lit and can be inhaled (e.g., pipes, cigars).

Univariable and Multivariable logistic regression analyses were performed to calculate the Odds ratios (ORs) and 95% Confidence intervals (CIs) for diabetes, HOMA-IR and obesity outcomes where the odds ratios for all subjects and noDM meds subjects

indicate the ratios of 4th quartile (Q4) vs. 1st quartile (Q1) of each analyte. Multivariable models included adjustments for traditional risk factors including age, gender, smoking, systolic blood pressure, high-density lipoprotein cholesterol (HDLc), low-density lipoprotein cholesterol (LDLc), triglycerides (TG), and C-reactive protein (CRP) (Figures 2, 3, S1, S2, 4). In addition to these risk factors, we further adjusted the multivariable models for HOMA-IR and BMI (for Diabetes), Diabetes status (Y/N) and BMI (for HOMA-IR), and HOMA-IR and Diabetes status (Y/N) (for obesity) in Figure S2. Subjects taking exogenous insulin (n=100), including Type I diabetics (n=9), were excluded from these analyses (Figure S2).

Multiple hypothesis testing was performed in this study either by Bonferroni correction or false discovery rate correction. We performed Spearman rank-correlation analyses to determine the relationship between individual BA species and biomarkers of diabetes and obesity (Figures S3) where we reported both unadjusted P-values and Bonferroni corrected p-values. In Figure S4, we also performed partial Spearman correlation to account for confounding variables such as age, gender, smoking, systolic blood pressure, body mass index (BMI), high-density lipoprotein cholesterol (HDLc), low-density lipoprotein cholesterol (LDLc), triglycerides (TG), and C-reactive protein (CRP) for DM-relevant phenotypes (glucose, insulin, glucose/insulin ratio, HbA1C, HOMA-IR and HOMA- β) and additional adjustments for HOMA-IR and diabetic status (Y/N) were also performed for obesity relevant phenotypes (weight and BMI). Both Spearman and partial Spearman correlation heatmaps were generated using R (R-4.1.3 for Windows 64-bit) with scripts developed in-house. In Table S3, Multiple testing was also performed when plasma BA concentrations were compared between subjects with and without

diabetes among subjects who are not on anti-diabetic medication through the Benjamini-Hochberg method controlling the false discovery rate (FDR).

1.3. Genotyping and imputation

Genome-wide genotyping was carried out with either the Affymetrix Genome-Wide Human Array 6.0 Chip (n=3,031) or the Illumina Infinium Global Screening Array-24 v2.0 (GSA) BeadChip (n=1,728). Prior to imputation, genomic coordinates of SNPs on each genotyping platform were first converted to GRCh37/hg19. Quality control steps were then used to remove duplicate SNPs as well as those with call rates <97%, minor allele frequencies (MAFs) <1%, and without chromosome and base pair position. Individuals with genotype call rates <90%, of African American ancestry, and outliers from PCA analysis were also excluded, resulting in 671,968 SNPs in 2,972 participants genotyped with the Affymetrix 6.0 Chip, and 539,533 SNPs in 1,624 participants genotyped with the GSA Chip. Imputation was carried out for unmeasured SNPs on the forward (+) strand using 1000 Genomes Project Phase 3 (Version 5) and Haplotype Reference Consortium (Version r1.1 2016) as reference panels through the University of Michigan Imputation Server (<https://imputationserver.sph.umich.edu>). After imputation, subjects with discordant sex and SNPs with duplicates, multiallelic and imputation quality scores <0.3, and MAFs <1% were removed. This resulted in 9,370,552 SNPs available on the 22 autosomes and X chromosome for analysis in 1,840 GeneBank subjects with plasma BA measurements.

1.4. Genetic and Mendelian randomization analyses

Genome-wide association study (GWAS) analyses were carried out for the 14 BAs that remained significantly associated with DM after adjustment for multiple testing (false discovery rate [FDR] $P < 0.05$). Levels of deoxycholic acid (DCA), taurodeoxycholic acid (TDCA), glycodeoxycholic acid (GDCA), and 7-Ketolithocholic acid (7-Keto-LCA) were analyzed by linear regression with normal inverse transformed values and adjustment for age, sex, and genotyping array using PLINK2[2]. BAs where >25% of subjects had values below the limit of quantification (LOQ) were dichotomized by grouping subjects above or below the LOQ, and analyzed by logistic regression with adjustment for age, sex, and genotyping array. The BAs and their LOQs that were analyzed as binary variables included lithocholenic acid (9.1 nM), 12-Ketolithocholic acid (9.6 nM), 23-nordeoxycholic acid (8.7 nM), hyocholic acid (11.9 nM), taurohyocholic acid (4.1 nM), glycohyocholic acid (4.9 nM), 6-Ketolithocholic acid (8.6 nM), hyodeoxycholic acid (15.7 nM), isolithocholic acid (10.0 nM), and taurohyodeoxycholic acid (3.6 nM). Loci that were significantly associated with BA levels at the genome-wide threshold ($P < 5.0 \times 10^{-8}$) were evaluated for associations with BMI and diabetes mellitus (DM) using the results of previous GWAS [3, 4]. A reciprocal analysis was also performed by evaluating the 338 previously-reported [4] DM loci for association with BA levels using the results of our GWAS analyses in Genebank. Mendelian randomization (MR) was carried out with the Wald ratio method using the Two-Sample MR package in R v4.0.3[5]. Exposure data for isolithocholic acid, DCA, and 6-Keto-LCA levels were based on the lead SNP(s) identified for each BA from the GWAS analyses in the GeneBank cohort, and publicly available summary data for BMI [3] and DM [4] were used for the outcome data.

Reverse two-sample MR analyses were also carried out using the effect sizes of DM susceptibility variants (**Table S7**) [4] as instrument variables and betas for Iso-LCA and 6-Keto LCA levels from our GWAS analyses in GeneBank as the outcomes. Steiger tests were performed to both verify the reliability and evaluate the directionality of our instrument variables which yielded significant p-values and provided consistent evidence for DCA, 6-Keto-LCA, and Iso-LCA being causal drivers of BMI and diabetes.

2. Supplemental Figures

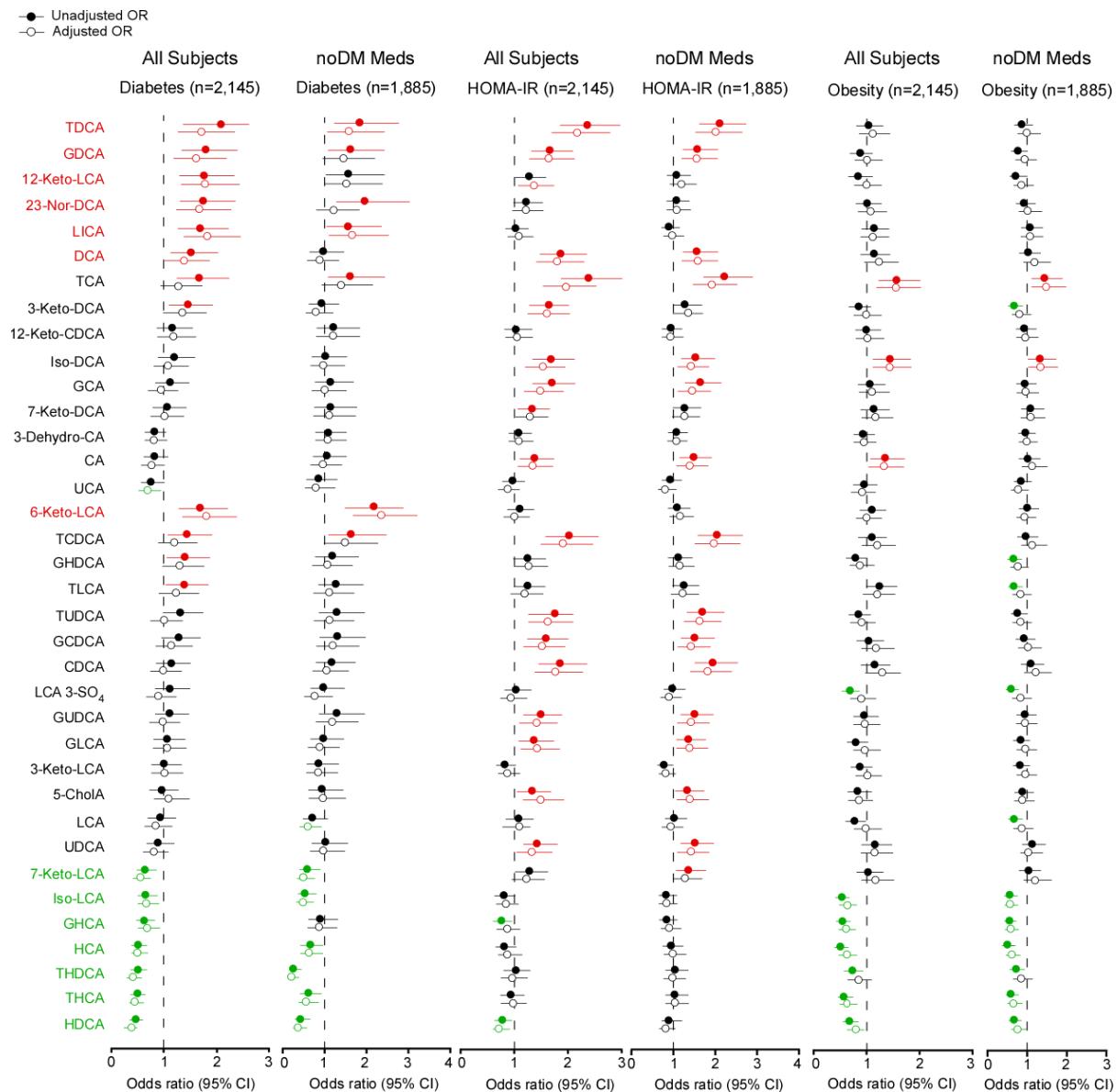


Figure S1. Association of BAs with DM and metabolic indices after accounting for anti-diabetic medications. Forest plots of odds ratios (OR) for diabetes, HOMA-IR, and obesity of 4th quartile (Q4) vs. 1st quartile (Q1) for cohort including subjects taking DM medication [All Subjects (n=2,145)] vs cohort excluding subjects taking DM medication [noDM Meds cohort (n=1,885)]. Bars represent 95% confidence interval (CI). Closed circles represent unadjusted OR, and open circles represent OR adjusted for

age, gender, smoking, systolic blood pressure, HDLc, LDLc, TG and CRP. When the association was statistically significant ($P < 0.05$), the OR (circle) and 95% CI (line) were colored red (positive association) or green (inverse association). If both the unadjusted and adjusted associations of BA with DM were significant, the label names for BAs were colored red (positive association) or green (inverse association), and black if just the unadjusted or adjusted association was significant or if both were not significant.

