Plasma metabolomics and incidence of atrial fibrillation: the Atherosclerosis Risk in Communities (ARIC) Study

Alvaro Alonso, MD, PhD,[1] Bing Yu, PhD,[2] Yan V. Sun, PhD,[1] Lin Y. Chen, MD, MS,[3] Laura R. Loehr, MD, PhD,[4] Wesley T. O'Neal, MD, MPH,[5] Elsayed Z. Soliman, MD, MSc, MS,[6] Eric Boerwinkle, PhD[2,7]

[1]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA. [2]University of Texas Health Sciences Center, Houston, TX. [3]Division of Cardiology, University of Minnesota Medical School, Minneapolis, MN. [4]University of North Carolina, Chapel Hill, NC. [5]Division of Cardiology, School of Medicine, Emory University, Atlanta, GA. [6]Department of Epidemiology and Prevention, Wake Forest School of Medicine, Winston-Salem, NC. [7]Human Genome Sequence Center, Baylor College of Medicine, Houston, TX

Corresponding author:

Alvaro Alonso, MD, PhD. Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Rd NE, Atlanta, GA 30322. Phone: +1 404 727 8714. Email: alvaro.alonso@emory.edu

Target journal: Journal of the American Heart Association

Disclosures: No relevant conflicts of interest

ABSTRACT

Background: Previously, we identified associations of two circulating secondary bile acids (glycocholenate and glycolithocolate sulfate) with atrial fibrillation (AF) risk among African Americans. We aimed to replicate these findings in an independent sample including both whites and African Americans, and performed a new metabolomic analysis in the combined sample.

Methods: We studied 3,922 participants from the ARIC cohort followed between 1987 and 2013. Of these, 1,919 had been included in the prior analysis and 2,003 were new samples. Metabolomic profiling was done in baseline serum samples using gas and liquid chromatography mass spectrometry. AF was ascertained from electrocardiograms, hospitalizations, and death certificates. We used multivariable Cox regression to estimate hazard ratios (HR) and 95% confidence intervals (95%CI) of AF by one standard deviation difference of metabolite levels.

Results: Over a mean follow-up of 20 years, 608 participants developed AF. Glycocholenate sulfate was associated with AF in the replication and combined samples (HR 1.10, 95%CI 1.00, 1.21 and HR 1.13, 95%CI 1.04, 1.22, respectively). Glycolithocolate sulfate was not related to AF risk in the replication sample (HR 1.02, 95%CI 0.92, 1.13). An analysis of 245 metabolites in the combined cohort identified three additional metabolites associated with AF after multiple-comparison correction: pseudouridine (HR 1.18, 95%CI 1.10, 1.28), uridine (HR 0.86, 95%CI 0.79, 0.93) and acisoga (HR 1.17, 95%CI 1.09, 1.26).

Conclusion: We replicated a prospective association between a previously identified secondary bile acid, glycocholenate sulfate, and AF incidence, and identified new metabolites involved in nucleoside and polyamine metabolism as markers of AF risk.

INTRODUCTION

Atrial fibrillation (AF), a common cardiac arrhythmia, is a major risk factor for stroke and other cardiovascular diseases.[1] Application of metabolomics, the systematic investigation of all small molecules in a biological system, to the study of AF risk could deepen our understanding of AF pathogenic pathways as well as contribute to the discovery of novel disease biomarkers.[2] To date, however, metabolomic studies in this area have been few and limited in sample size. In an analysis of metabolomic data from 1,919 African-American participants in the community-based Atherosclerosis Risk in Communities (ARIC) study, including 183 who were newly diagnosed with AF, we reported an association of higher circulating levels of two secondary bile acids, glycolithocholate sulfate and glycocholenate sulfate, with incidence of AF, but no replication in independent cohorts was available.[3] More recently, a report from the mostly European-American Framingham Heart Study including 2,458 participants with targeted metabolomic profiling, of which 156 developed AF, did not identify any molecule significantly associated with AF incidence after adjustment for multiple comparisons.[4] Additional studies are required to replicate previous findings and increase statistical power for novel discoveries.

In this manuscript, as a follow-up to our previous study in the ARIC cohort, we extend the metabolomics assessment to 2,003 additional ARIC participants. We aimed to replicate the findings from the prior ARIC analysis in the additional ARIC participants and to conduct a new hypothesis-generating analysis in the combined sample of 3,922 participants.

METHODS

Study population

In 1987-89, the ARIC study examined 15,792 men and women 45-64 years of age recruited from four communities in the United States (Forsyth County, NC; Jackson, MS; Minneapolis suburbs, MN; Washington County, MD).[5] Participants were mostly white in the Minneapolis and Washington County sites, white and African American in Forsyth County, while only African Americans were recruited in Jackson. After their baseline exam, participants underwent follow-up visits in 1990-92, 1993-95, 1996-98, 2011-13, and 2016-17. Participants have been followed up via annual phone calls (semiannual since 2012). For the current analysis, we included 3,922 participants with available metabolomics data and without evidence of AF at baseline.

Metabolomic profiling

As previously described, 1,977 randomly selected African Americans in the Jackson field center had serum metabolomic profiling performed in 2010 in samples obtained at study baseline in 1987-89.[6] The samples had been stored at -80°C and were assayed with an untargeted, gas chromatography/mass spectrometry and liquid chromatography/mass spectrometry-based metabolomic quantification protocol by Metabolon, Inc. (Durham, NC). Similarly in 2014, serum samples from an additional 2,055 randomly selected participants (76% white, 24% African-American) collected in 1987-89 and stored since then at -80°C were assayed by Metabolon, Inc. using the same protocol. We selected a set of 97 samples to measure their metabolome profiles using baseline serum samples at both 2010 and 2014. We calculated the Pearson correlation coefficients (r) between the 97 pairs for shared metabolites. For the present study, we limited the analysis to metabolites with: 1) no more than 25% missing values, and 2) Pearson correlation coefficients ≥ 0.3 between 2010 and 2014 measurements. After applying these criteria, 245 named metabolites were included.

Ascertainment of atrial fibrillation

We have described elsewhere the details about AF ascertainment in the ARIC cohort.[7] Briefly, we identified AF cases through the end of 2013 from three sources: electrocardiograms (ECG) done at scheduled study visits, discharge diagnosis codes from hospitalizations, and death certificates. At all study visits, participants underwent a standard 12-lead 10-second ECG, which was transmitted electronically to the ARIC ECG reading center at EPICARE (Wake Forest School of Medicine, Winston-Salem, NC) for review and analysis using the GE Marquette 12-SL program (GE Marquette, Milwaukee, WI). A computer algorithm identified the presence of AF in the ECG, with a cardiologist confirming the diagnosis.

Participants' hospitalizations during follow-up were identified through phone calls and surveillance of local hospitals. Trained abstractors collected information from these hospitalizations, including all discharge codes. We considered AF present if ICD-9-CM codes 427.31 or 427.32 were listed as discharge diagnoses in any given hospitalization. We excluded AF cases associated with open cardiac surgery. We and others have demonstrated adequate validity of this approach for the ascertainment of AF.[7, 8] Finally, we also defined AF from death certificates if ICD-9 427.3 or ICD-10 I48 were listed as any cause of death.

