Original Contribution

Associations Between Metabolomic Compounds and Incident Heart Failure Among African Americans: The ARIC Study

Yan Zheng, Bing Yu, Danny Alexander, Teri A. Manolio, David Aguilar, Josef Coresh, Gerardo Heiss, Eric Boerwinkle, and Jennifer A. Nettleton*

* Correspondence to Dr. Jennifer A. Nettleton, Suite E641, 1200 Hermann Pressler Street, Houston, TX 77030 (e-mail: jennifer.a.nettleton@uth.tmc.edu).

Heart failure is more prevalent among African Americans than in the general population. Metabolomic studies among African Americans may efficiently identify novel biomarkers of heart failure. We used untargeted methods to measure 204 stable serum metabolites and evaluated their associations with incident heart failure hospitalization ($n = 276$) after a median follow-up of 20 years (1987–2008) by using Cox regression in data from 1,744 African Americans aged 45–64 years without heart failure at baseline from the Jackson, Mississippi, field center of the Atherosclerosis Risk in Communities (ARIC) Study. After adjustment for established risk factors, we found that 16 metabolites (6 named with known structural identities and 10 unnamed with unknown structural identities, the latter denoted by using the format X-12345) were associated with incident heart failure ($P < 0.0004$ based on a modified Bonferroni procedure). Of the 6 named metabolites, 4 are involved in amino acid metabolism, 1 (prolylhydroxyproline) is a dipeptide, and 1 (erythritol) is a sugar alcohol. After additional adjustment for kidney function, 2 metabolites remained associated with incident heart failure (for metabolite X-11308, hazard ratio = 0.75, 95% confidence interval: 0.65, 0.86; for metabolite X-11787, hazard ratio = 1.23, 95% confidence interval: 1.10, 1.37). Further structural analysis revealed X-11308 to be a dihydroxy docosatrienoic acid and X-11787 to be an isoform of either hydroxyleucine or hydroxyisoleucine. Our metabolomic analysis revealed novel biomarkers associated with incident heart failure independent of traditional risk factors.

heart failure; metabolomics; risk factors

Abbreviations: ARIC, Atherosclerosis Risk in Communities; eGFR, estimated glomerular filtration rate; PUFA, polyunsaturated fatty acid.

Heart failure is a leading cause of morbidity and mortality and its prevalence continues to rise in the United States, partly because of a shift in the age distribution of the population (1, 2). Identification of novel risk factors for heart failure may reduce the burden on the health-care system and improve quality of life (1). Several risk factors for heart failure have been identified, including age, African-American race, male sex, prevalent coronary heart disease, left ventricular hypertrophy, diabetes, hypertension, kidney dysfunction, low physical activity, dyslipidemia, cigarette smoking, and obesity (1, 3–13).

Metabolomic technologies allow for global characterization of metabolic networks by untargeted quantification of many low-molecular-weight metabolites present in a biolog-

ical sample (14). Thus, metabolomic analysis may identify single biomarkers of heart failure or collections of metabolites pointing to dysregulated metabolic pathways underlying heart failure etiology. Although some studies have evaluated metabolomic changes following heart failure onset (15, 16), few have sought to identify upstream physiological factors that predict future heart failure. Furthermore, to date, no study has explored the metabolomic antecedents of heart failure in African Americans, a race group with the highest rate of hospitalization for heart failure and cardiovascular disease mortality (17). Although socioeconomic status and health care accessibility likely play important roles (18), the physiological factors underlying this risk disparity are not well understood.

Therefore, we evaluated whether baseline metabolomic biomarkers were associated with incident hospitalized heart failure independent of established risk factors during approximately 20 years of follow-up in a well-characterized, population-based sample of African Americans from the Atherosclerosis Risk in Communities (ARIC) Study.

MATERIALS AND METHODS

Study design and population

The ARIC Study is designed to study the causes and outcomes of cardiovascular disease in a prospective cohort of 15,792 individuals from 4 communities (Forsyth County, North Carolina; Jackson, Mississippi; the suburbs of Minneapolis, Minnesota; and Washington County, Maryland). Detailed descriptions of its design, objectives, and procedures have been published elsewhere (19). Metabolomic profiles were measured in a subsample of African-American participants, including 1,977 individuals randomly selected from the Jackson, Mississippi, field center who provided consent for genetic research, provided dietary quality data, and had a stored fasting ($\geq$8 hours) serum sample at the baseline examination. For the current analysis, we excluded participants with baseline prevalent heart failure ($n = 97$), missing follow-up information on incident heart failure ($n = 30$), or missing values for the baseline covariates ($n = 106$). The final sample size for the results presented here was 1,744.

Assessment of baseline covariates and metabolites

Prevalent coronary heart disease was defined by evidence of previous myocardial infarction by electrocardiogram at baseline, history of physician-diagnosed myocardial infarction, or previous coronary reperfusion procedure (19). Current cigarette smoking status, physical activity (continuous variable) (20), and educational level (6-level ordinal variable) (21) were self-reported. A standardized protocol was used to measure blood pressure (22). Antihypertensive and lipid-lowering medication use during the past 2 weeks was ascertained by inventory. Diabetes was defined as a fasting serum glucose level $\geq$126 mg/dL (23) or the use of diabetes medication. Left ventricular hypertrophy was measured by electrocardiography according to the Cornell voltage criteria (24). Estimated glomerular filtration rate (eGFR) was calculated by using the Chronic Kidney Disease Epidemiology Collaboration equation (25).