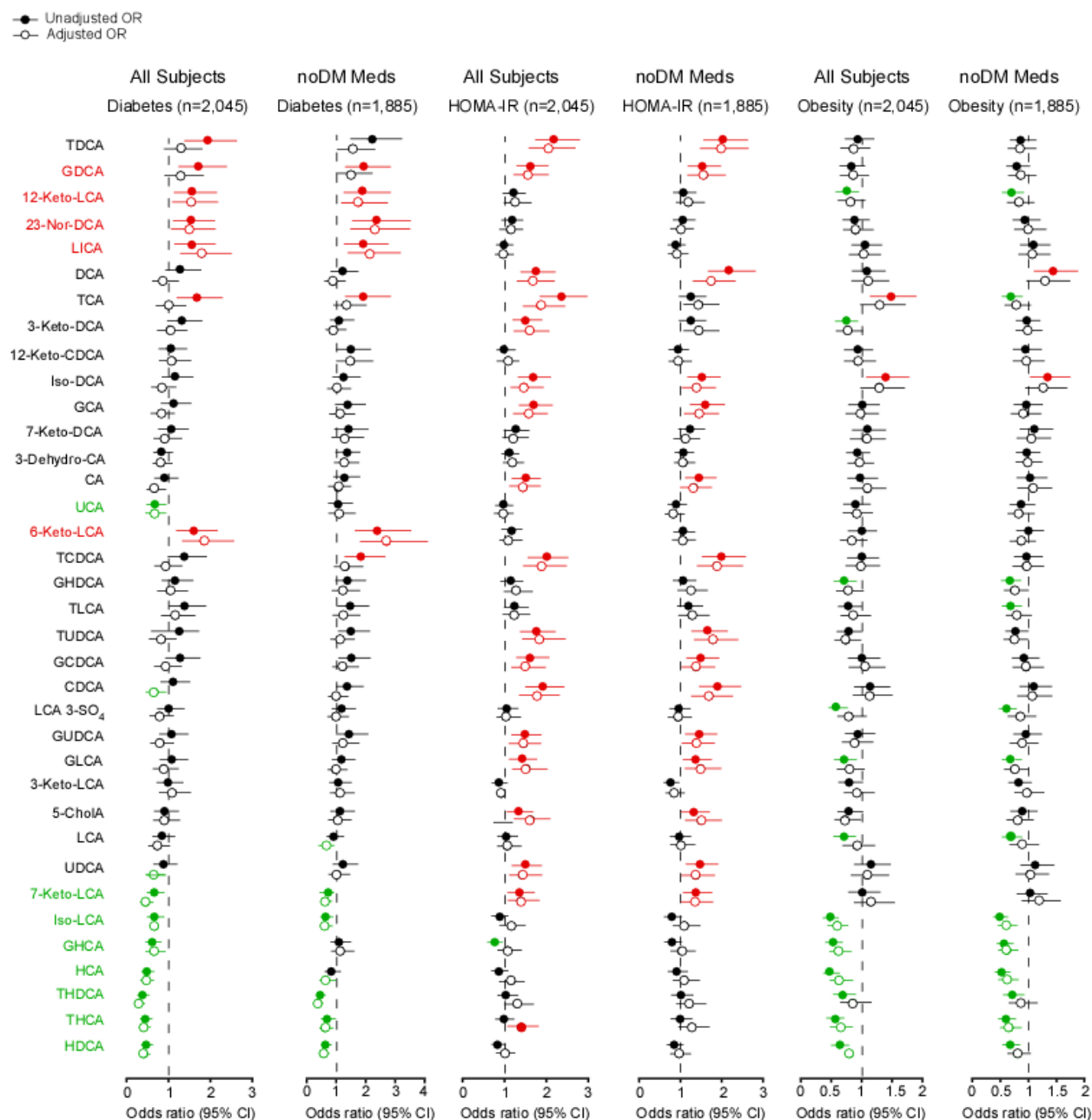


Figure S2: Association of BAs with DM and metabolic indices after accounting for HOMA-IR and BMI and excluding subjects taking exogenous insulin and Type I diabetics. Forest plots of odds ratios (OR) for diabetes, HOMA-IR, and obesity of 4th quartile (Q4) vs. 1st quartile (Q1) for: all subjects after excluding any subject taking exogenous insulin [All Subjects (n=2,045)]; and only subjects taking no DM medication

[noDM Meds cohort (n=1,885)]. Bars represent 95% confidence interval (CI). Closed circles represent unadjusted OR, and open circles represent OR adjusted for age, gender, smoking, body mass index, systolic blood pressure, HDLc, LDLc, TG, CRP in addition to HOMA-IR and BMI for diabetes; BMI and diabetic status (Yes/No) for HOMA-IR; diabetic status (Yes/No) and HOMA-IR for obesity. When the association was statistically significant ($P < 0.05$), the OR (circle) and 95% CI (line) were colored red (positive association) or green (inverse association). If both the unadjusted and adjusted associations of BA with DM were significant, the label names for BAs were colored red (positive association) or green (inverse association), and black if just the unadjusted or adjusted association was significant or if both were not significant.

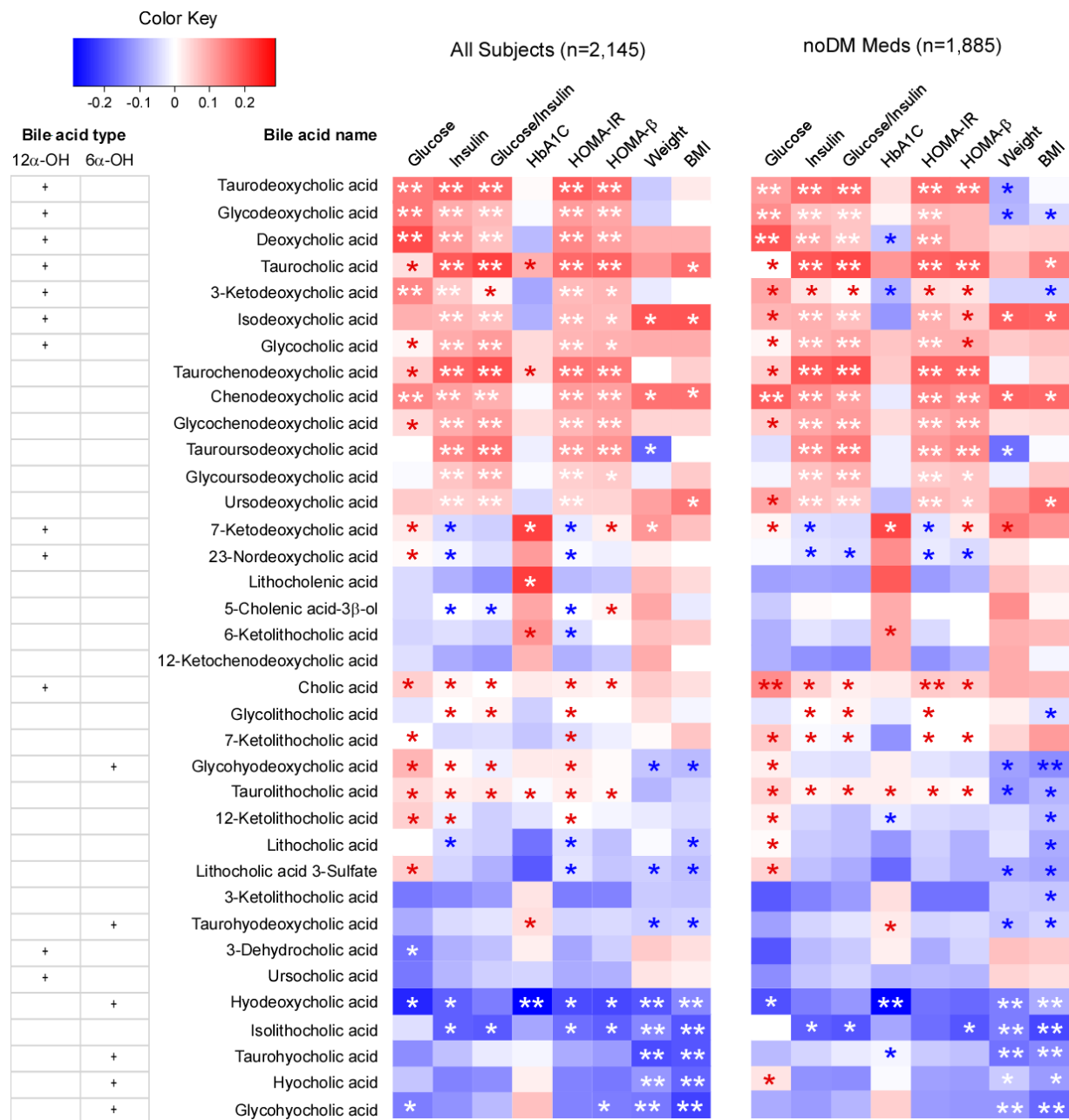


Figure S3. Relationship between grouped plasma BA concentrations and biomarkers of diabetes and obesity. Heat map showing Spearman correlation between BAs and fasting glucose, insulin, glucose to insulin ratio, hemoglobin A1c (HbA1c), homeostatic model assessment for insulin resistance (HOMA-IR), homeostatic model assessment for β -cell function (HOMA- β), weight, and body mass index (BMI). Heat maps for all subjects (n=2,145) and subjects not taking anti-diabetic medication

(n=1,885) are shown for comparison. The correlation strength (ρ) is shown by the color key; red represents a positive association, blue represents a negative association and white no association. * indicates unadjusted $P < 0.05$ and ** represents Bonferroni corrected $P < 0.05$. Darker color boxes indicate $\text{Rho } (\rho) > 0.1$ (red) or $\text{Rho } (\rho) < -0.1$ (blue). Two columns were indicated on the left side of the bile acid names to show whether each BA represents a 12α -hydroxylated BA or a 6α -hydroxylated BA using a “+” sign.