Covariates

During the baseline visit, participants self-reported age, sex, race, and smoking history and underwent a physical exam that included measurements of blood pressure, weight, and height. Blood glucose and lipid concentrations were measured using standard methods in baseline samples. Estimated glomerular filtration rate (eGFR) was calculated from serum creatinine using the CKD-EPI equation.[9] Diabetes was defined if the participant had fasting blood glucose ≥126 mg/dL, non-fasting blood glucose >200 mg/dL, used antidiabetic medications, or reported a physician-diagnosis of diabetes. Prevalent heart failure was defined according to Gothenburg criteria,[10] while prevalent coronary heart disease was based on self-reported information. Also at baseline, participants underwent a standard 12-lead 10-second electrocardiogram, which was processed at EPICARE (Wake Forest School of Medicine, Winston-Salem, NC). PR duration, P wave axis and P wave terminal force in V1 were all automatically measured. Abnormal P wave axis was defined as any P wave axis value outside 0 to 75 degress, while elevated P wave terminal force in V1 was defined if P wave terminal force was >4,000 μV*ms. Genome-wide and exome genotyping of ARIC participants has been done using the Affymetrix 6.0 and the Illumina HumanExome Beadchip v1.0, as described elsewhere.[11]

Statistical analysis

We conducted two separate sets of analyses. In the first one, we aimed to replicate the findings from our prior ARIC publication, estimating the association of glycolithocholate sulfate and glycocholenate sulfate with AF incidence in 2,003 participants without AF at baseline not included in our published analysis. A second analysis combined participants from the two metabolomic assessment batches (n = 3,922). We used a modified Bonferroni correction to determine statistical significance.[12] Using this approach, p-values less than 3.538×10^{-4} were considered statistically significant for 245 tested metabolites.

For all analyses, the association of individual metabolites with the incidence of AF was estimated with Cox proportional hazards regression. Time of follow-up was defined as the time in days from the baseline visit to incidence of AF, death, loss to follow-up or December 31, 2013, whichever occurred earlier. Metabolites were mean centered and modeled as continuous variables in standard deviation units. Missing values were imputed with the lowest detected value in each batch. We ran three separate models with increasing number of covariates. A first model adjusted for age, sex, race, center, and batch (when applicable). A second model additionally adjusted for smoking, body mass index, systolic blood pressure, hypertension medications, diabetes mellitus, history of heart failure, and history of coronary heart disease. A final model additionally adjusted for eGFR. We selected model covariates based on

prior knowledge of risk factors for AF.[13] We assessed effect measure modification by race and sex using stratified analysis. The dose-response shape of the association between metabolite concentration and AF incidence was evaluated modeling metabolites using a restricted cubic spline with five knots. To test the robustness of the observed significant associations, we conducted a series of sensitivity analyses, adjusting for blood lipids and lipid-lowering medications and excluding participants with a prior history of prevalent coronary heart disease or heart failure, as well as adjusting for aspartate aminotransferase (AST) and alanine aminotransferase (ALT), measured in visit 2 samples, in the analyses of bile acids.

We conducted several additional analyses to explore potential mechanisms of the association between metabolites and AF incidence. First, we evaluated the association of statistically significant metabolites with electrocardiographic endophenotypes of AF risk using linear regression (PR duration, in ms) or logistic regression (abnormal P wave axis and elevated P wave terminal force in V1). Second, we evaluated the association of statistically significant metabolites with 23 single nucleotide polymorphisms (SNPs) associated with AF in a prior genome-wide association study (GWAS) from the AFGen consortium, and a genetic score calculated by adding the number of risk alleles weighted by the beta coefficient from the published genome-wide study.[11] Finally, we explored whether variation in rs2272996 in gene *VNN1*, a SNP previously related to circulating concentrations of acisoga (one of the metabolites associated with AF incidence in this analysis),[14] was associated with AF incidence in the latest GWAS of AF.

RESULTS

Of 15,792 participants in the ARIC cohort, the present analysis included 3,922 with available metabolomic data and free of AF at baseline, 1,919 of them included in our previous publication and 2,003 with newly available data. Participants were followed up for a mean (standard deviation) of 20.4 (7.0) years, during which 608 AF events were identified (incidence rate, 7.6 cases per 1,000 person-years). Table 1 reports participants' characteristics overall and by AF incidence status during follow-up. As expected, participants who developed AF during follow-up were older, had higher systolic blood pressure and worse kidney function at baseline. They were also more likely to be white, male and have a baseline diagnosis of diabetes, heart failure or coronary heart disease.

In an initial analysis, we aimed to replicate the findings from our previous publication showing that higher levels of glycolithocholate sulfate and glycocholenate sulfate were associated with increased risk of AF. In an age and sex-adjusted analysis including 2,003 participants and 386 incident AF events, higher levels of glycocholenate sulfate but not of glycolithocholate sulfate were associated with AF

incidence in the replication analysis (Table 2, Model 1). The association of glycocholenate sulfate with incidence of AF became weaker after multivariable adjustment (HR 1.10, 95%CI 1.00, 1.21 per 1-SD difference; Table 2, Model 2). Given the strong attenuation after multivariable adjustment, we explored if any individual covariate was responsible for this change. Adding each covariate to Model 1 individually did not point to any particular variable as responsible for the attenuation (Supplementary Figure 1). The hazard ratio (HR) and 95% confidence interval (CI) of AF per 1-standard deviation (SD) difference in glycocholenate sulfate in the combined derivation and replication samples was 1.23 (95%CI 1.14-1.32, p = 9.5×10^{-8}) in minimally adjusted models and 1.13 (95%CI 1.04, 1.22, p = 0.003) after additional adjustment for cardiovascular risk factors. Additional adjustment for concentrations of ALT and AST in 3,401 participants with available information on liver enzymes did not modify the associations (HR 1.15, 95%CI 1.07, 1.23, p = 2.5×10^{-5}). Analysis stratified by race and sex showed a weaker association between glycolithocholate sulfate and AF in whites compared to African Americans (HR 1.04, 95%CI 0.94, 1.16 versus HR 1.19, 95%CI 1.10, 1.28, p for interaction = 0.05). No other interactions were identified (Supplementary Figures 2 and 3).

Subsequently, we performed a metabolome-wide, hypothesis-free analysis combining the two study samples. Of the 245 studied metabolites, 9 were associated with the incidence of AF with p-values <0.001 after multivariable adjustment (Table 3, Model 2). These metabolites included molecules involved in the metabolism of pyrimidines (pseudouridine and uridine), polyamines (acisoga), amino acids (N-acetylalanine and N-acetylthreonine), and bile acids (glycoursodeoxycholate and glycochenodeoxycholate), as well as one lisolipid (1-docosahexaenoylglycerophosphocholine), and a xenobiotic (O-sulfo-L-tyrosine). Three of these molecules, pseudouridine, acisoga, and uridine, were significantly associated with AF with p-values < 3.538×10^{-4}. Specifically, higher levels of pseudouridine and acisoga were associated with higher rates of AF (HR 1.18, 95%CI 1.10, 1.28 and 1.17, 95%CI 1.09, 1.26, respectively) while higher uridine levels were associated with reduced AF rates (HR 0.86, 95%CI 0.79, 0.93). Complete results for the 245 metabolites are available as a supplementary file. The correlation matrix of the 9 metabolites is shown in Supplementary Table 1. Uridine was not correlated with pseudouridine (r = -0.02) or acisoga (r = -0.03), though there was a modest association between pseudouridine and acisoga (r = 0.42). Associations for pseudouridine and acisoga weakened, but were still present, in a model including the 3 metabolites simultaneously (HR 1.16, 95%CI 1.06, 1.26 for pseudouridine, HR 1.11, 95%CI 1.02, 1.20 for acisoga). The inverse association between uridine and AF risk did not change after adjustment for pseudouridine and acisoga (HR 0.85, 95%CI 0.79, 0.92). The association remained essentially unchanged after adjustment for blood lipids and in those without CVD

(Supplementary Table 2). Figure 1 presents the dose-response associations of pseudouridine, acisoga, and uridine with AF risk, which were approximately linear for the three molecules. Multivariable adjustment led to meaningful attenuation in the association of pseudouridine with AF. None of the individual covariates in the multivariable model seemed particularly responsible for this attenuation, as evaluated by adding each covariate individually to the minimally adjusted model (Supplementary Figure 1). Associations were similar across race and sex groups (Supplementary Figures 2 and 3).