Metabolite profiling of fasting serum samples that had been stored at $-80°$C since collection at baseline in 1986–1987 was completed in June 2010. Detection and quantification of 602 metabolites were completed by Metabolon, Inc. (Durham, North Carolina) by using an untargeted, gas chromatography–mass spectrometry and liquid chromatography–mass spectrometry–based metabolomic quantification protocol (26). This untargeted approach identifies and quantifies named compounds whose chemical identities are known ($n = 361$), as well as additional unnamed compounds that do not currently have chemical standards ($n = 241$) (27). These unnamed compounds are tagged beginning with "X" and followed by numbers (e.g., X-12345). A rigorous assessment

of the metabolomic data was done on the basis of 2 criteria. First, 44 metabolites were excluded because more than 80% of the samples had missing values or values below the detection limit of the technology. Considering the appropriateness of population-based referencing, we preferred a relatively stable biomarker with a small intrasubject variance compared with the between-subject variance (28). Second, 386 metabolites were excluded because the medium-term reliability coefficient based on a comparison of 2 samples collected 4–6 weeks apart from 60 individuals was less than 0.60. Of note, 32 metabolites were excluded because they had more than 80% missing values and reliability coefficients of less than 0.60. Thus, the present report is based on an evaluation of 204 stable metabolites, of which 118 had known and 86 had unknown structural identities (reliability coefficient $\geq$0.60; median reliability coefficient, 0.74). The values of 187 metabolites with reliability coefficients of 0.60 or higher and missing/below-the-detection-limit values of less than 50% were analyzed as continuous variables, and missing/below-the-detection-limit values were assigned the lowest detected value for that metabolite in all samples. Seventeen metabolites with reliability coefficients of 0.60 or higher and missing/below-the-detection-limit values between 50% and 80% were analyzed as ordinal variables with the following levels: 1) missing/below-the-detection-limit values, 2) detected values below the median; and 3) detected values at or higher than the median.

Instrument variability was determined by calculating the median relative standard deviation for the internal standards that were added to each sample prior to injection into the mass spectrometers. Overall process variability was determined by calculating the median relative standard deviation for all endogenous metabolites (i.e., noninstrument standards) present in 100% of the technical replicate samples, which were created from a homogeneous pool of human plasma. Because this study spanned multiple days, a data normalization step was performed to correct variation resulting from differences in instrument tuning from day to day. Essentially, each compound was corrected in run-day blocks by registering the medians to equal 1.00 and normalizing each data point proportionately.

Compounds were identified by comparison with an in-house–generated authentic standard library that includes retention times, molecular weights, preferred adducts, in-source fragments, and associated fragmentation spectra of the intact parent ions. The database allows for rapid and high-confidence identification of experimentally detected molecules on a multiparameter match basis without the need for additional analyses. Unnamed compounds of interest were subjected to detailed analysis by mass spectrometry and mass spectrometry[n] (i.e., the accurate monoisotopic mass and fragmentation patterns of the primary ion, as well as higher order fragmentations) by using an Orbitrap Elite instrument (Thermo Fisher Scientific, Inc., Waltham, Massachusetts). Resulting empirical formulas were queried against the ChemSpider database (http://www.chemspider.com/) search engine for potential structure matches.

Follow-up and outcome events

Individuals were considered to have prevalent heart failure if they had stage-3 heart failure or "manifest heart failure" at baseline according to the Gothenburg criteria (29, 30) or

medication use for heart failure during the last 2 weeks before baseline (19, 31). In individuals free of prevalent heart failure, incident heart failure was defined as the first occurrence of either a hospitalization that included an *International Classification of Diseases, Ninth Revision*, discharge code of 428 (428.0–428.9) in any position or a death certificate with a code of 428 (heart failure) or *International Classification of Diseases, Tenth Revision*, code I50 (heart failure) in any position (6, 31). Participants who were lost to follow-up were censored at the date of last contact. Individuals were followed up for events until death or December 31, 2008. There were 276 incident heart failure cases during approximately 20 years of follow-up.

Statistical analysis

Baseline data are presented as means and standard errors for the continuous variables and as numbers and percentages for the categorical variables. Baseline characteristics were compared by using the χ^2 test for categorical variables and a 2-sample Student's t test for continuous variables. Cox proportional hazards regression was used to assess associations between each metabolite and incident hospitalization for heart failure. Covariates were selected on the basis of published reports (32–34). Four hierarchical models were used. Model 1 adjusted for unmodifiable factors, including age, sex, and prevalent coronary heart disease. Model 2 additionally adjusted for the following lifestyle factors: physical activity, education, and current smoking status. Model 3 additionally adjusted for the following physiological factors: body mass index (weight (kg)/height (m)^2), systolic blood pressure, antihypertensive medication use, fasting serum glucose levels, prevalent diabetes, ratio of serum total cholesterol to high-density lipoprotein cholesterol, lipid-lowering medication use, and left ventricular hypertrophy. Model 4 additionally adjusted for eGFR (25) because kidney dysfunction is a known risk factor for heart failure (35) and may also reciprocally act on serum metabolite concentrations (i.e., kidney function may affect the levels of certain metabolites, and certain metabolites, especially the exogenous metabolites from medication with nephrotoxicity, may be associated with altered kidney function) (36, 37). Therefore, kidney function may be a confounder or a mediator in the association between the human metabolome and heart failure. Finally, metabolites identified as being associated with heart failure in the fullest model (model 4) were included together in a single model to determine whether their associations with heart failure were mutually independent.

An exploratory analysis that used a composite metabolomic score was also conducted. The metabolomic score was created by summing the quartile ranks of metabolites that were associated with heart failure in model 3 (the quartile ranks were reversed for metabolites inversely associated with heart failure). The association between this metabolomic score and incident heart failure was assessed by adjusting for all covariates in model 4.