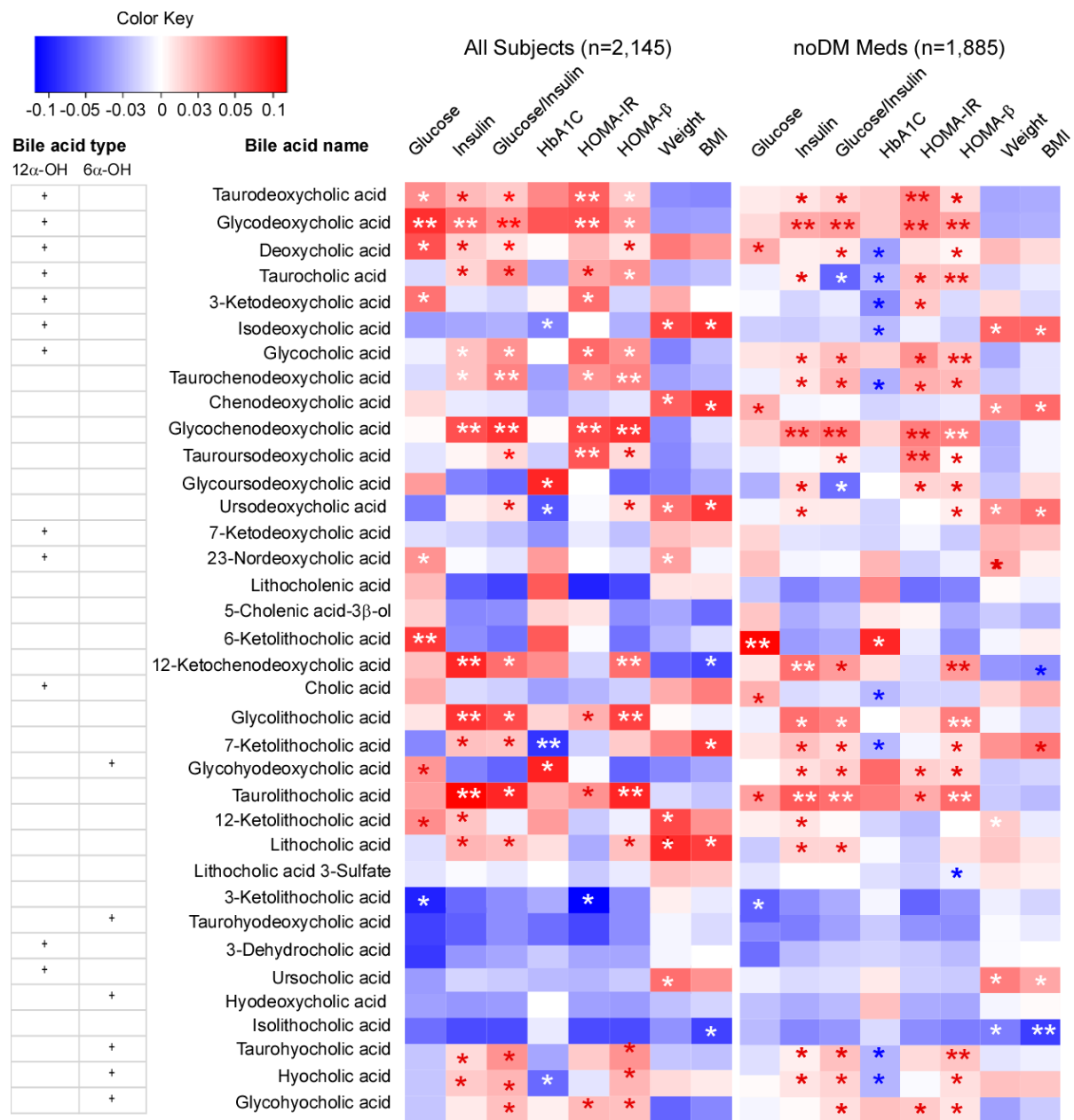


Figure S4: Relationship between grouped plasma BA concentrations and biomarkers of diabetes and obesity after adjusting for confounders. Heat map showing Partial Spearman correlation between BAs and fasting glucose, insulin, glucose to insulin ratio, hemoglobin A1c (HbA1c), homeostatic model assessment for insulin resistance (HOMA-IR), homeostatic model assessment for β -cell function (HOMA- β), weight, and body mass index (BMI). Heat maps for all subjects (n=2,145)

and subjects not taking anti-diabetic medication (n=1,885) are shown for comparison. All partial Spearman correlation analyses were adjusted for age, gender, body mass index, smoking, systolic blood pressure, low-density lipoprotein, high-density lipoprotein, triglycerides, and C-reactive protein. The correlation strength (ρ) is shown by the color key; red represents a positive association, blue represents a negative association, and white no association. * indicates partial Spearman correlation $P < 0.05$. ** indicates partial Spearman correlation Bonferroni corrected $P < 0.05$. Darker color boxes indicate $Rho(\rho) > 0.05$ (red) or $Rho(\rho) < -0.05$ (blue). Two columns were indicated on the left side of the bile acid names to show whether each BA represents a 12α -hydroxylated BA or a 6α -hydroxylated BA using a “+” sign.

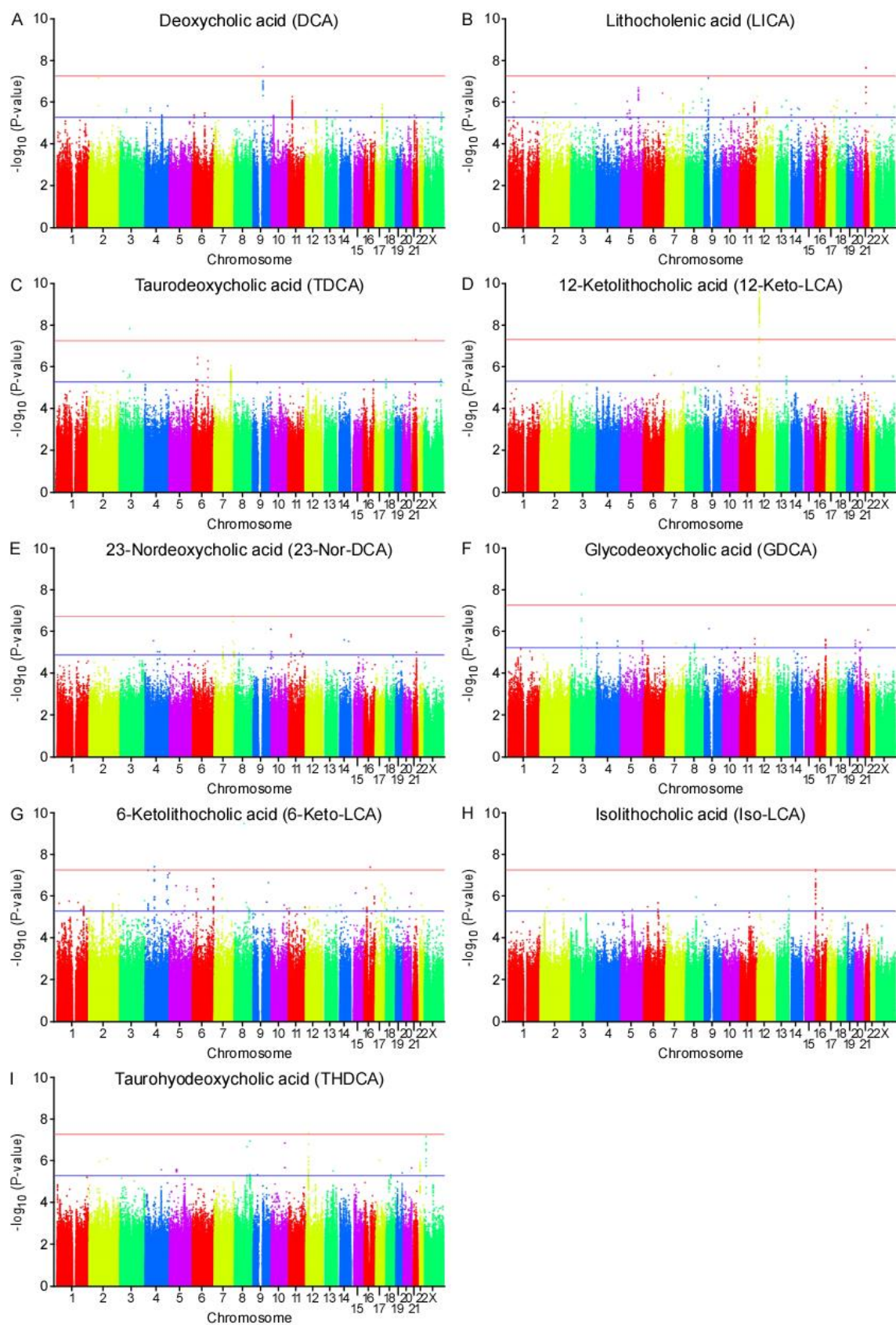
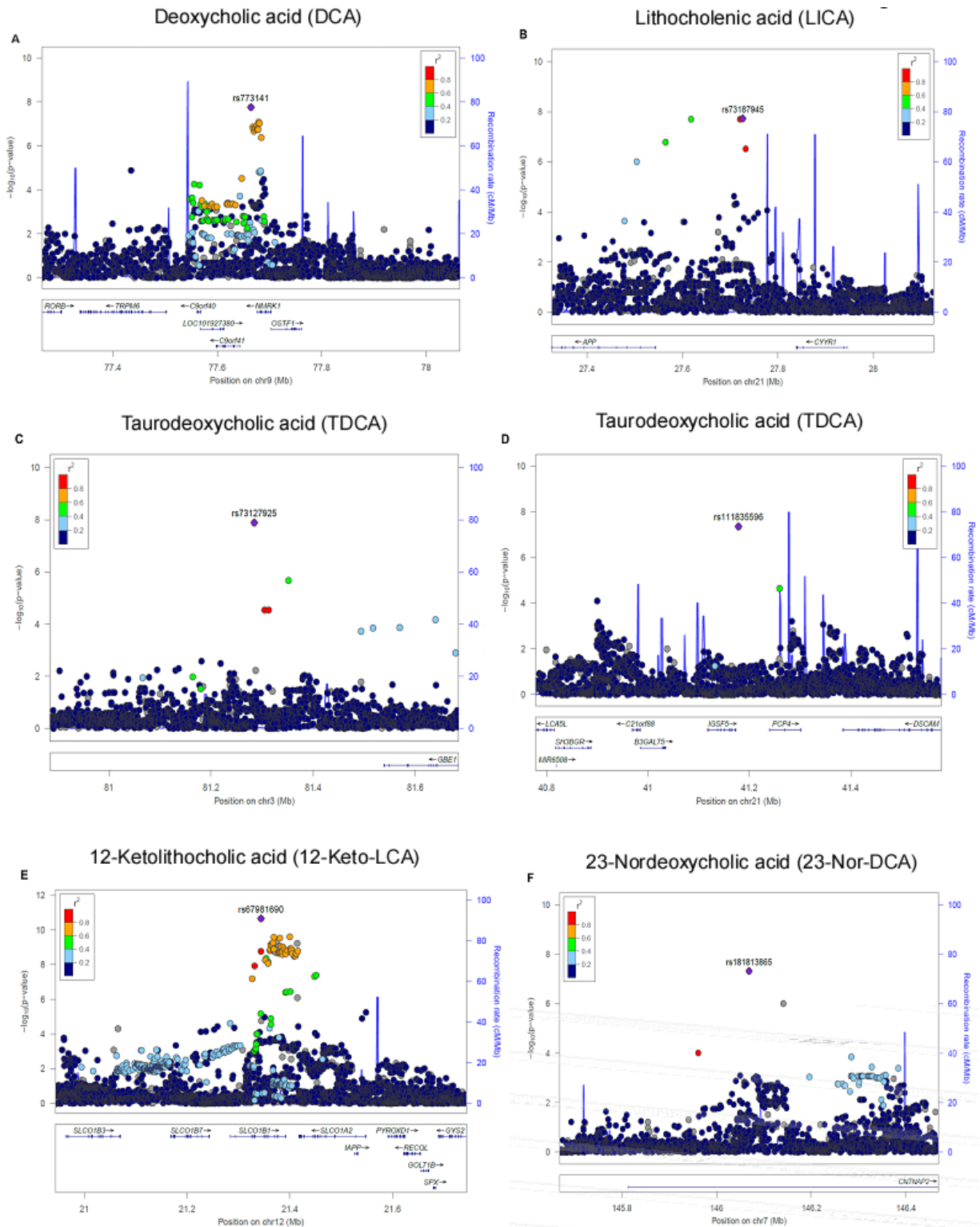


Figure S5. GWAS results for BAs yielding significantly associated loci. Manhattan plots show loci significantly associated with **(A)** DCA, **(B)** LICA, **(C)** TDCA, **(D)** 12-Keto-LCA, **(E)** 23-Nor-DCA, **(F)** GDCA, **(G)** 6-Keto-LCA, **(H)** Iso-LCA, and **(I)** THDCA. Genome-wide thresholds for significant ($P=5.0 \times 10^{-8}$) and suggestive ($P=5.0 \times 10^{-6}$) association are indicated by the horizontal red and blue lines, respectively.