To characterize in more detail the association of the three metabolites with AF, we explored their cross-sectional association with selected intermediate phenotypes of AF (PR interval, elevated P wave terminal force in V1, abnormal P wave axis) (Table 4). None of the three metabolites were associated with the odds of abnormal P wave axis or elevated P wave terminal force in V1. The results were suggestive of a possible association of higher pseudouridine and acisoga with shorter PR interval [beta (95% CI), -0.9 ms (-1.9, 0.1), and -0.9 ms (-1.8, -0.1), respectively) and higher uridine with longer PR interval [0.6 ms (-0.2, 1.4)].

We assessed whether any of the AF-related genetic variants identified in a recent GWAS of AF among individuals of European ancestry were associated with levels of pseudouridine, acisoga or uridine among white participants with genomic data (N = 1421). In this analysis, neither the individual genetic variants nor the AFGen genetic risk score predicted serum levels of these three metabolites (Supplementary Table 3).

Finally, variation in rs2272996 in gene *VNN1*, previously associated with circulating levels of acisoga, was not predictive of AF risk (p = 0.88 in the most recent GWAS from the AFGen consortium).

DISCUSSION

In this metabolomic study of 3,922 men and women from a diverse prospective cohort we replicated a previously described association of glycocholenate sulfate, a secondary bile acid, with the incidence of AF. Also, we identified three additional metabolites (two related to pyrimidine metabolism, pseudouridine and uridine, and one related to polyamine metabolism, acisoga) associated with incidence of AF. Several additional analyses showing lack of association of these metabolites with AF electrical endophenotypes and AF-related genes suggest that these metabolites may affect AF pathogenesis through alternative mechanisms.

Bile acids and AF

Consistent with our prior analysis of the ARIC cohort,[3] we found an association of circulating glychocholenate sulfate with increased incidence of AF. The previously described association of another secondary bile acid, glycholithocholate sulfate, with AF was not replicated in this new analysis. In addition, we identified two additional secondary bile acids, glycoursodeoxycholate and glycochenodeoxycholate, associated with AF incidence, though these associations did not achieve the multiple comparison-corrected threshold for statistical significance. Glychocholenate sulfate is possibly derived from 3-beta-hydroxy-5-cholenoic acid (cholenate). Prior literature has described elevations of cholenate in patients with liver disease,[15] while both glycoursodeoxycholate and glycochenodeoxycholate are elevated in patients with liver cirrhosis.[16] Thus, liver injury, which has been associated with AF previously, could explain the association of bile acids with incident AF. However, adjustment for biomarkers of liver damage (ALT and AST) did not materially change the associations. Alternative mechanisms, including the cardiometabolic implications of systemic activation of farnesoid X receptor by circulating bile acids[17] or changes in the gut microbiota,[18] instrumental in bile acid metabolism, could underlie the described associations. Our results, together with a prior study describing potential arrhythmogenic effects of bile acids,[19] provide the rationale for future work exploring the impact of bile acids on the development of AF.

Pseudouridine and uridine

Pseudouridine and uridine are nucleosides involved in RNA synthesis and metabolism. Pseudouridine results from enzymatic posttranscriptional modification of uridine in RNA, with stress conditions influencing the occurrence of this process.[20] In turn, RNA pseudouridylation can affect gene expression regulation through mRNA stability and proteome diversity.[21] Because of its physiological roles, circulating or urinary pseudouridine is considered a marker of RNA degradation and cell turnover.[22] Prior studies have reported higher concentrations of circulating pseudouridine in patients with pulmonary arterial hypertension,[23] heart failure,[24] impaired kidney function,[25] end-stage renal disease,[26] and cancer.[27] The relationships between circulating pseudouridine and posttranscriptional pseudouridylation of RNA and what role, if any, pseudouridine has in processes contributing to AF risk, requires further investigation.

Uridine is a ribonucleoside potentially involved in modulation of the metabolism of multiple systems and critical for cellular function and survival, though its specific targets have not been identified.[28] Recent studies indicate that plasma uridine plays a key role in energy homeostasis and thermoregulation, modulating leptin signaling and potentially affecting glucose and insulin

metabolism.[29] Given the involvement of obesity and diabetes in the development of AF, deeper understanding of the physiological role of uridine in cardiometabolic disorders is needed. In fact, prior epidemiologic and clinical evidence has shown beneficial associations with higher plasma uridine, with higher levels of uridine associated with reduced mortality in the ARIC cohort,[30] and reduced pulse wave velocity in the Twins UK Registry.[31] In the Framingham Heart Study, higher concentrations of uridine were associated with a nonsignificant lower risk of AF (HR 0.84, 95%CI 0.70, 1.00, p = 0.05, per 1-standard deviation higher concentrations).[4]

Acisoga

Acisoga (N-(3-acetamidopropyl)pyrrolidin-2-one) is a catabolic product of spermidine formed from N1-acetylspermidine, and involved in the metabolism of polyamines.[32] Its precise role is unknown, but two prior studies have found associations of elevated acisoga concentrations with higher body mass index,[33, 34] and a potential association with the incidence of diabetes mellitus in the ARIC study.[35] Concentrations of acisoga were part of a metabolomic-score predicting mortality in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention study cohort.[36] Polyamines are key players in a range of processes, including cell-cell interactions, cellular signaling, and ion channel regulation.[37] Acisoga, as an end product of polyamine metabolism, may be a marker of dysregulation in this pathway.

Strengths and limitations

Our study has important strengths, including the inclusion of a large and diverse cohort with excellent follow-up, an adequate number of AF cases to identify associations, and the availability of extensive covariates to reduce confounding. Moreover, we have considered only metabolites that passed rigorous quality control criteria. However, the method of AF ascertainment—relying predominantly on hospital discharge diagnoses—has probably led to missed events, including asymptomatic AF and AF managed exclusively in outpatient settings. Other limitations include the risk of false negatives, due to the limited number of events, and the absence of an independent sample for replication.

Future directions

Our findings identify potential fruitful avenues of research. Additional studies that aim to evaluate the role played by the metabolism of bile acids, uridine and polyamines in processes leading to AF are warranted. Replicating findings from the ARIC cohort in independent samples is also needed. Combining metabolomic data with those coming from other omic levels (genomics, transcriptomics, and

proteomics) and exploring associations with intermediate phenotypes of AF (e.g. left atrial abnormalities) could be particularly rewarding.

Conclusions

This study identified several molecules involved in a range of metabolic pathways associated with the incidence of AF. Our findings demonstrate the value that metabolomic approaches in large epidemiologic studies has for biomarker discovery and advancing our understanding of the pathogenesis of complex diseases.

ACKNOWLEDGEMENTS

The authors thank the staff and participants of the ARIC study for their important contributions.