Results of the Cox proportional hazards models are reported as hazard ratios with corresponding P values. All hazard ratios were calculated and reported per standard deviation increase for continuous variables or per 1-category change for categorical

variables. The proportionality assumption was evaluated and accepted by using the log-log survival curves for categorical covariates and Schoenfeld's goodness-of-fit test (38) for continuous covariates.

A modified stepwise Bonferroni procedure, the Dubey/Armitage-Parmar algorithm (39), was used to correct for multiple comparisons. This adjustment takes into account the full correlation matrix of metabolites and uses the mean correlation among the metabolites in the formula, where the new α-level for the kth hypothesis for $k = 1, 2, \ldots, K$ is readjusted for each individual metabolite as follows:

$$\alpha_k = 1 - (1 - \alpha)^{1/m_k},$$

where

$$m_k = K^{1-r_k}, \quad r_k = (K-1)^{-1} \sum_{j \neq k}^{K} r_{jk},$$

and r_{jk} is the correlation coefficient between the jth and kth metabolites. When the average of the correlation coefficients is 0, this adjustment is identified with the Bonferroni test, and when it is 1, the adjusted and the unadjusted P values are the same. When applying this method to each individual test of association within the current metabolomics data, we considered a 2-tailed $P < 0.0004$ to be statistically significant.

All statistical analyses were performed with SAS, version 9.2, software (SAS Institute, Inc., Cary, North Carolina).

RESULTS

Among the 1,744 study participants, 16% ($n = 276$) developed hospitalized heart failure during a median follow-up of 20 years (incidence rate: 9.3 per 1,000 person-years). Participants who developed heart failure were more likely to be older, current smokers, obese, hypertensive, diabetic, less physically active and to have prevalent coronary heart disease (Table 1). Except for sex, all the baseline risk factors differed between those who developed heart failure and those who did not during the follow-up period ($P < 0.05$).

In models 1, 2, 3, and 4, there were 56, 43, 16, and 2 metabolites, respectively, associated with incident heart failure ($P < 0.0004$). As covariates were added, the number of statistically significant metabolites was progressively reduced, suggesting that lifestyle, physiological factors, and kidney function are confounders in the association between the human metabolome and incident heart failure. In model 3, the 16 significant metabolites included 6 named metabolites with known structural identities and 10 unnamed metabolites with unknown structural identities. These 16 metabolites were analyzed as continuous variables because of a high reliability coefficient and a small percentage of missing values or those below the detection limit. The multivariable Cox proportional hazards models for these 16 metabolites are shown in Table 2. Metabolite X-11308 had the largest effect size and was the only metabolite inversely associated with heart failure (hazard ratio = 0.75 per standard deviation; $P = 8.34 \times 10^{-5}$ in model 3). Four of the 6 named metabolites (*N*-acetylalanine, *p*-cresol sulfate, phenylacetylglutamine, pyroglutamine) are involved in amino acid metabolism; 1

Table 1. Distribution of Baseline Established Risk Factors by Incident Heart Failure Among 1,744 African Americans in the Jackson, Mississippi, Field Center of the ARIC Study, 1987–2008

Risk Factor	Total Sample (n = 1,744)			Nonincident Heart Failure (n = 1,468)			Incident Heart Failure (n = 276)		
	No.	%	Mean (SE)	No.	%	Mean (SE)	No.	%	Mean (SE)
Unmodifiable factors									
Age, years***			52.8 (5.7)			52.3 (5.6)			55.3 (5.6)
Male sex	633	36.3		530	36.1		103	37.3	
Prevalent coronary heart disease***	57	3.3		31	2.1		26	9.4	
Lifestyle factors									
Current smoker***	498	28.6		397	27.0		101	36.6	
Physical activity [a], ***			6.6 (1.4)			6.7 (1.4)			6.1 (1.5)
Educational level***									
Grade school or 0 years of education	328	18.8		235	16.0		93	33.7	
High school but no degree	369	21.2		313	21.3		56	20.3	
High school graduate	367	21.1		316	21.5		51	18.5	
Vocational school	119	6.8		106	7.2		13	4.7	
College	308	17.7		268	18.3		40	14.5	
Graduate or professional school	253	14.5		230	15.7		23	8.3	
Physiology factors									
Body mass index[b], ***			29.5 (6.0)			29.2 (5.8)			31.1 (6.6)
Systolic blood pressure, mm Hg***			128.1 (21.3)			126.7 (20.4)			135.2 (24.5)
Antihypertensive medication use**	618	35.4		475	32.4		143	51.2	
Fasting glucose, mmol/L***			6.2 (2.5)			5.9 (2.0)			7.6 (4.1)
Prevalent diabetes***	270	15.5		170	11.6		100	36.2	
Serum total cholesterol to high-density lipoprotein cholesterol ratio***			4.2 (1.5)			4.1 (1.5)			4.5 (1.5)
Lipid-lowering medication use*	17	1.0		11	0.8		6	2.2	
Left ventricular hypertrophy***	83	4.8		50	3.4		33	12.0	
eGFR, mL/min/1.73 m²***			104.5 (18.1)			105.5 (16.9)			99.2 (22.5)

Abbreviations: ARIC, Atherosclerosis Risk in Communities; eGFR, estimated glomerular filtration rate; SE, standard error.

* $P < 0.05$; **$P < 0.01$; ***$P < 0.001$ (P values for the test for difference in each risk factor between the nonincident heart failure group and the incident heart failure group).

[a] Physical activity was calculated from a modified Baecke questionnaire on habitual physical activity by integrating the types and intensity of physical activity via principal components analysis; therefore, there is no unit of measurement.

[b] Body mass index is weight (kg)/height (m)².