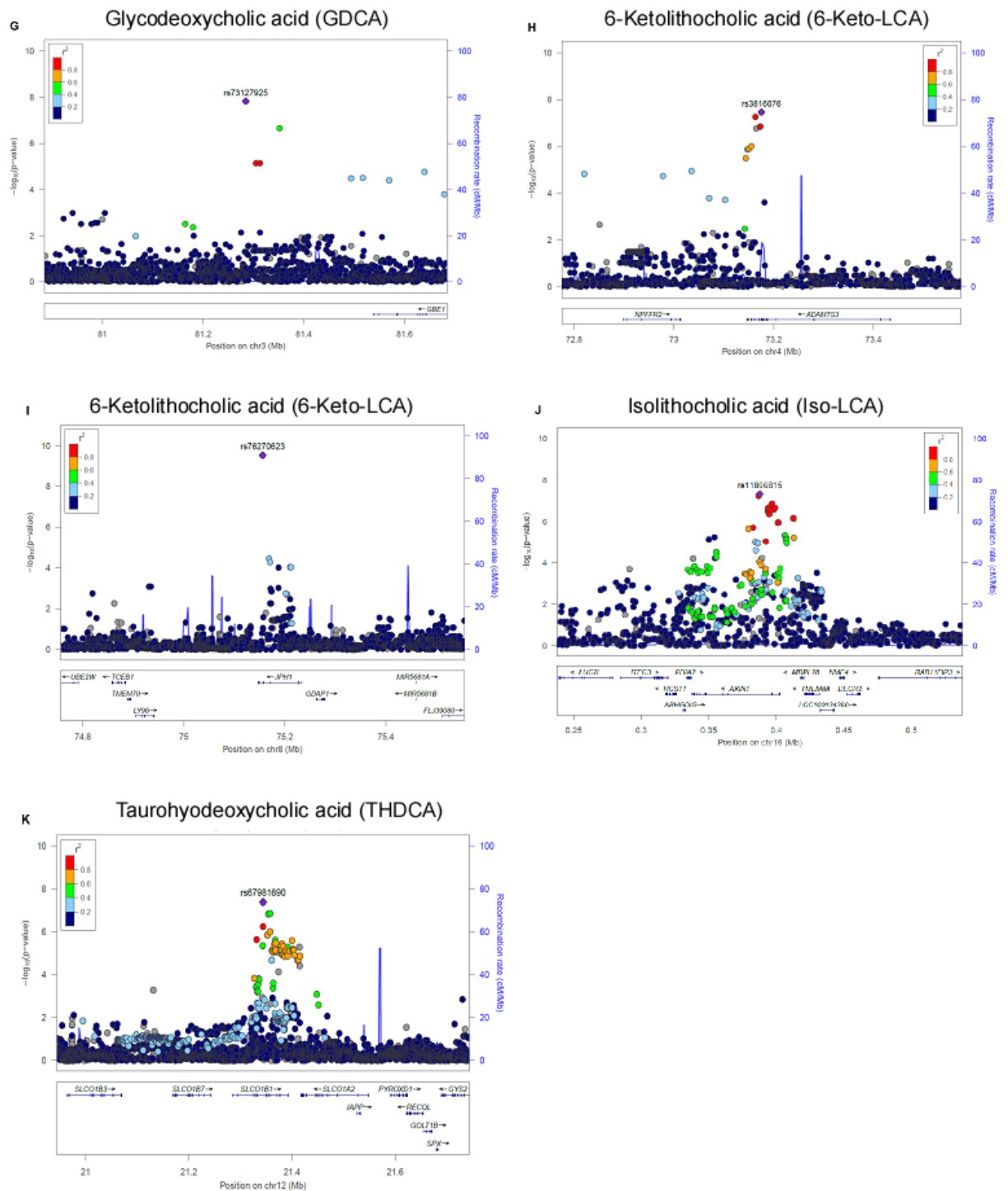


Figure S6. Regional plots of significant loci for levels of DM-associated BAs. The 800kb interval centered on the lead SNP (purple diamond) is shown for each locus significantly associated with levels of **(A)** DCA, **(B)** LICA, **(C and D)** TDCA, **(E)** 12-Keto-

LCA, **(F)** 23-Nor-DCA, **(G)** GDCA, **(H and I)** 6-Keto-LCA, **(J)** Iso-LCA, and **(K)** THDCA. Genes located within the interval are shown below the plot, and the degrees of linkage disequilibrium (LD) between the lead SNP and other variants are shown as r^2 values according to the color-coded legend. The locus on chromosome 16 for 6-Keto-LCA is not shown, as the lead variant (rs535408158) was not available in the reference European ancestry population used for making regional plots.

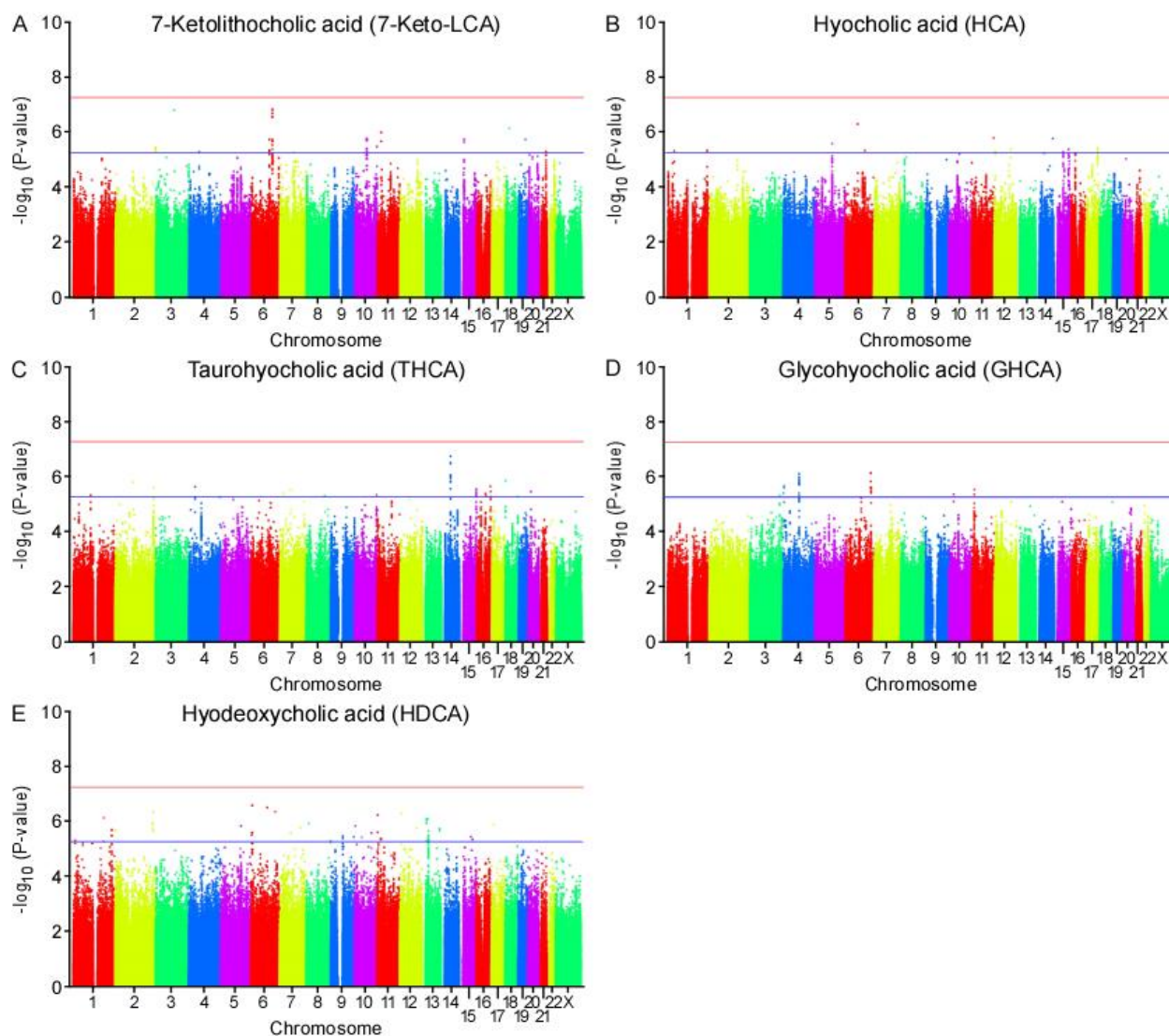
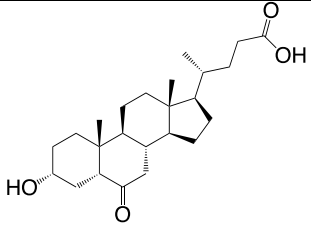
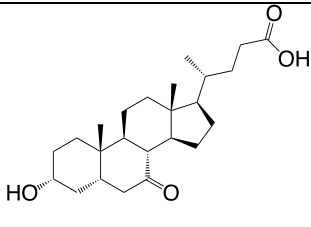
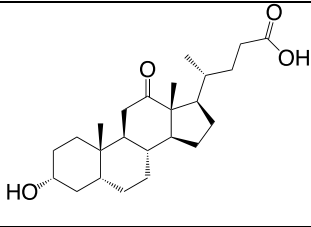
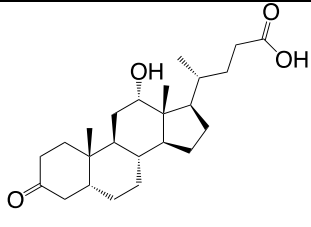
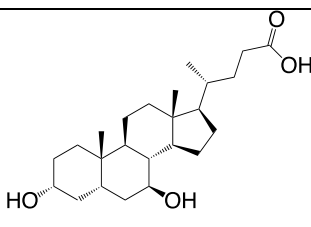
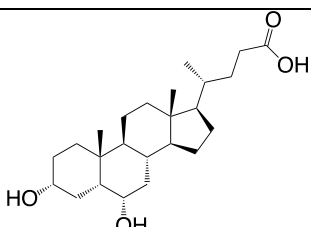
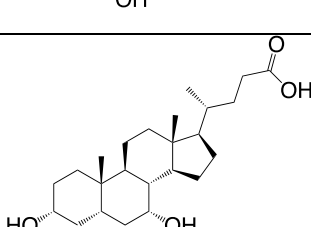


Figure S7. GWAS results for BAs yielding suggestively-associated loci. Manhattan plots show loci suggestively associated with **(A)** 7-Keto-LCA, **(B)** HCA, **(C)** THCA, **(D)** GHCA, and **(E)** HDCA. Genome-wide thresholds for significant ($P=5.0 \times 10^{-8}$) and suggestive ($P=5.0 \times 10^{-6}$) association are indicated by the horizontal red and blue lines, respectively.