FUNDING

The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. (HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700005I, HHSN268201700004I). The metabolomics research was sponsored by the National Human Genome Research Institute (3U01HG004402-02S1). This work was additionally supported by American Heart Association grant 16EIA26410001 (Alonso). Dr. Yu is supported in part by American Heart Association (17SDG33661228) and the National Heart, Lung, and Blood Institute (HL141824 and HL142003).

REFERENCES

1. Benjamin EJ, Blaha MJ, Chiuve SE, et al. Heart disease and stroke statistics—2017 update: a report from the American Heart Association. *Circulation.* 2017;135:e146-e603.
2. Ussher JR, Elmariah S, Gerszten R E, Dyck RF. The emerging role of metabolomics in the diagnosis and prognosis of cardiovascular disease. *J Am Coll Cardiol.* 2016;68:2850-2870.
3. Alonso A, Yu B, Qureshi WT, et al. Metabolomics and incidence of atrial fibrillation in African Americans: the Atherosclerosis Risk in Communities (ARIC) Study. *PLoS One.* 2015;10:e0142610.
4. Ko D, Riles EM, Marcos EG, et al. Metabolomic profiling in relation to new-onset atrial fibrillation (from the Framingham Heart Study). *Am J Cardiol.* 2016;118:1493-1496.
5. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) study: design and objectives. *Am J Epidemiol.* 1989;129:687-702.
6. Zheng Y, Yu B, Alexander D, et al. Associations between metabolomic compounds and incident heart failure among African Americans: the ARIC Study. *Am J Epidemiol.* 2013;178:534-542.
7. Alonso A, Agarwal SK, Soliman EZ, et al. Incidence of atrial fibrillation in whites and African-Americans: the Atherosclerosis Risk in Communities (ARIC) study. *Am Heart J.* 2009;158:111-117.
8. Jensen PN, Johnson K, Floyd J, Heckbert SR, Carnahan R, Dublin S. A systematic review of validated methods for identifying atrial fibrillation using administrative data. *Pharmacoepidemiol Drug Saf.* 2012;21 Suppl 1:141-147.
9. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med.* 2012;367:20-29.
10. Loehr LR, Rosamond WD, Chang PP, Folsom AR, Chambless LE. Heart failure incidence and survival (from the Atherosclerosis Risk in Communities Study). *Am J Cardiol.* 2008;101:1016-1022.
11. Christophersen IE, Rienstra M, Roselli C, et al. Large-scale analysis of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat Genet.* 2017;49:946-952.
12. Sankoh AJ, Huque MF, Dubey SD. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Stat Med.* 1997;16:2529-2542.
13. Alonso A, Krijthe BP, Aspelund T, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF Consortium. *J Am Heart Assoc.* 2013;2:e000102.
14. Yu B, de Vries PS, Metcalf GA, et al. Whole genome sequence analysis of serum amino acid levels. *Genome Biology.* 2016;17:237.
15. Minder EI, Karlaganis G, Paumgartner G. Radioimmunological determination of serum 3beta-hydroxy-5-cholenoic acid in normal subjects and patients with liver disease. *J Lipid Res.* 1979;20:986-993.
16. Wang X, Xie G, Zhao A, et al. Serum bile acids are associated with pathological progression of hepatitis B-induced cirrhosis. *Journal of Proteome Research.* 2016;15:1126-1134.
17. Schaap FG, Trauner M, Jansen PLM. Bile acid receptors as targets for drug development. *Nat Rev Gastroenterol Hepatol.* 2014;11:55-67.
18. Le Chatelier E, Nielsen T, Qin J, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature.* 2013;500:541-546.
19. Rainer PP, Primessnig U, Harenkamp S, et al. Bile acids induce arrhythmias in human atrial myocardium--implications for altered serum bile acid composition in patients with atrial fibrillation. *Heart.* 2013;99:1685-1692.
20. Li X, Zhu P, Ma S, et al. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nature Chemical Biology.* 2015;11:592-597.

21. Karijolich J, Yi C, Yu YT. Transcriptome-wide dynamics of RNA pseudouridylation. *Nature Reviews Molecular Cell Biology.* 2015;16:581-585.
22. Sander G, Topp H, Wieland J, Heller-Schöch G, Schöch G. Possible use of urinary modified RNA metabolites in the measurement of RNA turnover in the human body. *Human Nutrition. Clinical Nutrition.* 1986;40:103-118.
23. Rhodes CJ, Ghataorhe P, Wharton J, et al. Plasma metabolomics implicates modified transfer RNAs and altered bioenergetics in the outcomes of pulmonary arterial hypertension. *Circulation.* 2017;135:460-475.
24. Dunn WB, Broadhurst DI, Deepak SM, et al. Serum metabolomics reveals many novel metabolic markers of heart failure, including pseudouridine and 2-oxoglutarate. *Metabolomics.* 2007;3:413-426.
25. Sekula P, Goek ON, Quaye L, et al. A metabolome-wide association study of kidney function and disease in the general population. *J Am Soc Nephrol.* 2016;27:1175-1188.
26. Niewczas MA, Sirich TL, Mathew AV, et al. Uremic solutes and risk of end-stage renal disease in type 2 diabetes: metabolomic study. *Kidney International.* 2014;85:1214-1224.
27. Feng B, Zheng MH, Zheng YF, et al. Normal and modified urinary nucleosides represent novel biomarkers for colorectal cancer diagnosis and surgery monitoring. *Journal of Gastroenterology and Hepatology.* 2005;20:1913-1919.
28. Connolly GP, Duley JA. Uridine and its nucleotides: biological actions, therapeutic potentials. *Trends in Pharmacological Sciences.* 1999;20:218-225.
29. Deng Y, Wang ZV, Gordillo R, et al. An adipo-biliary-uridine axis that regulates energy homeostasis. *Science.* 2017;355:eaaf5375.
30. Yu B, Heiss G, Alexander D, Grams ME, Boerwinkle E. Associations between the serum metabolome and all-cause mortality among African Americans in the Atherosclerosis Risk in Communities (ARIC) study. *Am J Epidemiol.* 2016;183:650-656.
31. Menni C, Mangino M, Cecelja M, et al. Metabolomic study of carotid-femoral pulse-wave velocity in women. *J Hypertens.* 2015;33:791-796.
32. van den Berg GA, Kingma AW, Elzinga H, Muskiet FA. Determination of N-(3-acetamidopropyl)pyrrolidin-2-one, a metabolite of spermidine, in urine by isotope dilution mass fragmentography. *Journal of Chromatography.* 1986;383:251-258.
33. Butte NF, Liu Y, Zakeri IF, et al. Global metabolomic profiling targeting childhood obesity in the Hispanic population. *Am J Clin Nutr.* 2015;102:256-267.
34. Moore SC, Playdon MC, Sampson JN, et al. A metabolomics analysis of body mass index and postmenopausal breast cancer risk. *Journal of the National Cancer Institute.* 2018;doi: 10.1093/jnci/djx244. [Epub ahead of print].
35. Rebholz CM, Yu B, Zheng Z, et al. Serum metabolomic profile of incident diabetes. *Diabetologia.* 2018;61:1046-1054.
36. Huang J, Weinstein SJ, Moore SC, et al. Serum metabolomic profiling of all-cause mortality: a prospective analysis in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study cohort. *Am J Epidemiol.* 2018;doi: 10.1093/aje/kwy017. [Epub ahead of print].
37. Pegg AE. Mammalian polyamine metabolism and function. *IUBMB Life.* 2009;61.