(prolylhydroxyproline) is a dipeptide; and 1 (erythritol) is a sugar alcohol. Except for pyroglutamine and prolylhydroxyproline, these metabolites were significantly associated with heart failure in all 3 models. Pyroglutamine was associated with heart failure only in model 3 with an 11% increase in the magnitude of its hazard ratio (from 1.17 in model 2 to 1.29 in model 3), a change attributable almost solely to adjustment for fasting serum glucose level and prevalent diabetes. Associations between the baseline established risk factors and the 16 metabolites that were significantly associated with incident heart failure in model 3 are shown in the Web Appendix (available at http://aje.oxfordjournals.org/). The correlations between each of these 16 metabolites and eGFR are shown in Table 3. Except for X-11308, heart failure–related metabolites were consistently negatively associated with eGFR (i.e., higher concentrations of metabolites were related to poorer kidney function). The associations of these 16 metabolites with the other established risk factors varied considerably across metabolites (Web Appendix).

After further adjustment for eGFR (in model 4), only X-11308 ($P = 5.36 \times 10^{-5}$) and X-11787 ($P = 2.84 \times 10^{-4}$) remained significantly associated with heart failure risk, whereas the positive association between pyroglutamine and heart failure risk was of borderline statistical significance ($P = 0.0006$). eGFR itself was associated with incident heart failure in model 4 (hazard ratio = 0.812, 95% confidence interval: 0.724, 0.911), showing the expected relationship between poor kidney function and heart failure risk. Web Figure 1 illustrates the cumulative hazard of heart failure by quartiles of these 3 metabolites. When included together in

Table 2. Adjusted Hazard Ratios[a] for Incident Heart Failure in 3 Models in Relationship to 16 Candidate Metabolites[b] in Model 3 Among 1,744 African Americans in the Jackson, Mississippi, Field Center of the ARIC Study, 1987–2008

Metabolite	Model 1[c]		Model 2[d]			Model 3[e]			Model 4[f]		
	HR	95% CI	HR	95% CI	Δβ[g] From Model 1, %	HR	95% CI	Δβ[g] From Model 2, %	HR	95% CI	Δβ[g] From Model 3, %
X-11429	1.22*	1.15, 1.29	1.16*	1.09, 1.23	−26.47	1.14*	1.08, 1.21	−10.70	1.10	1.02, 1.18	−29.01
X-11687_200	1.27*	1.20, 1.35	1.24*	1.16, 1.32	−12.33	1.18*	1.10, 1.26	−23.24	1.13	1.03, 1.22	−27.50
Erythritol	1.29*	1.22, 1.35	1.24*	1.18, 1.31	−13.46	1.16*	1.09, 1.23	−32.94	1.10	1.01, 1.20	−32.43
N-Acetylalanine	1.18*	1.11, 1.25	1.16*	1.09, 1.23	−11.88	1.15*	1.08, 1.23	−1.67	1.11	1.02, 1.20	−29.20
X-12096	1.28*	1.21, 1.36	1.22*	1.15, 1.30	−18.93	1.15*	1.08, 1.23	−28.51	1.10	1.01, 1.20	−34.25
X-11787[h]	1.36*	1.23, 1.51	1.36*	1.23, 1.51	1.30	1.27*	1.14, 1.41	−23.87	1.23*	1.10, 1.37	−13.19
X-11334	1.18*	1.12, 1.25	1.17*	1.11, 1.23	−7.70	1.13*	1.07, 1.19	−22.31	1.09	1.02, 1.16	−30.00
Pyroglutamine	1.20	1.06, 1.36	1.17	1.03, 1.32	−15.26	1.29*	1.15, 1.46	67.14	1.25	1.10, 1.41	−14.52
X-11423	1.22*	1.16, 1.29	1.17*	1.10, 1.23	−23.13	1.13*	1.07, 1.19	−21.98	1.08	1.00, 1.16	−37.32
Phenylacetylglutamine	1.20*	1.13, 1.27	1.18*	1.11, 1.25	−11.57	1.15*	1.08, 1.23	−12.49	1.10	1.01, 1.20	−33.26
X-04499	1.36*	1.22, 1.53	1.31*	1.17, 1.47	−12.66	1.25*	1.12, 1.40	−17.44	1.19	1.06, 1.34	−21.81
X-11308[i]	0.67*	0.58, 0.78	0.72*	0.63, 0.83	−18.67	0.75*	0.65, 0.87	−11.28	0.75*	0.65, 0.86	1.46
X-11333	1.23*	1.16, 1.30	1.17*	1.10, 1.24	−23.16	1.13*	1.06, 1.20	−20.71	1.08	1.00, 1.17	−39.75
p-Cresol sulfate	1.26*	1.15, 1.38	1.21*	1.11, 1.33	−17.00	1.18*	1.08, 1.28	−14.35	1.11	1.00, 1.23	−36.29
X-11564	1.31*	1.21, 1.42	1.23*	1.14, 1.34	−23.50	1.16*	1.07, 1.26	−29.13	1.09	0.99, 1.20	−40.09
Prolylhydroxyproline	1.24*	1.12, 1.37	1.18	1.07, 1.31	−22.33	1.17*	1.07, 1.27	−7.52	1.11	1.01, 1.22	−31.60

Abbreviations: ARIC, Atherosclerosis Risk in Communities; CI, confidence interval; HR, hazard ratio.

* $P < 0.0004$.

[a] Hazard ratios are calculated for 1 standard deviation increase in continuous variables or a transfer from 1 level to another in categorical variables.

[b] Unnamed metabolites with unknown structural identities are denoted by using the format X-12345.

[c] Model 1 is adjusted for unmodifiable factors.