3. Supplemental Tables

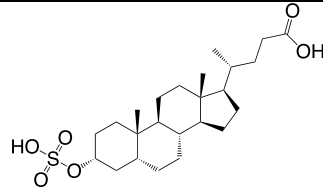
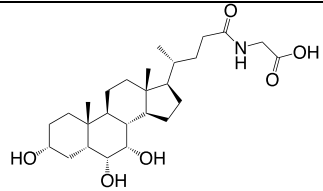
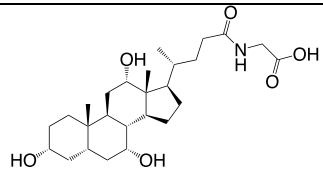
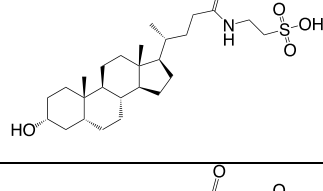
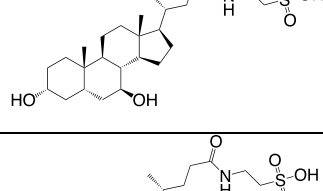
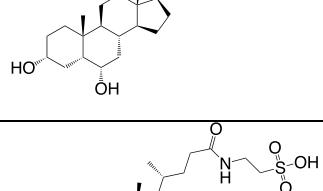
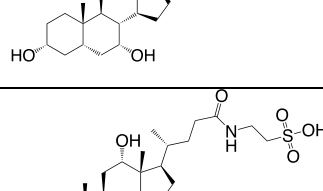
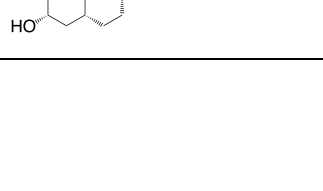
Table S1. List of BAs' common names, systematic names, InChI keys, chemical formulas and structures.

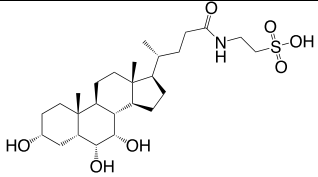
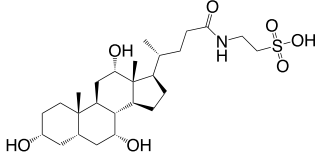
S. No	Common Name	Systematic Name	InChI Key	Formula	Structure
1	5-Cholenic acid-3 β -ol (5-Chol-A)	3 β -Hydroxy-chol-5-en-24-oic acid	HIAJCGFYHIANNA-QIZZZRFXSA-N	C ₂₄ H ₃₈ O ₃	
2	Lithocholic acid (LICA)	3 α -Hydroxy-5 β -chol-11-en-24-oic acid	FEGCPHPSRBREU-REBGREMMSA-N	C ₂₄ H ₃₈ O ₃	
3	3-Ketolithocholic acid (3-Keto-LCA) or Dehydrolithocholic acid (Dehydro-LCA)	3-Oxo-5 β -cholan-24-oic acid	KIQFUORWRVZTHT-OPTMKGCMMSA-N	C ₂₄ H ₃₈ O ₃	
4	Isolithocholic acid (Iso-LCA)	3 β -Hydroxy-5 β -cholan-24-oic acid	SMEROWZSTRWXGI-WFVDQZAMSA-N	C ₂₄ H ₄₀ O ₃	
5	Lithocholic acid (LCA)	3 α -Hydroxy-5 β -cholan-24-oic acid	SMEROWZSTRWXGI-HVATVPOCSA-N	C ₂₄ H ₄₀ O ₃	
6	23-Nordeoxycholic acid (23-Nor-DCA)	24-Nor-3 α ,12 α -dihydroxy-5 β -cholan-23-oic acid	PLRQOCVIINWCFA-AHFDLSHQSA-N	C ₂₃ H ₃₈ O ₄	

7	6-Ketolithocholic acid (6-Keto-LCA)	3 α -Hydroxy-6-oxo-5 β -cholan-24-oic acid	JWZBXKZZDYMDCJ -IJPFKRJSSA-N	C ₂₄ H ₃₈ O ₄	
8	7-Ketolithocholic acid (7-Keto-LCA)	3 α -Hydroxy-7-oxo-5 β -cholan-24-oic acid	DXOCDBGWDZAYR Q-AURDAFMXSA-N	C ₂₄ H ₃₈ O ₄	
9	12-Ketolithocholic acid (12-Keto-LCA)	3 α -Hydroxy-12-oxo-5 β -cholan-24-oic acid	CVNYHSDFZXHMM J-VPUMZWJWSA-N	C ₂₄ H ₃₈ O ₄	
10	3-Ketodeoxycholic acid (3-Keto-DCA)	12 α -Hydroxy-3-oxo-5 β -cholan-24-oic acid	WMUMZOAFCDOT RW-OVEHVULHSA-N	C ₂₄ H ₃₈ O ₄	
11	Ursodeoxycholic acid (UDCA)	3 α ,7 β -Dihydroxy-5 β -cholan-24-oic acid	RUDATBOHQWOJD D-UZVSRGJWSA-N	C ₂₄ H ₄₀ O ₄	
12	Hyodeoxycholic acid (HDCA)	3 α ,6 α -Dihydroxy-5 β -cholan-24-oic acid	DGABKXLVXPYZII- SIBKNMCHSA-N	C ₂₄ H ₄₀ O ₄	
13	Chenodeoxycholic acid (CDCA)	3 α ,7 α -Dihydroxy-5 β -cholan-24-oic acid	RUDATBOHQWOJD D-BSWAIDMHSA-N	C ₂₄ H ₄₀ O ₄	

14	Deoxycholic acid (DCA)	3 α ,12 α -Dihydroxy-5 β -cholan-24-oic acid	KXGVEGMKQFWNS R-LLQZFEROSA-N	C ₂₄ H ₄₀ O ₄	
15	Isodeoxycholic acid (Iso-DCA)*	7 α ,12 α -Dihydroxy-5 β -cholan-24-oic acid	ZHCAAZIHTDCFJX- QLEQUTGBSA-N	C ₂₄ H ₄₀ O ₄	
		3 β ,12 α -Dihydroxy-5 β -cholan-24-oic acid	KXGVEGMKQFWNS R-OFYXWCICSA-N	C ₂₄ H ₄₀ O ₄	
16	12-Ketodeoxycholic acid (12-Keto-CDCA)	3 α ,7 α -Dihydroxy-12-oxo-5 β -cholan-24-oic acid	MIHNUBCEFJLAGN- DMMBONCOSA-N	C ₂₄ H ₃₈ O ₅	
17	7-Ketodeoxycholic acid (7-Keto-DCA)	3 α ,12 α -Dihydroxy-7-oxo-5 α -cholan-24-oic acid	RHCPKKNRWFXMA T-RRWYKFPJSA-N	C ₂₄ H ₃₈ O ₅	
18	3-Dehydrocholic acid (3-Dehydro-CA)	7 α ,12 α -Dihydroxy-3-oxo-5 α -cholan-24-oic acid	OEKUSRBIIIZNLHZ- DJDNIQJZSA-N	C ₂₄ H ₃₈ O ₅	
19	Ursocholic acid (UCA)	3 α ,7 β ,12 α -Trihydroxy-5 β -cholan-24-oic acid	BHQCQFFYRZLCQ Q-UTLSPDKDSA-N	C ₂₄ H ₄₀ O ₅	

20	Hyocholic acid (HCA)	3 α ,6 α ,7 α -Trihydroxy-5 β -cholan-24-oic acid	DKPMWHFRUGMU KF-KWXDGCAGSA-N	C ₂₄ H ₄₀ O ₅	
21	Cholic acid (CA)	3 α ,7 α ,12 α -trihydroxy-5 β -cholan-24-oic acid	BHQCQFFYRZLCQ Q-OELDTZBJSA-N	C ₂₄ H ₄₀ O ₅	
22	Glycolithocholic acid (GLCA)	N-(3 α , -hydroxy-5 β -cholan-24-oyl)-glycine	XBSQTYHEGZTYJE -OETIFKLTSA-N	C ₂₆ H ₄₃ NO ₄	
23	Glycoursodeoxycholic acid (GUDCA)	N-(3 α ,7 β -dihydroxy-5 β -cholan-24-oyl)-glycine	GHCZAUBVMUEKK P-XROMFQGDSA-N	C ₂₆ H ₄₃ NO ₅	
24	Glycohyodeoxycholic acid (GHDCA)	N-(3 α ,6 α -dihydroxy-5 β -cholan-24-oyl)-glycine	SPOIYSFQOFYOFZ- BRDORRHWSA-N	C ₂₆ H ₄₃ NO ₅	
25	Glycochenodeoxycholic acid (GCDCA)	N-(3 α ,7 α -dihydroxy-5 β -cholan-24-oyl)-glycine	GHCZAUBVMUEKK P-GYPHWSFCSA-N	C ₂₆ H ₄₃ NO ₅	
26	Glycodeoxycholic acid (GDCA)	N-(3 α ,12 α -dihydroxy-5 β -cholan-24-oyl)-glycine	WVULKSPCQVQLC U-BUXLTGKBSA-N	C ₂₆ H ₄₃ NO ₅	

27	Lithocholic acid 3-Sulfate (LCA 3-SO ₄)	3 α -Sulfooxy-5 β -cholan-24-oic acid	AXDXVEYHEODSP N-HVATVPOCSA-N	C ₂₄ H ₃₉ O ₆ S ⁻	
28	Glycohyocholic acid (GHCA)	N-(3 α ,6 α ,7 α -trihydroxy-5 β -cholan-24-oyl)-glycine	ZQYUKJFJPJDMMR -ZDWCHQGWSA-N	C ₂₆ H ₄₃ NO ₆	
29	Glycocholic acid (GCA)	N-(3 α ,7 α ,12 α -trihydroxy-5 β -cholan-24-oyl)-glycine	RFDAIACWWDRED C-FRVQLJSFSA-N	C ₂₆ H ₄₃ NO ₆	
30	Taurolithocholic acid (TLCA)	N-(3 α , -hydroxy-5 β -cholan-24-oyl)-taurine	QBYUNVOYXHFK C-GBURMNQMSA-N	C ₂₆ H ₄₅ NO ₅ S	
31	Tauroursodeoxycholic acid (TUDCA)	N-(3 α ,7 β -dihydroxy-5 β -cholan-24-oyl)-taurine	BHTRKEVKTKCXO H-LBSADWJPSA-N	C ₂₆ H ₄₄ NO ₆ S	
32	Taurohyodeoxycholic acid (THDCA)	N-(3 α ,6 α -dihydroxy-5 β -cholan-24-oyl)-taurine	HMXPOCDLAFAN T-BHYUGXBJSA-N	C ₂₆ H ₄₅ NO ₆ S	
33	Taurochenodeoxycholic acid (TCDCA)	N-(3 α ,7 α -dihydroxy-5 β -cholan-24-oyl)-taurine	BHTRKEVKTKCXO H-BJLOMENOSA-N	C ₂₆ H ₄₅ NO ₆ S	
34	Taurodeoxycholic acid (TDCA)	N-(3 α ,12 α -dihydroxy-5 β -cholan-24-oyl)-taurine	AWDRATDZQPNJF N-VAYUFCLWSA-N	C ₂₆ H ₄₄ NO ₆ S	

35	Taurohyocholic acid (THCA)	N-(3 α ,6 α ,7 α -trihydroxy-5 β -cholan-24-oyl)-taurine	XSOLDPYUICCHJX-QZEPYOAJSA-N	C ₂₆ H ₄₅ NO ₇ S	
36	Taurocholic acid (TCA)	N-(3 α ,7 α ,12 α -trihydroxy-5 β -cholan-24-oyl)-taurine	WBWWGRHZICKQ-GZ-HZAMXZRMSA-N	C ₂₆ H ₄₅ NO ₇ S	

**3 β , 12 α -Dihydroxy-5 β -cholan-24-oic acid and 7 α , 12 α -Dihydroxy-5 β -cholan-24-oic acid are both referred to as isodeoxycholic acid in the literature[1, 6]; In the present study, 3 β , 12 α -Dihydroxy-5 β -cholan-24-oic acid was analyzed.*

Table S2. Classification of all measured BAs.