Table 1. Selected baseline characteristics by atrial fibrillation (AF) status during follow-up in 3,922 participants with available metabolomic data and free of AF at baseline, ARIC study, 1987-89

Baseline	Overall	No incident AF	Incident AF
N	3,922	3,314	608
Age, years	54 (6)	53 (6)	56 (6)
Women, %	60	62	50
Race			
African American, %	61	64	47
White, %	39	36	53
Body mass index, kg/m^2	29 (6)	29 (6)	30 (6)
Current smoker, %	28	27	29
Systolic blood pressure, mmHg	125 (21)	124 (21)	129 (22)
Anti-hypertensive medication, %	32	31	39
Diabetes, %	14	13	19
eGFR, mL/min/1.73 m^2	99 (18)	100 (18)	94 (19)
Prevalent heart failure, %	5.1	4.5	8.4
Prevalent coronary heart disease, %	4.8	4.0	9.2
Values correspond to mean (standard deviation) or percentages. eGFR: estimated glomerular filtration rate			

Table 2. Association of two secondary bile acids (glycocholenate sulfate and glycolithocholate sulfate) with incidence of AF, by analytical batch. Hazard ratios per 1-standard deviation difference. ARIC study, 1987-2013

	First batch (N = 1919; AF = 222)		Second batch (N = 2003; AF = 386)		Combined sample (N = 3,922; AF = 608)	
	HR (95%CI)	p-value	HR (95%CI)	p-value	HR (95%CI)	p-value
Glycocholenate sulfate						
Model 1	1.27 (1.16, 1.39)	1.9×10^{-7}	1.21 (1.10, 1.33)	0.0001	1.23 (1.14, 1.32)	9.5×10^{-8}
Model 2	1.20 (1.08, 1.33)	0.0006	1.10 (1.00, 1.21)	0.05	1.13 (1.04, 1.22)	0.003
Glycolithocholate sulfate						
Model 1	1.22 (1.13, 1.31)	1.4×10^{-7}	1.02 (0.93, 1.13)	0.69	1.09 (1.01, 1.17)	0.03
Model 2	1.21 (1.11, 1.31)	5.5×10^{-6}	1.02 (0.92, 1.13)	0.67	1.07 (0.99, 1.15)	0.11
Model 1 adjusted for age, sex and race, center and batch where applicable. Model 2 additionally adjusted for smoking, body mass index, systolic blood pressure, use of antihypertensive medication, diabetes, prevalent heart failure, and prevalent coronary heart disease.						

Table 3. Association of individual metabolites with incidence of atrial fibrillation, ARIC study, 1987–2013. Hazard ratios per 1-standard deviation difference. Only metabolites that were statistically significant at p<0.001 in the multivariable model 2 are shown.

Metabolite	Model 1		Model 2		Model 3	
	HR (95%CI)	P-value	HR (95%CI)	P-value	HR (95%CI)	P-value
Pseudouridine	1.31 (1.22, 1.41)	4.5×10^{-13}	1.18 (1.10, 1.28)	1.7×10^{-5}	1.16 (1.06, 1.27)	9.6×10^{-4}
Acisoga	1.20 (1.12, 1.30)	1.3×10^{-6}	1.17 (1.09, 1.26)	4.0×10^{-5}	1.15 (1.06, 1.24)	3.7×10^{-4}
Uridine	0.82 (0.75, 0.88)	5.4×10^{-7}	0.86 (0.79, 0.93)	1.3×10^{-4}	0.86 (0.79, 0.93)	1.7×10^{-4}
1-docosahexaenoylglycerophosphocholine	0.82 (0.75, 0.90)	2.2×10^{-5}	0.85 (0.77, 0.93)	3.6×10^{-4}	0.85 (0.77, 0.93)	4.0×10^{-4}
O-sulfo-L-tyrosine	1.18 (1.09, 1.28)	5.4×10^{-5}	1.16 (1.07, 1.25)	4.0×10^{-4}	1.12 (1.03, 1.23)	0.01
Glycoursodeoxycholate	1.15 (1.08, 1.23)	3.0×10^{-5}	1.13 (1.05, 1.20)	5.2×10^{-4}	1.13 (1.05, 1.20)	5.4×10^{-4}
Glycochenodeoxycholate	1.16 (1.08, 1.24)	1.8×10^{-5}	1.13 (1.05, 1.21)	5.8×10^{-4}	1.13 (1.06, 1.21)	4.8×10^{-4}
N-acetylalanine	1.22 (1.14, 1.32)	5.6×10^{-8}	1.14 (1.06, 1.23)	6.0×10^{-4}	1.11 (1.02, 1.21)	0.02
N-acetylthreonine	1.21 (1.12, 1.31)	7.3×10^{-7}	1.14 (1.05, 1.23)	9.2×10^{-4}	1.11 (1.02, 1.21)	0.02

Model 1: Proportional hazards model adjusted for age, sex, race, study site, and batch. Model 2: As Model 1, additionally adjusted for smoking, body mass index, systolic blood pressure, use of antihypertensive medication, diabetes mellitus, prevalent heart failure and prevalent coronary heart disease. Model 3: As Model 2, additionally adjusted for eGFR.

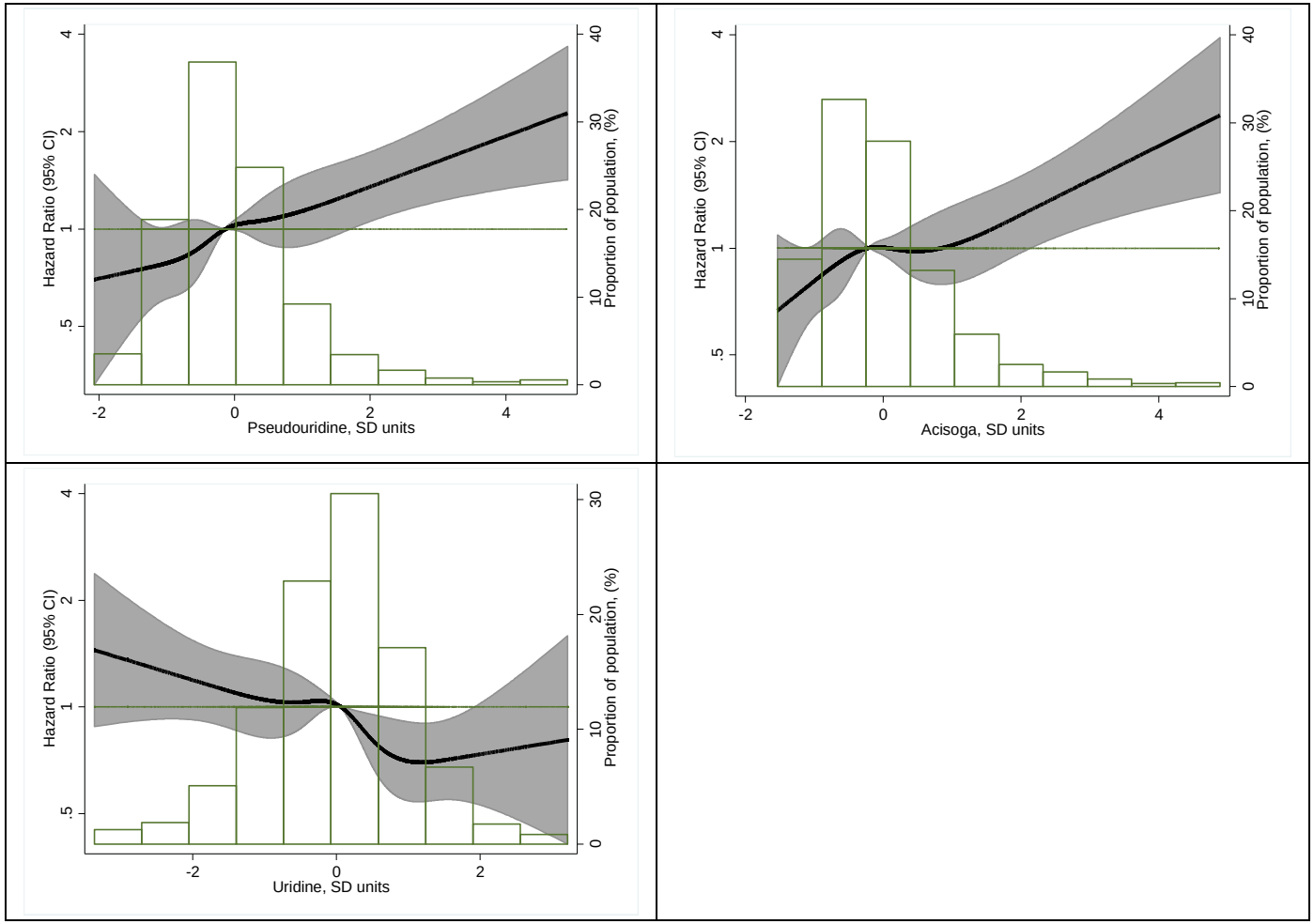
Table 4. Association of pseudouridine, acisoga and uridine with selected ECG measures, ARIC study, 1987-1989