[d] Model 2 is adjusted for unmodifiable factors and lifestyle factors.

[e] Model 3 is adjusted for unmodifiable factors, lifestyle factors, and physiology factors.

[f] Model 4 is adjusted for unmodifiable factors, lifestyle factors, physiology factors, and estimated glomerular filtration rate.

[g] Change in the β regression coefficient.

[h] An isoform of either hydroxyleucine or hydroxyisoleucine.

[i] A dihydroxy docosatrienoic acid.

model 4, the 3 metabolites, X-11308, X-11787, and pyroglutamine, were each independently associated with heart failure ($P = 0.0004$, 0.001, and 0.004, respectively). Sensitivity analyses excluding cases during the first 3, 5, and 8 years of follow-up were performed, and X-11308 and X-11787 were consistently identified as the top 2 metabolites in each analysis (data not shown).

We further identified with good confidence the general structures of the 2 unnamed metabolites, X-11308 and X-11787. The unnamed metabolite, X-11308, has a monoisotopic accurate mass of 366.27736 g/mol and an empirical formula of $C_{22}H_{38}O_4$ with 4 double bonds based on perfect matches in both mass and isotope pattern. Detailed analyses of fragmentation pathways revealed it to be consistent with dihydroxy docosatrienoic acid (22:3), a long-chain polyunsaturated fatty acid (PUFA). For the other unnamed metabolite, X-11787, the neutral accurate mass was measured as 147.08952 g/mol with an empirical molecular formula of $C_6H_{13}NO_3$ and 1 double bond. Analysis of fragmentation spectra suggests this unknown metabolite is an isoform of

either hydroxyleucine or hydroxyisoleucine. The metabolomic score created from the 16 metabolites identified in model 3 ranged from 2 to 48 (mean = 23.84; standard deviation, 8.47). Each higher standard deviation in the metabolomic score was associated with a 38% greater risk of heart failure in model 4 (hazard ratio = 1.38, 95% confidence interval: 1.18, 1.62; $P = 5.36 \times 10^{-5}$). A histogram of the metabolomic score by incident heart failure status is shown in Figure 1.

To better understand how the identified metabolites fit into larger metabolomic pathways, we calculated the age- and sex-adjusted Pearson correlations among all of the metabolites (Web Figure 2A). Strong correlations were observed in the X-11308 and X-11372 pair ($r = 0.66$), the X-11308 and X-11880 pair ($r = 0.60$), and the pyroglutamine and creatine pair ($r = -0.44$). Except in these pairs, the correlations between the metabolites associated with heart failure and the other metabolites were weak (metabolite-metabolite r ranged from −0.14 to 0.36). To illustrate the pairwise correlations among the identified potential biomarkers of heart failure, we show

Table 3.   Associations Between Each of 16 Candidate Metabolites[a] and eGFR Among 1,744 African Americans in the Jackson, Mississippi, Field Center of the ARIC Study, 1987–2008

Metabolite	Pearson Correlation Coefficient[b] With eGFR	P Value
X-11429	−0.389	1.78×10^{-69}
X-11687_200	−0.335	8.73×10^{-50}
Erythritol	−0.426	9.90×10^{-86}
N-Acetylalanine	−0.387	1.06×10^{-68}
X-12096	−0.391	2.68×10^{-70}
X-11787[c]	−0.163	5.33×10^{-12}
X-11334	−0.294	1.27×10^{-37}
Pyroglutamine	−0.188	2.27×10^{-15}
X-11423	−0.447	2.59×10^{-96}
Phenylacetylglutamine	−0.329	5.72×10^{-48}
X-04499	−0.225	5.31×10^{-22}
X-11308[d]	−0.008	0.75
X-11333	−0.380	1.14×10^{-65}
p-Cresol sulfate	−0.342	5.42×10^{-52}
X-11564	−0.361	1.23×10^{-58}
Prolylhydroxyproline	−0.264	3.09×10^{-30}

Abbreviations: ARIC, Atherosclerosis Risk in Communities; eGFR, estimated glomerular filtration rate.

[a] Unnamed metabolites with unknown structural identities are denoted by using the format X-12345.

[b] Correlation coefficients are calculated with adjustment for age and sex.

[c] An isoform of either hydroxyleucine or hydroxyisoleucine.

[d] X-11308, a dihydroxy docosatrienoic acid, is the only heart failure–related metabolite that was not significantly associated with eGFR.

the partial metabolite-metabolite correlations of the 16 significant metabolites identified in model 3 in Web Figure 2B.

DISCUSSION

By using a mass spectrometry–based metabolite profiling platform (26), we found that some metabolomic compounds measured in prospectively collected and stored serum samples were significantly associated with incident heart failure in a well-characterized, population-based sample of African Americans. Associations between 16 metabolites and incident heart failure were noted after adjustment for established risk factors. Each 1–standard deviation increase in a metabolomic score, consisting of the sum of the quartile ranks of the 16 metabolites, was associated with a 38% increased risk of heart failure in the fullest model (95% confidence interval: 1.18, 1.62). Two metabolites, X-11308 and X-1178, remained associated and 1 metabolite (pyroglutamine) retained a borderline significant association with incident heart failure after adjustment for established heart failure risk factors and eGFR. The metabolites that were no longer associated after adjustment for eGFR may be related to incident heart failure

via pathways related to kidney dysfunction. For example, p-cresol sulfate, a significant predictor in model 3 with adverse cardiovascular effects (40, 41), is cleared by the kidneys and was no longer a significant predictor of incident heart failure after adjustment for eGFR.