S. No	BA	Free	Taurine Conjugated	Glycine Conjugated	12 α -Hydroxylated	6 α -Hydroxylated
1	Chenodeoxycholic acid (CDCA)	X				
2	Glycochenodeoxycholic acid (GCDCA)			X		
3	Taurochenodeoxycholic acid (TCDCA)		X			
4	Cholic acid (CA)	X			X	
5	Glycocholic acid (GCA)			X	X	
6	Taurocholic acid (TCA)		X		X	
7	Hyocholic acid (HCA)	X				X
8	Glycohyocholic acid (GHCA)			X		X
9	Taurohyocholic acid (THCA)		X			X
10	5-Cholenic acid,3 β -ol (5-CholA)					
11	Lithocholic acid (LCA)	X				
12	3-Ketolithocholic acid (3-Keto-LCA)					
13	Glycolithocholic acid (GLCA)			X		
14	Taurolithocholic acid (TLCA)		X			
15	Lithocholic acid 3-Sulfate (LCA 3-SO ₄)					
16	6-Ketolithocholic acid (6-Keto-LCA)					
17	7-Ketolithocholic acid (7-Keto-LCA)					
18	Ursodeoxycholic acid (UDCA)	X				
19	Glycoursodeoxycholic acid (GUDCA)			X		
20	Tauroursodeoxycholic acid (TUDCA)		X			
21	Isolithocholic acid (Iso-LCA)					
22	Deoxycholic acid (DCA)	X			X	
23	Glycodeoxycholic acid (GDCA)			X	X	
24	Taurodeoxycholic acid (TDCA)		X		X	
25	23-Nordeoxycholic acid (23-Nor-DCA)				X	
26	7-Ketodeoxycholic acid (7-Keto-DCA)				X	
27	Ursocholic acid (UCA)	X			X	
28	12-Ketochenodeoxycholic acid (12-Keto-CDCA)					
29	12-Ketolithocholic acid (12-Keto-LCA)					
30	3-Ketodeoxycholic acid (3-Keto-DCA)				X	
31	3-Dehydrocholic acid (3-Dehydro-CA)				X	
32	Isodeoxycholic acid (Iso-DCA)				X	
33	Lithocholenic acid (LICA)					
34	Hyodeoxycholic acid (HDCA)	X				X
35	Glycohyodeoxycholic acid (GHDCA)			X		X
36	Taurohyodeoxycholic acid (THDCA)		X			X

Table S3. Plasma BA concentrations in subjects not on anti-diabetic medications (n=1,885) stratified by DM status.

Compounds	Subjects without diabetes Median (IQR), nM (n=1,670)	Subjects with Diabetes Median (IQR), nM (n=215)	P value	FDR corrected P value
Primary BAs and conjugated forms				
Chenodeoxycholic acid (CDCA)	68.2 (26.6 - 184.9)	73.7 (26.9 - 212.2)	0.508	0.653
Glycochenodeoxycholic acid (GCDCA)	311.6 (148.2 - 640.0)	378.5 (170.2 - 706.1)	0.040	0.103
Taurochenodeoxycholic acid (TCDCA)	41.0 (20.1 - 90.4)	54.0 (26.5 - 108.3)	0.006	0.027
Cholic acid (CA)*	37.2 (18.7 - 96.2)	41.5 (17.9 - 110.9)	0.821	0.953
Glycocholic acid (GCA)*	95.5 (47.3 - 204.1)	99.6 (49.5 - 224.8)	0.427	0.591
Taurocholic acid (TCA)*	12.6 (7.2 - 29.7)	15.1 (8.2 - 36.4)	0.034	0.094
Hyochoolic acid (HCA)**	9.5 (5.0 - 20.6)	8.7 (4.5 - 17.9)	0.102	0.216
Glycohychoolic acid (GHCA)**	5.7 (2.8 - 11.6)	5.4 (2.5 - 12.2)	0.705	0.875
Taurohychoolic acid (THCA)**	1.4 (0.6 - 3.1)	1.0 (0.5 - 2.5)	0.008	0.032
Secondary BAs derived from CA				
Deoxycholic acid (DCA)*	230.9 (113.8 - 408.1)	257.3 (128.2 - 374.4)	0.412	0.600
Glycodeoxycholic acid (GDCA)*	106.9 (46.7 - 244.4)	137.3 (60.9 - 307.4)	0.006	0.031
Taurodeoxycholic acid (TDCA)*	17.8 (8.1 - 42.3)	24.30 (10.3 - 61.5)	<0.001	0.007
23-Nordeoxycholic acid (23-Nor-DCA)*	1.9 (0.1 - 4.1)	2.1 (1.0 - 4.7)	0.028	0.084
7-Ketodeoxycholic acid (7-Keto-DCA)*	1.9 (0.8 - 6.0)	2.00 (1.1 - 5.3)	0.198	0.339
Ursocholic acid (UCA)*	0.8 (0.1 - 2.3)	0.60 (0.1 - 1.5)	0.101	0.227
12-Ketochenodeoxycholic acid (12-Keto-CDCA)	10.4 (5.8 - 17.2)	10.8 (6.8 - 17.3)	0.282	0.441
12-Ketolithocholic acid (12-Keto-LCA)	4.4 (1.7 - 9.6)	5.1 (2.5 - 10.4)	0.016	0.058
3-Ketodeoxycholic acid (3-Keto-DCA)*	48.9 (25.5 - 85.0)	48.9 (23.5 - 87.8)	0.929	0.956
3-Dehydrocholic acid (3-Dehydro-CA)*	0.1 (0.1 - 5.7)	0.1 (0.05 - 5.7)	0.379	0.570
Isoodeoxycholic acid (Iso-DCA)*	84.65 (43.7 - 161.2)	96.9 (44.7 - 161.5)	0.893	0.974
Lithocholenic acid (LICA)	1.1 (0.05 - 3.00)	1.5 (0.30 - 2.9)	0.022	0.072
Secondary BAs derived from CDCA				
5-Cholenic acid,3 β -ol (5-Chol-A)	7.9 (4.9 - 11.2)	7.8 (5.5 - 10.6)	0.946	0.948
Lithocholic acid (LCA)	28.2 (19.5 - 39.2)	25.6 (17.7 - 36.9)	0.061	0.146
3-Ketolithocholic acid (3-Keto-LCA)	1.4 (0.1 - 3.3)	1.2 (0.1 - 2.7)	0.478	0.637
Glycolithocholic acid (GLCA)	7.9 (4.2 - 17.3)	7.7 (4.1 - 17.5)	0.824	0.927
Taurolithocholic acid (TLCA)	4.0 (3.0 - 6.1)	4.3 (3.2 - 6.3)	0.166	0.299
Lithocholic acid 3-Sulfate (LCA 3-SO4)	7.5 (3.4 - 15.3)	6.7 (3.5 - 15.1)	0.772	0.928
7-Ketolithocholic acid (7-Keto-LCA)	12.7 (8.2 - 22.4)	10.6 (7.3 - 18.4)	0.005	0.030
6-Ketolithocholic acid (6-Keto-LCA)	0.7 (0.1 - 1.5)	1.0 (0.5 - 1.5)	<0.001	0.000
Ursodeoxycholic acid (UDCA)	19.8 (7.2 - 58.8)	19.3 (7.9 - 55.6)	0.961	0.961
Glycoursodeoxycholic acid (GUDCA)	55.9 (23.0 - 129.0)	58.00 (27.4 - 144.1)	0.130	0.260
Tauroursodeoxycholic acid (TUDCA)	3.5 (1.9 - 6.9)	3.90 (2.1 - 7.5)	0.133	0.252
Isolithocholic acid (Iso-LCA)	16.9 (8.8 - 30.3)	12.20 (5.6 - 24.7)	<0.001	0.000
Secondary BAs derived from HA				
Hyodeoxycholic acid (HDCA)**	0.3 (0.1 - 3.7)	0.05 (0.1 - 1.8)	<0.001	0.000
Glycohyodeoxycholic acid (GHDCA)**	8.2 (3.9 - 6.0)	8.6 (4.7 - 15.5)	0.222	0.363
Taurohyodeoxycholic acid (THDCA)**	0.7 (0.2 - 2.5)	0.5 (0.2 - 1.1)	<0.001	0.000

P-values were calculated using Wilcoxon rank-sum tests. $P < 0.05$ is considered statistically significant; FDR correction was performed for multiple hypothesis testing and FDR-corrected p -values are reported.

**12 α -Hydroxylated; **6 α -Hydroxylated.*

Table S4. Association of lead variants with levels of other BAs.

BA	rs77314 1	rs731879 45	rs731279 25	rs111835 596	rs679816 90	rs181813 865	rs38160 76	rs76270 623	rs53540 8158	rs11866 815
*Deoxycholic acid (DCA)	1.7E-08	0.99	3.0E-03	0.19	2.0E-03	0.64	0.44	0.26	0.92	0.11
Lithocholenic acid (LICA)	0.85	1.8E-08	0.29	0.55	0.67	NA	0.96	0.05	0.46	0.04
*Taurodeoxycholic acid (TDCA)	0.01	0.99	1.3E-08	4.4E-08	0.01	0.52	0.62	0.2	0.18	0.24
12-Ketolithocholic acid (12-Keto-LCA)	0.01	0.91	2.1E-03	0.06	2.2E-11	0.02	0.32	0.06	0.72	0.62
23-Nordeoxycholic acid (23-Nor-DCA)	9.1E-04	0.79	0.62	0.5	0.96	4.7E-08	0.15	0.33	0.17	0.8
*Glycodeoxycholic acid (GDCA)	1.4E-05	0.82	1.5E-08	7.5E-07	2.9E-03	0.23	0.67	0.31	0.66	0.12
7-Ketolithocholic acid (7-Keto-LCA)	0.18	0.72	0.02	0.92	0.78	0.8	0.83	0.08	0.24	0.74
Hyochoolic acid (HCA)	0.64	0.52	0.48	0.2	0.7	0.61	0.1	0.26	0.83	0.37
Taurohyochoolic acid (THCA)	0.38	0.3	0.6	0.09	0.38	0.69	0.43	0.3	0.04	0.83
Glycohyochoolic acid (GHCA)	0.17	0.12	0.01	4.4E-04	0.37	0.99	0.34	0.1	0.08	0.39
Hyodeoxycholic acid (HDCA)	0.47	N/A	0.79	0.38	0.87	0.21	0.83	0.01	0.02	0.42
Taurohyodeoxycholic acid (THDCA)	0.04	0.39	0.01	0.51	4.2E-08	0.02	0.53	0.19	0.03	0.74
6-Ketolithocholic acid (6-Keto-LCA)	0.17	N/A	N/A	0.93	0.47	NA	3.4E-08	2.9E-10	3.6E-08	0.9
Isolithocholic acid (Iso-LCA)	0.01	0.29	0.35	0.04	0.12	0.2	0.88	0.02	0.6	4.7E-08

P-values are shown for association of indicated SNP with all other BAs;

Bold indicates P-value for associations that were less than the Bonferroni-corrected threshold for testing 10 variants ($P=5.0E-03$);

**Results for DCA, TDCA, GDCA, and 7-Keto-LCA are based on normal inverse transformed data whereas remaining BAs were dichotomized prior to analysis;*

N/A, not available.