		PR duration, ms[a]		Abnormal P wave axis		Elevated P wave terminal force in V1	
		Diff (95%CI)	P-value	OR (95%CI)	P-value	OR (95%CI)	P-value
Pseudouridine	Model 1	0.15 (-0.66, 0.97)	0.71	0.86 (0.74, 1.00)	0.04	1.11 (1.03, 1.20)	0.01
	Model 2	-0.56 (-1.39, 0.27)	0.18	0.92 (0.79, 1.06)	0.25	1.03 (0.94, 1.12)	0.55
	Model 3	-0.90 (-1.87, 0.06)	0.07	0.91 (0.77, 1.08)	0.27	1.00 (0.91, 1.11)	0.92
Acisoga	Model 1	-0.52 (-1.31, 0.27)	0.19	1.04 (0.92, 1.18)	0.54	1.07 (0.99, 1.16)	0.09
	Model 2	-0.81 (-1.60, -0.02)	0.05	1.01 (0.89, 1.15)	0.86	1.03 (0.95, 1.11)	0.51
	Model 3	-0.90 (-1.71, -0.09)	0.03	1.02 (0.89, 1.16)	0.79	1.02 (0.94, 1.11)	0.68
Uridine	Model 1	0.79 (0.01, 1.57)	0.05	0.92 (0.81, 1.04)	0.17	0.96 (0.88, 1.04)	0.30
	Model 2	0.58 (-0.21, 1.37)	0.15	1.00 (0.88, 1.15)	0.95	1.00 (0.92, 1.08)	0.93
	Model 3	0.59 (-0.20, 1.38)	0.15	1.00 (0.88, 1.15)	0.96	1.00 (0.92, 1.09)	0.98

Model 1: Adjusted for age, sex, race, study site, and batch. Model 2: As Model 1, additionally adjusted for smoking, body mass index, systolic blood pressure, use of antihypertensive medication, diabetes mellitus, prevalent heart failure and prevalent coronary heart disease. Model 3: As Model 2, additionally adjusted for eGFR

[a] Models additionally adjusted for resting heart rate

Figure 1. Association of concentrations of pseudouridine (top left panel), acisoga (top right panel) and uridine (bottom right panel) with incidence of atrial fibrillation presented as hazard ratio (HR; solid line) and 95% confidence intervals (CI; shaded area). Results from Cox proportional hazards model with metabolites modeled using restricted cubic splines (knots at 5th, 27.5th,50th, 72.5th, and 95th percentiles), adjusted for age, sex, race, batch, study site, body mass index, smoking, diabetes, systolic blood pressure, use of antihypertensive medication, prevalent coronary heart disease, and prevalent heart failure. Median value of the metabolite was considered the reference (HR = 1). The histograms represent the frequency distribution of metabolites levels. ARIC study, 1987–2013



ONLINE SUPPLEMENT

Supplementary Table 1. Partial Pearson correlation between 9 top metabolites adjusted for batch effect.

	Pseudouridine	Acisoga	Uridine	1-docosahexaenoylglycerophosphocholine	O-sulfo-L-tyrosine	Glycoursodeoxycholate	Glycochenodeoxycholate	N-acetylalanine	N-acetylthreonine
Pseudouridine	1.00	0.42	-0.02	-0.01	0.61	0.08	0.11	0.68	0.58
Acisoga	0.42	1.00	-0.03	0.06	0.28	0.03	0.09	0.33	0.29
Uridine	-0.02	-0.03	1.00	0.08	-0.01	-0.06	-0.12	0.06	-0.05
1-docosahexaenoylglycerophosphocholine	-0.01	0.06	0.08	1.00	0.02	-0.01	-0.03	0.03	0.01
O-sulfo-L-tyrosine	0.61	0.28	-0.01	0.02	1.00	0.08	0.06	0.57	0.55
Glycoursodeoxycholate	0.08	0.03	-0.06	-0.01	0.08	1.00	0.58	0.05	0.05
Glycochenodeoxycholate	0.11	0.09	-0.12	-0.03	0.06	0.58	1.00	0.07	0.09
N-acetylalanine	0.68	0.33	0.06	0.03	0.57	0.05	0.07	1.00	0.69
N-acetylthreonine	0.58	0.29	-0.05	0.01	0.55	0.05	0.09	0.69	1.00

Supplementary Table 2. Association of selected individual metabolites with incidence of atrial fibrillation after adjustment for blood lipids or exclusion of participants with prevalent cardiovascular disease at baseline, ARIC study, 1987-2013.

Metabolite	Adjustment for blood lipids[a] (N = 3,873; AF = 606)		Excluding participants with prevalent HF or CHD[b] (N = 3,567; AF = 514)	
	HR (95%CI)	P-value	HR (95%CI)	P-value
Pseudouridine	1.15 (1.06, 1.26)	0.002	1.19 (1.08, 1.31)	0.0006
Acisoga	1.14 (1.05, 1.23)	0.001	1.10 (1.01, 1.20)	0.03
Uridine	0.87 (0.81, 0.94)	0.0007	0.86 (0.79, 0.94)	0.0007