Though the exact identification of unknown metabolites remains a challenge (42), the putative chemical structures for the 2 unnamed candidate metabolites were identified. The unnamed metabolite, X-11308, was identified as a dihydroxy docosatrienoic acid (22:3), which is a PUFA. Long-chain PUFAs are considered to be associated with lower cardiovascular risk, including heart failure (43). PUFAs may reverse adverse mechanical/electrical remodeling via a multitude of mechanisms, such as improving hemodynamics, promoting antiinflammatory activation, and increasing cardiac output. Therefore, PUFAs have hypothesized cardioprotective effects that subsequently reduce the risk of heart failure (44). Among the PUFAs, eicosapentaenoic acid, docosahexaenoic acid, and α-linolenic acid are well known for their benefits to cardiovascular health. The results of this study suggest that docosatrienoic acid may also be associated with a lower risk of heart failure. The other unnamed metabolite, X-11787, was identified as an isoform of either hydroxyleucine or hydroxyisoleucine. Hydroxyleucines are known oxidized end products of leucine or hydroperoxyleucines. This modification is irreversible, making hydroxyleucines useful in vivo markers of protein oxidation (45). Oxidative damage to cellular proteins and membranes is an obvious mechanism through which myocardial oxidative stress impairs cardiac function, thereby inducing cellular dysfunction or death through apoptosis and necrosis (46). Oxidation of skeletal muscle actin, tropomyosin, and myosin occurs in chronic heart failure, and it further depresses muscle function and exercise capacity (47). To our knowledge, there are no previous reports of a possible mechanism for a cardiac effect of hydroxyisoleucine.

Strengths and limitations

Strengths of this study include the use of a well-characterized African-American cohort with detailed clinical characterization. The long period of prospective follow-up enabled us to show that the circulating metabolomic differences can occur well before any alteration is detectable in pathophysiology of heart failure development by standard clinical measures. Earlier research results (48) suggest that error can occur when single measurements of serum metabolites are used to categorize individuals because of high intraindividual variation; therefore, strict quality standards for the metabolomic data were applied here to ensure a valid and reliable conclusion. This study is among the first to evaluate the medium-term reliability of metabolomic data in human samples from a prospective cohort study although, as a consequence, only about one-third of the metabolites satisfied our criterion for reliability to be included in this analysis. As an exploratory exercise, we reanalyzed incident heart failure beginning with all 602 metabolites, and the 2 significant metabolites identified in model 4 remained significant (data not shown). The reliability of these 2 metabolites over time and the consistency of their association with incident heart failure bode well for
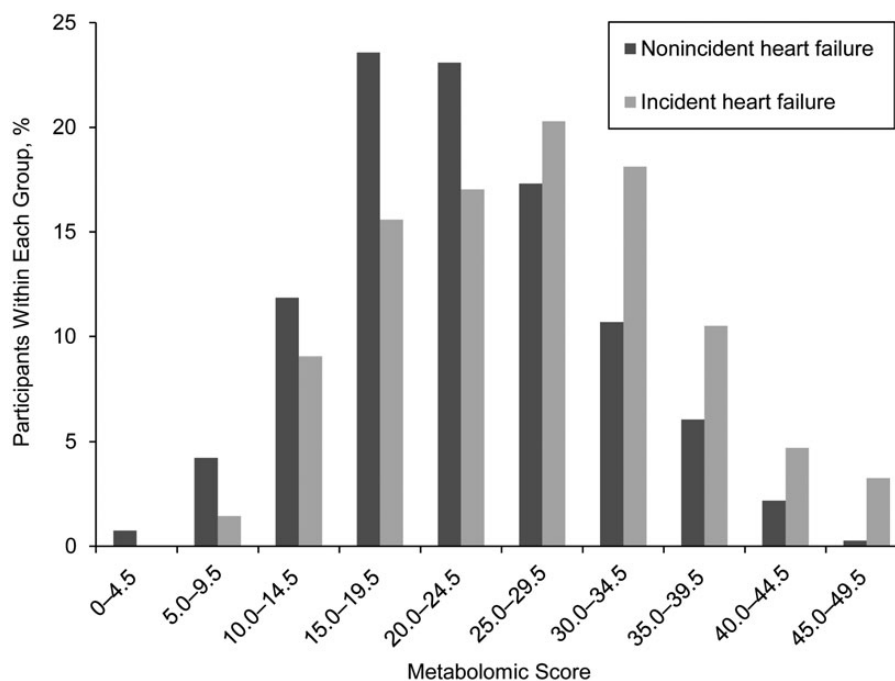
Figure 1.   A histogram of metabolomic score by incident heart failure status among 1,744 African Americans from the Jackson, Mississippi, field center of the Atherosclerosis Risk in Communities (ARIC) Study, 1987–2008.

their potential as applied biomarkers contributing to the body of knowledge regarding heart failure etiology or predicting future heart failure. A growing number of studies have used mass spectrometry as a tool for biomarker discovery, but these studies have been largely cross-sectional (49, 50), providing limited information regarding the relationships of metabolomic biomarkers to the development of future disease (14). No study, to our knowledge, has tested whether metabolomic profiling of prospectively collected serum samples in large cohorts can identify novel biomarkers of future incident heart failure. The present study identified novel candidate metabolites that were associated with future heart failure by using untargeted metabolite profiling in African Americans in the ARIC Study over approximately 20 years of follow-up.