Table S5. Multi-tissue eQTLs at loci identified for BAs.

BA	SNP	EA/OA	eQTL Gene	Tissue	P	Direction	N	PMID
*Deoxycholic acid (DCA)	rs773141	A/G	<i>CARNMT1</i>	Brain - Cerebellum	5.4E-17	-	209	23715323
				Brain - Cerebellar Hemisphere	3.0E-11	-	175	23715323
				Muscle - Skeletal	3.0E-10	-	706	23715323
			<i>C9orf40</i>	Muscle - Skeletal	8.8E-08	-	706	23715323
				Colon - Transverse	1.5E-07	-	368	23715323
			<i>NMRK1</i>	Thyroid	4.2E-13	+	574	23715323
				Skin - Not Sun Exposed (Suprapubic)	1.8E-10	+	517	23715323
				Skin - Sun Exposed (Lower leg)	4.7E-10	+	605	23715323
				Whole Blood	4.8E-240	+	25,674	34475573
				Whole Blood	5.2E-09	+	670	23715323
				Whole Blood	1.0E-07	+	N/A	27540175
				Adipose - Subcutaneous	1.1E-08	+	581	23715323
				Lung	1.3E-08	+	515	23715323
				Artery - Tibial	8.1E-08	+	584	23715323
				Colon - Transverse	2.6E-07	+	368	23715323
				Adipose - Visceral (Omentum)	3.3E-07	+	469	23715323
				Esophagus - Mucosa	6.7E-07	+	497	23715323
				Esophagus - Muscularis	2.3E-06	+	465	23715323
				Liver	4.0E-06	+	N/A	27540175
Isolithocholic acid (Iso-LCA)	rs11866815	T/C	<i>ARHGDIG</i>	Stomach	1.7E-06	-	324	23715323
				Pancreas	3.5E-06	-	305	23715323
			<i>AXIN1</i>	Whole Blood	5.1E-24	-	27,399	34475573
				Testis	5.2E-11	+	322	23715323
			<i>DECR2</i>	Whole Blood	6.0E-37	-	27,399	34475573
			<i>HBM</i>	Whole Blood	1.3E-32	+	26,687	34475573
			<i>ITFG3</i>	Whole Blood	1.6E-18	-	27,399	34475573
			<i>METTL26</i>	Whole Blood	1.0E-06	-	27,399	34475573
			<i>MRPL28</i>	Esophagus - Mucosa	5.9E-13	+	497	23715323

	Skin - Not Sun Exposed (Suprapubic)	3.4E-11	+	517	23715323
	Skin - Sun Exposed (Lower leg)	1.8E-08	+	605	23715323
	Colon - Transverse	3.7E-06	+	368	23715323
<i>NME4</i>	Whole Blood	7.5E-154	-	27,185	34475573
	Artery - Tibial	3.5E-06	+	584	23715323
<i>PDIA2</i>	Nerve - Tibial	3.2E-06	-	532	23715323
<i>TMEM8A</i>	Whole Blood	3.9E-95	-	27,399	34475573
	Thyroid	1.1E-08	+	574	23715323
	Cells - Cultured fibroblasts	1.5E-08	+	483	23715323

Only *cis*-eQTLs with $P \leq 5.0E-06$ are listed.

Table S6. Association of loci identified for BA levels with DM and BMI.

BA	SNP	EA/OA	Association with BMI		Association with DM	
			Beta (SE)	P-value	OR (95% CI)	P-value
*Deoxycholic acid (DCA)	rs773141	A/G	0.009 (0.002)	3.2E-07	1.00 (0.99-1.01)	0.48
Lithocholenic acid (LICA)	rs73187945	G/A	0.004 (0.006)	0.47	1.02 (0.98-1.06)	0.34
*Taurodeoxycholic acid (TDCA)	rs73127925	C/T	0.016 (0.007)	0.02	0.99 (0.95-1.04)	0.75
*Taurodeoxycholic acid (TDCA)	rs111835596	A/G	N/A	N/A	N/A	N/A
12-Ketolithocholic acid (12-keto-LCA)	rs67981690	G/A	0.001 (0.003)	0.56	1.00 (0.99-1.02)	0.72
23-Nordeoxycholic acid (23-Nor-DCA)	rs181813865	A/G	N/A	N/A	N/A	N/A
*Glycodeoxycholic acid (GDCA)	rs73127925	C/T	0.016 (0.007)	0.02	0.99 (0.95-1.04)	0.75
6-Ketolithocholic acid (6-Keto-LCA)	rs3816076	C/T	-0.009 (0.006)	0.10	1.00 (0.98-1.02)	0.88
6-Ketolithocholic acid (6-Keto-LCA)	rs76270623	G/T	0.011 (0.008)	0.24	1.06 (1.03-1.09)	9.5E-06
6-Ketolithocholic acid (6-Keto-LCA)	rs535408158	CA/C	N/A	N/A	N/A	N/A
Isolithocholic acid (Iso-LCA)	rs11866815	T/C	-0.015 (0.002)	1.7E-11	0.97 (0.96-0.98)	5.3E-07
Taurohyodeoxycholic acid (THDCA)	rs67981690	G/A	0.001 (0.003)	0.56	1.00 (0.99-1.02)	0.72

Results are taken from previously published GWAS analyses for BMI [3] and DM [4].

Betas and odds ratios (OR) correspond to the BA-raising effect allele (EA).

Significant associations at the Bonferroni-corrected threshold for the number of variants tested ($p=0.05/10=0.005$) are shown in bold; OA, other allele; N/A, not available.

Table S8. Limit of Detection (LOD), Limit of Quantification (LOQ), Intra-day coefficients of variation, precision, and accuracy of BAs in human serum.

S. No	Compounds	LOD	LOQ	Intra-day coefficients of variation								
				QC Level 1			QC Level 2			QC Level 3		
				Mean (nM)	Precision (%CV)	Accuracy (%)	Mean (nM)	Precision (%CV)	Accuracy (%)	Mean (nM)	Precision (%CV)	Accuracy (%)
Primary BAs and conjugated forms												
1	Chenodeoxycholic acid (CDCA)	5.0	16.6	1381.5	3.7	12.8	391.5	4.8	1.8	804.3	4.3	0.4
2	Glycochenodeoxycholic acid (GCDCA)	0.6	2.1	1119.2	4.0	21.6	1023.6	5.7	1.7	3009.9	6.3	1.2
3	Taurochenodeoxycholic acid (TCDCA)	0.6	2.1	239.9	3.8	17.0	322.1	6.8	3.8	492.5	5.7	7.2
4	Cholic acid (CA)	2.1	7.0	917.3	3.4	11.1	298.0	5.2	2.1	335.2	3.8	9.4
5	Glycocholic acid (GCA)	0.9	3.0	395.9	4.8	17.7	679.3	6.2	2.5	1286.1	4.8	1.5
6	Taurocholic acid (TCA)	1.1	3.8	83.6	3.3	23.0	340.0	5.4	7.5	344.4	7.5	8.7
7	Hyochoolic acid (HCA)	3.6	11.9				142.3	4.8	3.5	239.5	4.7	5.0
8	Glycohyochoolic acid (GHCA)	1.5	4.9	31.8	5.3	21.6	129.6	6.6	5.3	259.6	6.2	9.5
9	Taurohyochoolic acid (THCA)	1.2	4.1	10.6	5.9	29.3	133.7	5.4	0.5	230.7	6.0	1.7
Secondary BAs derived from CA												
10	Deoxycholic acid (DCA)	4.6	15.3	941.9	3.9	7.0	619.0	5.7	5.2	896.8	5.3	2.4
11	Glycodeoxycholic acid (GDCA)	0.5	1.8	493.0	4.2	18.1	584.2	5.4	4.2	1106.6	6.8	4.8
12	Taurodeoxycholic acid (TDCA)	0.6	2.0	97.0	4.0	18.3	235.0	4.6	0.8	317.7	4.1	6.4
13	23-Nordeoxycholic acid (23-Nor-DCA)	2.6	8.7				123.4	5.9	3.5	218.7	5.1	11.4
14	7-Ketodeoxycholic acid (7-Keto-DCA)	3.4	11.3	36.0	5.0	9.1	124.5	4.9	4.6	266.3	4.2	8.3
15	Ursocholic acid (UCA)	3.3	11.0				120.2	12.7	0.7	250.1	5.4	1.9
16	12-Ketochenodeoxycholic acid (12-Keto-CDCA)	11.0	36.6				135.2	4.8	1.9	262.6	4.4	5.4
17	12-Ketolithocholic acid (12-Keto-LCA)	2.9	9.6				118.1	3.8	2.1	246.0	4.4	11.1
18	3-Ketodeoxycholic acid (3-Keto-DCA)	7.2	23.9	77.5	4.8	4.5	330.1	4.2	2.9	604.0	4.4	2.9
19	3-Dehydrocholic acid (3-Dehydro-CA)	13.0	43.2				255.4	4.8	5.4	1072.7	4.9	0.2
20	Isodeoxycholic acid (Iso-DCA)	2.6	8.6	132.1	3.6	11.8	272.7	5.1	3.7	322.2	5.9	0.5
21	Lithocholenic acid (LICA)	2.7	9.1				55.9	3.3	11.0	138.5	3.0	3.7
Secondary BAs derived from CDCA												
22	5-Cholenic acid-3 β -ol (5-Chol-A)	3.7	12.4				140.3	5.1	1.6	192.0	4.9	1.9
23	Lithocholic acid (LCA)	4.7	15.8	38.7	4.2	6.0	180.5	5.8	6.5	231.6	3.5	1.6
24	3-Ketolithocholic acid (3-Keto-LCA)	4.2	13.9				106.4	5.1	3.2	152.2	3.9	5.5
25	Glycolithocholic acid (GLCA)	0.7	2.2	27.0	5.2	24.9	139.9	6.8	12.2	224.2	6.7	8.3
26	Taurolithocholic acid (TLCA)	1.0	1.6				112.6	7.2	5.7	244.2	6.2	4.8