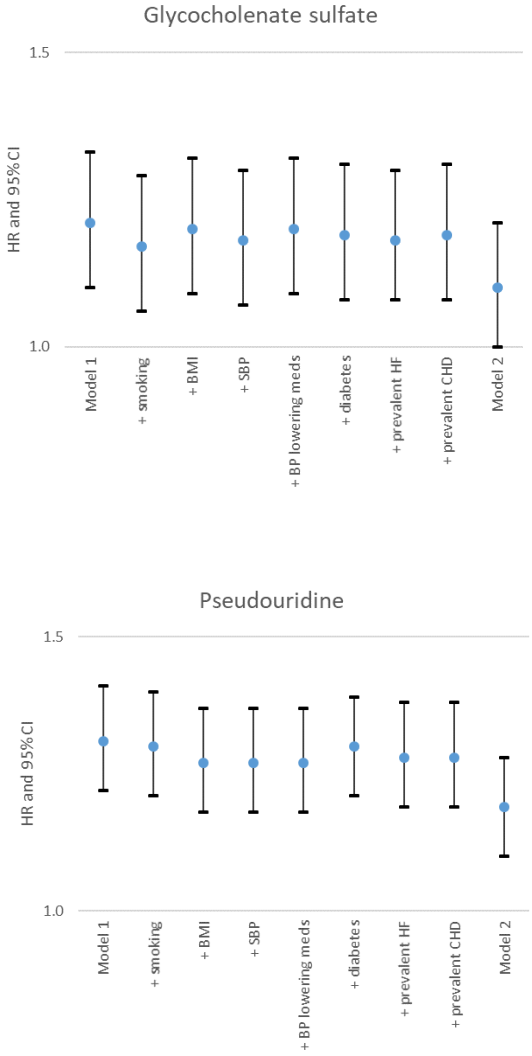
[a] Model adjusted for age, sex, race, study site, batch, smoking, body mass index, systolic blood pressure, use of antihypertensive medication, diabetes mellitus, prevalent heart failure, prevalent coronary heart disease, eGFR, total serum cholesterol, total HDL cholesterol, and total triglycerides.

[b] Model adjusted for age, sex, race, study site, batch, smoking, body mass index, systolic blood pressure, use of antihypertensive medication, diabetes mellitus, prevalent heart failure, prevalent coronary heart disease, and eGFR.

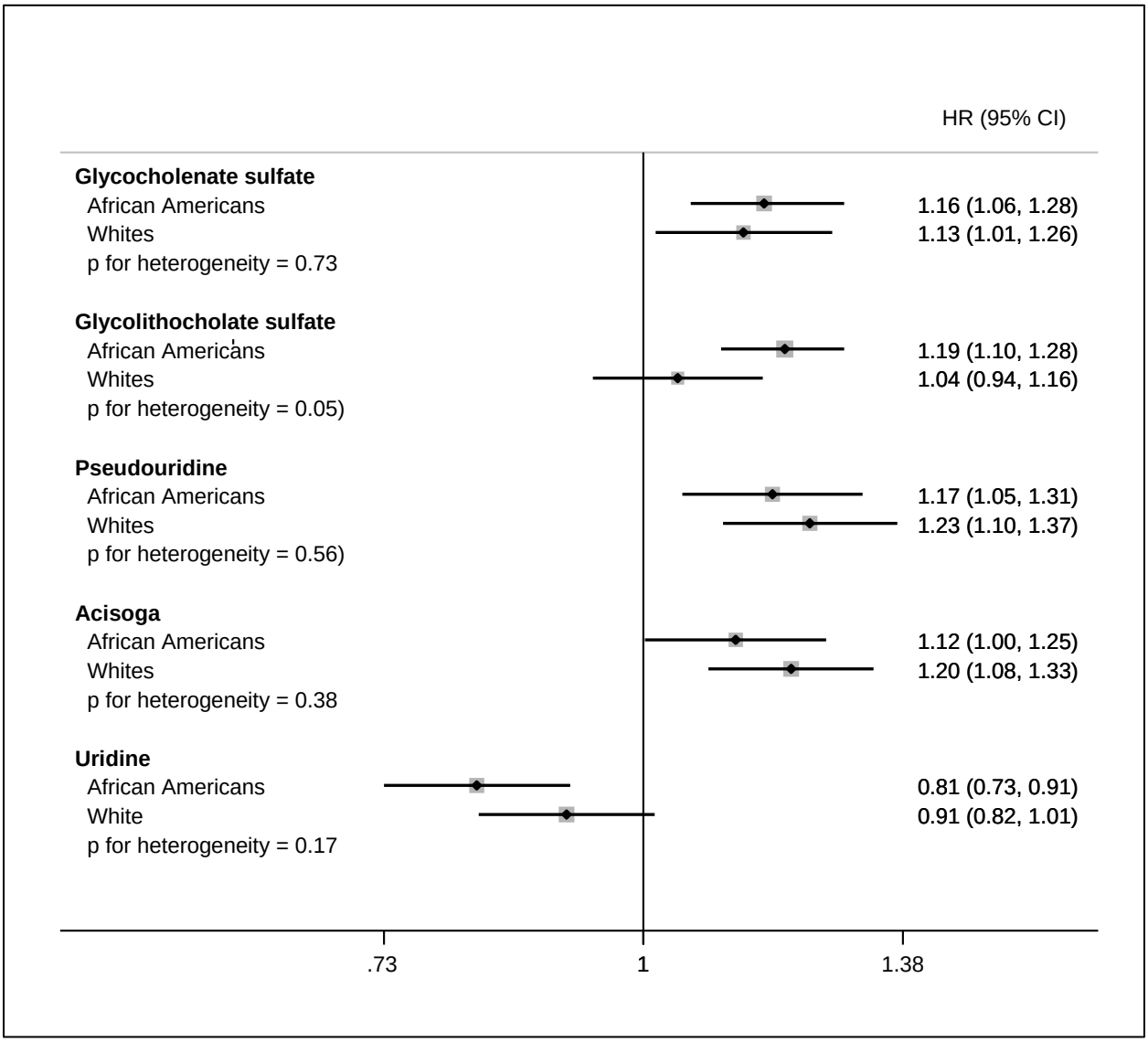
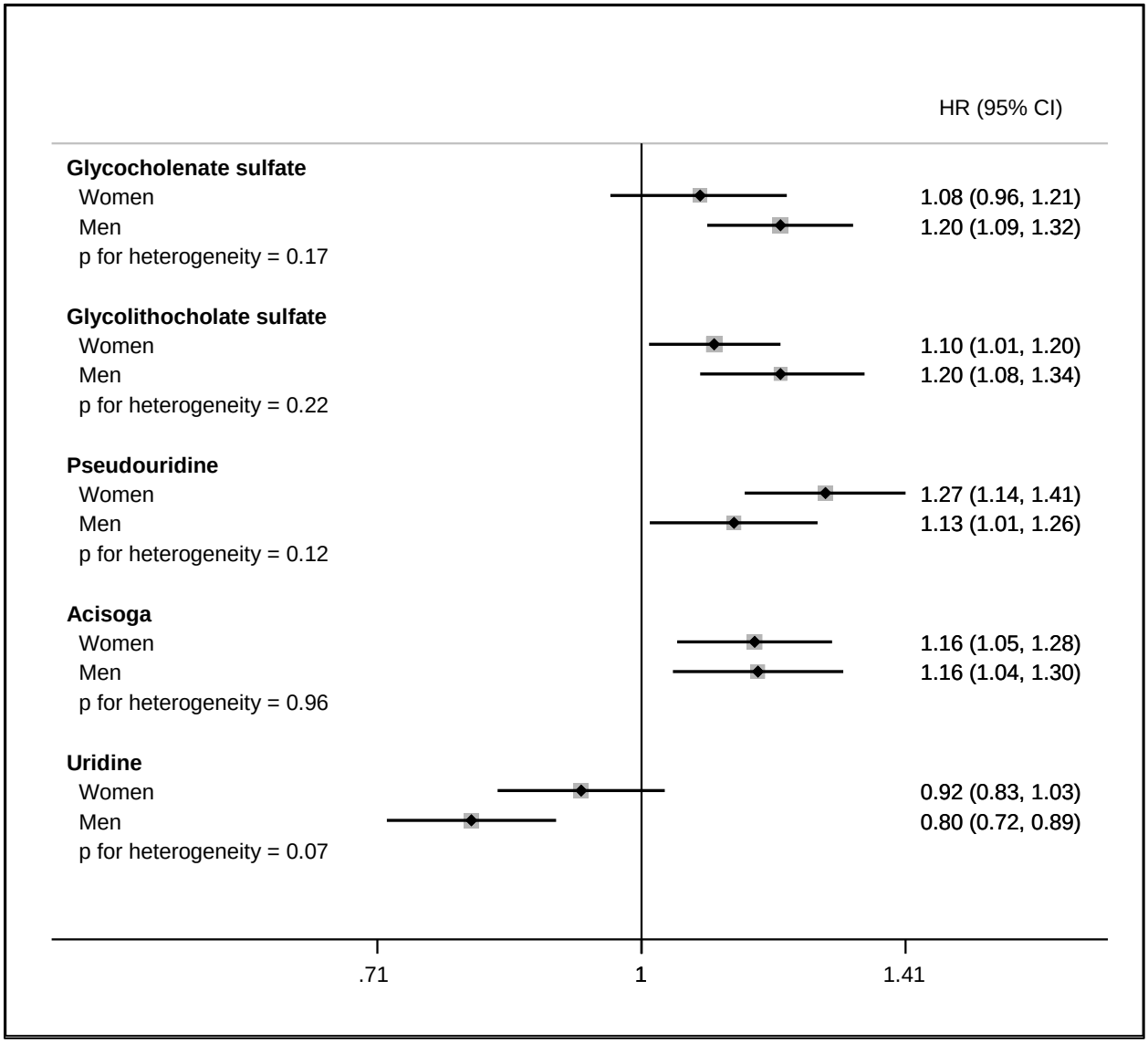
Supplementary Table 3. Association of AFGen score and SNPs with metabolites in white participants of the ARIC study (N = 1481), 1987-89. Results from multiple linear regression models adjusted for age, sex and study site.