There are many possible extensions to this work. First, further biological confirmation of the candidate metabolites (dihydroxy docosatrienoic acid (X-11308) and hydroxyleucine or hydroxyisoleucine (X-11787)) should be performed. Second, the candidate metabolites of interest from this study should be measured in an independent replication sample of African Americans. Third, genetic factors associated with levels of these candidate metabolites can be explored, as well as the genotype-environment interactions as they influence levels of these metabolites and, ultimately, heart failure. Genome-wide association studies on these incident heart failure–related metabolites will be worthwhile, because they can quantify the association between genome-wide significant single-nucleotide polymorphisms and incident heart failure and therefore help

clarify the whole genetic pathway of heart failure by integrating genetics and metabolomics. Finally, metabolomics is an emerging field, and the degree to which the human serum metabolome may or may not be stable over several decades of freezing at −80°C is unknown. We would expect, however, that the degree to which there is instability in the metabolome would not be related to the development of future heart failure.

The degree to which the results presented here are ethnically specific is unknown. The prevalence of heart failure is higher in African Americans compared with the general population (51, 52), and this difference can be partially explained by a higher prevalence of risk factors, such as hypertension (53). The factors underlying these disparities are not well understood. Genetic factors influence the risk of heart failure (54, 55), and population differences in the frequency and effects of these genetic factors also likely have a role. To our knowledge, there are no published large studies comparing the metabolic profile of African Americans with that of other populations, and this would be an important area for future investigations.

Conclusions

From a panel of 204 stable metabolites, 3 potential novel biomarkers were associated with incident heart failure, independent of traditional risk factors. Further investigation is warranted to test whether serum metabolite measurements help identify candidates for interventions to reduce heart failure

risk and elucidate the biological mechanisms by which certain metabolites promote heart failure.

REFERENCES

1. Bahrami H, Bluemke DA, Kronmal R, et al. Novel metabolic risk factors for incident heart failure and their relationship with obesity: the MESA (Multi-Ethnic Study of Atherosclerosis) Study. *J Am Coll Cardiol*. 2008;51(18):1775–1783.
2. Martins H, Pedro N, Castellano M, et al. Cardio-renal syndrome: the challenge in heart failure treatment [in Portuguese]. *Acta Med Port*. 2011;24(2):285–292.
3. Gottdiener JS, Arnold AM, Aurigemma GP, et al. Predictors of congestive heart failure in the elderly: the Cardiovascular Health Study. *J Am Coll Cardiol*. 2000; 35(6):1628–1637.
4. Wang Y, Tuomilehto J, Jousilahti P, et al. Lifestyle factors in relation to heart failure among Finnish men and women. *Circ Heart Fail*. 2011;4(5):607–612.
5. Hu G, Jousilahti P, Antikainen R, et al. Joint effects of physical activity, body mass index, waist circumference, and waist-to-hip ratio on the risk of heart failure. *Circulation*. 2010;121(2):237–244.
6. Blecker S, Matsushita K, Kottgen A, et al. High-normal albuminuria and risk of heart failure in the community. *Am J Kidney Dis*. 2011;58(1):47–55.
7. Kannel WB, Belanger AJ. Epidemiology of heart failure. *Am Heart J*. 1991;121(3 Pt 1):951–957.
8. Levy D, Larson MG, Vasan RS, et al. The progression from hypertension to congestive heart failure. *JAMA*. 1996;275(20): 1557–1562.
9. Eriksson H, Svardsudd K, Larsson B, et al. Risk factors for heart failure in the general population: the study of men born in 1913. *Eur Heart J*. 1989;10(7):647–656.
10. Ingelsson E, Arnlov J, Sundstrom J, et al. Novel metabolic risk factors for heart failure. *J Am Coll Cardiol*. 2005;46(11): 2054–2060.
11. Kenchaiah S, Evans JC, Levy D, et al. Obesity and the risk of heart failure. *N Engl J Med*. 2002;347(5):305–313.
12. Ho KK, Pinsky JL, Kannel WB, et al. The epidemiology of heart failure: the Framingham Study. *J Am Coll Cardiol*. 1993;22(4 suppl A):6A–13A.
13. Kannel WB, Hjortland M, Castelli WP. Role of diabetes in congestive heart failure: the Framingham Study. *Am J Cardiol*. 1974;34(1):29–34.
14. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med*. 2011;17(4): 448–453.
15. Tuunanen H, Ukkonen H, Knuuti J. Myocardial fatty acid metabolism and cardiac performance in heart failure. *Curr Cardiol Rep*. 2008;10(2):142–148.
16. Obrzut S, Tiongson J, Jamshidi N, et al. Assessment of metabolic phenotypes in patients with non-ischemic dilated cardiomyopathy undergoing cardiac resynchronization therapy. *J Cardiovasc Transl Res*. 2010;3(6):643–651.
17. Mensah GA, Mokdad AH, Ford ES, et al. State of disparities in cardiovascular health in the United States. *Circulation*. 2005;111(10):1233–1241.
18. Clark LT. Issues in minority health: atherosclerosis and coronary heart disease in African Americans. *Med Clin North Am*. 2005;89(5):977–1001.
19. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*. 1989;129(4):687–702.
20. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *Am J Clin Nutr*. 1982;36(5):936–942.
21. Din-Dzietham R, Liao D, Diez-Roux A, et al. Association of educational achievement with pulsatile arterial diameter change of the common carotid artery: the Atherosclerosis Risk in Communities (ARIC) Study, 1987–1992. *Am J Epidemiol*. 2000;152(7):617–627.
22. Sharrett AR, Hubbard LD, Cooper LS, et al. Retinal arteriolar diameters and elevated blood pressure: the Atherosclerosis Risk in Communities Study. *Am J Epidemiol*. 1999;150(3): 263–270.
23. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2011;34(1 suppl): 62S –69S.
24. Casale PN, Devereux RB, Alonso DR, et al. Improved sex-specific criteria of left ventricular hypertrophy for clinical and computer interpretation of electrocardiograms: validation with autopsy findings. *Circulation*. 1987;75(3):565–572.
25. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009; 150(9):604–612.
26. Perkel JM. *Metabolomics*. South San Francisco, CA: Biocompare; 2011. (http://www.biocompare.com/Articles/

EditorialArticle/1275/Metabolomics.html). (Accessed January 4, 2012).