27	Lithocholic acid 3-sulfate (LCA 3-SO ₄)	1.0	2.0	13.5	4.1	6.0	102.5	6.4	13.1	228.4	3.9	7.8
28	6-Ketolithocholic acid (6-Keto-LCA)	2.6	8.6				103.8	5.5	5.5	194.5	4.7	6.6
29	7-Ketolithocholic acid (7-Keto-LCA)	2.2	7.5				300.0	5.2	3.4	226.4	5.3	5.2
30	Ursodeoxycholic acid (UDCA)	2.4	8.1	65.1	4.8	14.4	169.5	5.3	4.6	244.3	4.5	3.9
31	Glycoursodeoxycholic acid (GUDCA)	1.0	3.2	102.1	5.0	15.1	219.6	5.7	3.4	302.5	6.7	8.4
32	Tauroursodeoxycholic acid (TUDCA)	1.3	4.2	7.2	5.0	2.5	117.0	5.3	3.0	256.1	6.1	13.5
33	Isolithocholic acid (Iso-LCA)	2.9	9.8	33.8	4.8	14.6	166.0	5.4	6.0	303.9	5.8	13.8
Secondary BAs derived from HA												
34	Hyodeoxycholic acid (HDCA)	4.7	15.7				97.1	4.9	4.5	202.6	3.6	5.6
35	Glycohyodeoxycholic acid (GHDCA)	1.2	4.1	15.9	4.5	5.7	125.6	4.9	3.3	239.2	4.9	6.5
36	Taurohyodeoxycholic acid (THDCA)	1.1	3.6				119.0	4.5	1.7	256.6	4.8	5.8
Other BAs												
37	Ursocholic acid	7.5	25.1				58.6	5.6	2.0	150.4	4.5	13.4
38	9(11),(5 β)-Cholenic acid-3 α -ol-12-one	3.4	11.4				113.7	4.5	6.2	231.1	4.5	6.2
39	3,7-Diketocholic acid	3.4	11.2				148.4	4.9	0.9	258.8	5.0	10.5
40	Apocholeic acid	1.8	5.9				119.8	4.7	2.9	225.3	5.3	7.1
41	Dehydrocholic acid	9.4	31.2				129.2	3.5	3.1	265.7	3.7	6.1
42	7,12-Diketolithocholic acid	5.9	19.8				128.0	4.7	6.4	276.8	5.4	6.7
43	6,7-Diketolithocholic acid	3.8	12.6				120.8	5.1	1.3	217.6	5.5	1.0
44	Glycodehydrocholic acid	5.9	19.5				132.2	5.8	4.9	262.9	5.1	10.6
45	Taurodehydrocholic acid	5.2	17.5				106.1	5.8	7.7	239.9	4.3	5.7

LOD = limit of detection; LOQ = limit of quantification; CV = coefficient of variation; Intra-day precision and accuracy were examined on three different QC human pooled serums [1]. Blank areas indicate metabolites in a given QC pool whose level was below the LOD;

The limit of detection (LOD) was defined as the lowest concentration of analyte in the sample matrix (e.g. serum) that generated a signal-to-noise ratio of ≥ 3 [1]. The limit of quantification (LOQ) was defined as the lowest concentration of analyte in the sample matrix that generated a signal-to-noise ratio of ≥ 10 [1]; The intra-day precision was calculated by injecting six analytical replicates of the three QCs in a single day and measuring the intra-day coefficients of variation (CVs). To quantify the precision for each detectable BA in human serum, 36 runs of three different pooled serum QC levels were performed, and a relative standard deviation was calculated.[1].

Intra-day accuracy was measured in triplicate for the three QC levels (acceptable concentration) using a standard addition method and compared to values obtained from six analytical replicates of the three QCs analyzed the same day using the methods calibration curve (experimental concentration) [1]. Accuracy was determined as a percentage using the following formula: (accepted concentration- experimental concentration)/accepted concentration) x100 = % accuracy [1].

Table S9. Inter-day coefficients of variation, precision, and accuracy of BAs in human serum.

S. No	Compounds	Inter-Day coefficients of variation								
		QC Level 1			QC Level 2			QC Level 3		
		Mean (nM)	Precision (%CV)	Accuracy (%)	Mean (nM)	Precision (%CV)	Accuracy (%)	Mean (nM)	Precision (%CV)	Accuracy (%)
Primary BAs										
1	Chenodeoxycholic acid (CDCA)	1358.7	3.9	11.0	378.0	5.3	1.7	831.8	5.4	3.8
2	Glycochenodeoxycholic acid (GCDCA)	1122.2	4.4	9.7	999.5	6.4	0.7	3133.8	7.4	5.4
3	Taurochenodeoxycholic acid (TCDCA)	241.2	3.9	12.3	312.8	7.0	0.8	501.4	6.2	9.2
4	Cholic acid (CA)	926.2	3.7	12.2	294.8	5.9	0.9	339.6	4.7	10.8
5	Glycocholic acid (GCA)	400.3	5.0	14.2	669.1	7.5	3.9	1327.8	5.8	4.8
6	Taurocholic acid (TCA)	84.2	3.7	14.7	332.1	6.4	5.0	355.6	8.2	12.3
7	Hyochoolic acid (HCA)				137.8	5.3	0.2	242.3	6.1	6.3
8	Glychohyocholic acid (GHCA)	32.3	5.6	15.2	129.9	7.8	5.1	260.7	6.8	10.0
9	Taurohyocholic acid (THCA)	10.0	7.4	9.9	130.7	8.0	1.7	240.6	8.4	2.5
Secondary BAs derived from CA										
10	Deoxycholic acid (DCA)	947.7	4.2	7.7	612.5	6.3	4.1	936.7	6.9	1.9
11	Glycodeoxycholic acid (GDCA)	490.2	4.6	8.1	556.0	6.9	8.8	1137.3	7.6	7.7
12	Taurodeoxycholic acid (TDCA)	99.3	4.6	12.9	229.3	4.8	3.3	322.2	4.5	7.9
13	23-Nordeoxycholic acid (23-Nor-DCA)				123.8	6.5	3.2	227.9	6.6	16.1
14	7-Ketodeoxycholic acid (7-Keto-DCA)	35.7	5.6	8.1	123.7	4.7	5.2	273.6	4.6	11.3
15	Ursocholic acid (UCA)				116.1	17.9	4.1	267.5	7.1	9.0
16	12-Ketochenodeoxycholic acid (12-Keto-CDCA)				129.7	5.4	5.9	269.6	4.7	2.9
17	12-Ketolithocholic acid (12-Keto-LCA)				118.1	4.2	2.1	253.4	6.0	14.5
18	3-Ketodeoxycholic acid (3-Keto-DCA)	77.7	5.3	4.8	321.6	4.5	5.4	655.5	6.8	5.4
19	3-Dehydrocholic acid (3-Dehydro-CA)				252.3	4.8	6.5	1084.4	6.2	0.9
20	Iso deoxycholic acid (Iso-DCA)	135.1	4.5	14.3	272.9	5.7	3.8	337.5	6.6	4.2
21	Lithocholenic acid (LICA)				53.7	7.3	14.4	134.0	8.6	6.8
Secondary BAs derived from CDCA										
22	5-Cholenic acid-3 β -ol (5-Chol-A)				131.8	7.4	4.5	199.7	6.8	6.0
23	Lithocholic acid (LCA)	39.1	4.7	7.1	180.4	5.7	6.4	233.4	4.4	2.4
24	3-Ketolithocholic acid (3-Keto-LCA)				103.5	5.8	5.9	149.9	5.2	6.9
25	Glycolithocholic acid (GLCA)	26.3	5.1	11.1	131.3	7.9	5.3	233.2	8.1	12.6
26	Taurolithocholic acid (TLCA)				108.0	7.8	9.5	252.7	6.9	8.5

27	Lithocholic acid 3-sulfate (LCA 3-SO ₄)	13.8	4.6	8.2	104.0	6.7	11.9	236.1	4.8	11.4
28	6-Ketolithocholic acid (6-Keto-LCA)				102.1	5.5	7.1	199.1	5.7	9.1
29	7-Ketolithocholic acid (7-Keto-LCA)				292.9	6.6	0.9	234.9	6.3	9.2
30	Ursodeoxycholic acid (UDCA)	65.4	4.9	14.9	168.7	5.3	4.1	249.0	5.3	5.8
31	Glyoursodeoxycholic acid (GUDCA)	106.0	5.5	10.9	214.8	6.3	1.1	311.1	7.5	11.5
32	Tauroursodeoxycholic acid (TUDCA)	7.4	5.9	0.7	113.9	6.2	5.5	261.5	7.5	15.9
33	Isolithocholic acid (Iso-LCA)	33.2	5.7	12.7	165.6	6.2	5.8	312.1	6.7	16.8
Secondary BAs derived from HA										
34	Hyodeoxycholic acid (HDCA)				95.5	4.9	6.0	211.8	4.9	1.3
35	Glycohyodeoxycholic acid (GHDCA)	15.9	4.6	5.8	124.8	5.1	3.9	247.9	6.2	10.5
36	Taurohyodeoxycholic acid (THDCA)				114.6	5.6	2.0	261.4	6.5	7.8
Other BAs										
37	Ursocholic acid				57.7	6.6	3.4	152.4	5.3	12.3
38	9(11),(5β)-Cholenic acid-3α-ol-12-one				115.0	5.7	5.1	244.0	6.1	12.1
39	3,7-Diketocholic acid				145.2	5.9	1.2	268.0	5.5	14.4
40	Apocholic acid				117.8	4.8	4.6	232.4	6.0	10.4
41	Dehydrocholic acid				122.9	4.6	7.8	269.0	5.2	7.4
42	7,12-Diketolithocholic acid				128.8	6.3	5.8	283.3	6.4	9.1
43	6,7-Diketolithocholic acid				119.2	5.1	2.6	227.9	6.3	5.7
44	Glycodehydrocholic acid				131.8	6.9	5.2	266.9	5.9	12.3
45	Taurodehydrocholic acid				103.9	6.9	9.6	255.0	6.0	12.3

CV = coefficient of variation; Inter-day precision and accuracy were examined on three different QC human pooled serums [1]. Blank areas indicate metabolites in a given QC pool whose level was below the LOD;

The inter-day precision was calculated by injecting six analytical replicates of the three QCs in a single day over a span of six different days and calculating the intra-day coefficients of variation (CVs). To quantify the precision for each detectable BA in human serum, 36 runs of three different pooled serum QC levels were performed, and relative standard deviation was calculated [1].

Inter-day accuracy was measured in triplicate for the three QC levels (acceptable concentration) using a standard addition method and compared to values obtained from six analytical replicates of the three QCs analyzed over six days using the methods calibration curve (experimental concentration)[1]. Accuracy was determined as a percentage using the following formula: (accepted concentration- experimental concentration)/accepted concentration) x100 = % accuracy [1].

4. References

1. Choucair, I., et al., *Quantification of bile acids: a mass spectrometry platform for studying gut microbe connection to metabolic diseases*. J Lipid Res, 2020. **61**(2): p. 159-177.
2. Chang, C.C., et al., *Second-generation PLINK: rising to the challenge of larger and richer datasets*. Gigascience, 2015. **4**: p. 7.
3. Loh, P.R., et al., *Mixed-model association for biobank-scale datasets*. Nat Genet, 2018. **50**(7): p. 906-908.
4. Mahajan, A., et al., *Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation*. Nat Genet, 2022. **54**(5): p. 560-572.
5. Hemani, G., et al., *The MR-Base platform supports systematic causal inference across the human phenome*. Elife, 2018. **7**.
6. Jover, A., et al., *Successful prediction of the hydrogen bond network of the 3-oxo-12 α -hydroxy-5 β -cholan-24-oic acid crystal from resolution of the crystal structure of deoxycholic acid and its three 3,12-dihydroxy epimers*. Steroids, 2004. **69**(6): p. 379-88.