	Pseudouridine		Acisoga		Uridine	
	Beta (95%CI)	P-value	Beta (95%CI)	P-value	Beta (95%CI)	P-value
AFGen score	-0.011 (-0.148, 0.126)	0.87	-0.013 (-0.152, 0.125)	0.85	0.041 (-0.101, 0.184)	0.57
rs11264280	-0.016 (-0.091, 0.060)	0.68	-0.024 (-0.100, 0.052)	0.54	-0.037 (-0.116, 0.041)	0.35
rs72700118	-0.082 (-0.185, 0.021)	0.12	-0.101 (-0.205, 0.003)	0.06	-0.016 (-0.123, 0.090)	0.76
rs520525	0.039 (-0.038, 0.116)	0.32	0.053 (-0.024, 0.131)	0.18	0.033 (-0.047, 0.113)	0.42
rs2540949	-0.001 (-0.072, 0.069)	0.97	-0.032 (-0.103, 0.039)	0.38	-0.007 (-0.080, 0.066)	0.86
rs3771537	0.023 (-0.045, 0.091)	0.51	0.010 (-0.059, 0.079)	0.77	0.014 (-0.057, 0.085)	0.70
rs2288327	-0.022 (-0.113, 0.070)	0.65	-0.035 (-0.128, 0.058)	0.46	0.064 (-0.031, 0.160)	0.19
rs11718898	0.069 (-0.005, 0.143)	0.07	0.042 (-0.33, 0.117)	0.27	-0.007 (-0.084, 0.070)	0.86
rs6843082	-0.025 (-0.109, 0.060)	0.57	0.027 (-0.058, 0.113)	0.53	0.011 (-0.076, 0.099)	0.80
rs337711	-0.014 (-0.083, 0.055)	0.69	-0.014 (-0.083, 0.056)	0.70	0.000 (-0.071, 0.072)	0.99
rs2967791	0.008 (-0.060, 0.076)	0.82	-0.031 (-0.100, 0.038)	0.38	-0.030 (-0.102, 0.041)	0.40
rs4946333	-0.008 (-0.077, 0.060)	0.81	-0.027 (-0.097, 0.042)	0.44	-0.005 (-0.077, 0.066)	0.88
rs12664873	0.029 (-0.045, 0.102)	0.44	-0.025 (-0.099, 0.050)	0.52	-0.055 (-0.131, 0.022)	0.16
rs1997572	0.026 (-0.046, 0.097)	0.48	0.041 (-0.032, 0.113)	0.27	0.015 (-0.059, 0.090)	0.69
rs7508	0.048 (-0.029, 0.125)	0.23	0.004 (-0.074, 0.082)	0.92	-0.021 (-0.101, 0.059)	0.61
rs7026071	-0.012 (-0.083, 0.058)	0.73	0.023 (-0.049, 0.094)	0.53	0.043 (-0.030, 0.117)	0.25
rs7915134	0.035 (-0.061, 0.131)	0.48	0.018 (-0.079, 0.115)	0.71	0.108 (0.009, 0.208)	0.03
rs11598047	0.014 (-0.080, 0.107)	0.78	-0.036 (-0.130, 0.059)	0.46	0.030 (-0.067, 0.127)	0.54
rs35176054	-0.036 (-0.135, 0.064)	0.48	0.099 (-0.001, 0.200)	0.05	0.029 (-0.074, 0.132)	0.58
rs75190942	0.121 (-0.033, 0.274)	0.12	-0.105 (-0.261, 0.050)	0.18	-0.075 (-0.234, 0.085)	0.36
rs883079	0.016 (-0.059, 0.090)	0.68	-0.012 (-0.087, 0.063)	0.76	0.005 (-0.072, 0.082)	0.90
rs1152591	-0.067 (-0.137, 0.002)	0.06	-0.048 (-0.118, 0.023)	0.18	-0.015 (-0.087, 0.058)	0.69
rs74022964	-0.037 (-0.134, 0.059)	0.45	0.044 (-0.054, 0.141)	0.38	0.029 (-0.072, 0.129)	0.57
rs2106261	-0.031 (-0.123, 0.061)	0.50	-0.035 (-0.127, 0.058)	0.47	-0.010 (-0.105, 0.086)	0.84

Supplementary Figure 1. Hazard ratios (HR) and 95% confidence intervals (95%CI) of atrial fibrillation by 1-standard deviation difference in levels of glycholenate sulfate (top panel) or pseudouridine (bottom panel), adding individual covariates to Model 1 (including age, sex, center, race and batch, if applicable). ARIC study, 1987-2013. BMI: body mass index; BP: blood pressure; CHD: coronary heart disease; HF: heart failure. Model 2 includes covariates in Model 1 and all other covariates in the figure.



Supplementary Figure 2. Association of selected metabolites with incidence of atrial fibrillation by race, ARIC study, 1987–2013. Cox proportional hazards model adjusted for age, sex, study site, batch, smoking, body mass index, systolic blood pressure, use of antihypertensive medication, diabetes mellitus, prevalent heart failure and prevalent coronary heart disease.



Supplementary Figure 3. Association of selected metabolites with incidence of atrial fibrillation by sex, ARIC study, 1987–2013. Cox proportional hazards model adjusted for age, race, study site, batch, smoking, body mass index, systolic blood pressure, use of antihypertensive medication, diabetes mellitus, prevalent heart failure and prevalent coronary heart disease.