27. Suhre K, Meisinger C, Doring A, et al. Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS One*. 2010;5(11):e13953.

28. Floegel A, Drogan D, Wang-Sattler R, et al. Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. *PLoS One*. 2011;6(6):e21103.

29. Eriksson H, Caidahl K, Larsson B, et al. Cardiac and pulmonary causes of dyspnoea—validation of a scoring test for clinical-epidemiological use: the study of men born in 1913. *Eur Heart J*. 1987;8(9):1007–1014.

30. Wilhelmsen L, Eriksson H, Svardsudd K, et al. Improving the detection and diagnosis of congestive heart failure. *Eur Heart J*. 1989;10(C suppl):13S–18S.

31. Loehr LR, Rosamond WD, Chang PP, et al. Heart failure incidence and survival (from the Atherosclerosis Risk in Communities Study). *Am J Cardiol*. 2008;101(7):1016–1022.

32. Butler J, Kalogeropoulos A, Georgiopoulou V, et al. Incident heart failure prediction in the elderly: the health ABC heart failure score. *Circ Heart Fail*. 2008;1(2):125–133.

33. Saunders JT, Nambi V, de Lemos JA, et al. Cardiac troponin T measured by a highly sensitive assay predicts coronary heart disease, heart failure, and mortality in the Atherosclerosis Risk in Communities Study. *Circulation*. 2011;123(13):1367–1376.

34. D'Agostino RB Sr, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*. 2008;117(6):743–753.

35. Chang A, Kramer H. Should eGFR and albuminuria be added to the Framingham risk score? Chronic kidney disease and cardiovascular disease risk prediction. *Nephron Clin Pract*. 2011;119(2):c171–c177; discussion c7–c8.

36. Rutkowski B, Rutkowski P, Slominska E, et al. Cellular toxicity of nicotinamide metabolites. *J Ren Nutr*. 2012;22(1):95–97.

37. Perazella MA. Renal vulnerability to drug toxicity. *Clin J Am Soc Nephrol*. 2009;4(7):1275–1283.

38. Schoenfeld D. Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika*. 1980;67(1):145–153.

39. Sankoh AJ, Huque MF, Dubey SD. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Stat Med*. 1997;16(22):2529–2542.

40. Meijers BK, Claes K, Bammens B, et al. p-Cresol and cardiovascular risk in mild-to-moderate kidney disease. *Clin J Am Soc Nephrol*. 2010;5(7):1182–1189.

41. Winchester JF, Hostetter TH, Meyer TW. p-Cresol sulfate: further understanding of its cardiovascular disease potential in CKD. *Am J Kidney Dis*. 2009;54(5):792–794.

42. Mayr M. Metabolomics: Ready for the prime time? *Circ Cardiovasc Genet*. 2008;1(1):58–65.

43. Yamagishi K, Nettleton JA, Folsom AR. Plasma fatty acid composition and incident heart failure in middle-aged adults: the Atherosclerosis Risk in Communities (ARIC) Study. *Am Heart J*. 2008;156(5):965–974.

44. Kazemian P, Kazemi-Bajestani SM, Alherbish A, et al. The use of omega-3 poly-unsaturated fatty acids in heart failure: a preferential role in patients with diabetes. *Cardiovasc Drugs Ther*. 2012;26(4):311–320.

45. Fu SL, Dean RT. Structural characterization of the products of hydroxyl-radical damage to leucine and their detection on proteins. *Biochem J*. 1997;324(Pt 1):41–48.

46. Grieve DJ, Shah AM. Oxidative stress in heart failure. More than just damage. *Eur Heart J*. 2003;24(24):2161–2163.

47. Dalla Libera L, Ravara B, Gobbo V, et al. Skeletal muscle myofibrillar protein oxidation in heart failure and the protective effect of carvedilol. *J Mol Cell Cardiol*. 2005;38(5):803–807.

48. Tangney CC, Shekelle RB, Raynor W, et al. Intra- and interindividual variation in measurements of beta-carotene, retinol, and tocopherols in diet and plasma. *Am J Clin Nutr*. 1987;45(4):764–769.

49. Huffman KM, Shah SH, Stevens RD, et al. Relationships between circulating metabolic intermediates and insulin action in overweight to obese, inactive men and women. *Diabetes Care*. 2009;32(9):1678–1683.

50. Lewis GD, Asnani A, Gerszten RE. Application of metabolomics to cardiovascular biomarker and pathway discovery. *J Am Coll Cardiol*. 2008;52(2):117–123.

51. Chobanian AV, Bakris GL, Black HR, et al. The seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 Report. *JAMA*. 2003;289(19):2560–2572.

52. Yancy CW. Heart failure in African Americans. *Am J Cardiol*. 2005;96(7B):3i–12i.

53. Schoots AC, Dijkstra JB, Ringoir SM, et al. Are the classical markers sufficient to describe uremic solute accumulation in dialyzed patients? Hippurates reconsidered. *Clin Chem*. 1988;34(6):1022–1029.

54. Lusis AJ, Mar R, Pajukanta P. Genetics of atherosclerosis. *Annu Rev Genomics Hum Genet*. 2004;5:189–218.

55. Smith NL, Felix JF, Morrison AC, et al. Association of genome-wide variation with the risk of incident heart failure in adults of European and African ancestry: a prospective meta-analysis from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. *Circ Cardiovasc Genet*. 2010;3(3):256–266.