



Published in final edited form as:

*Genet Epidemiol.* 2013 December ; 37(8): 840–845. doi:10.1002/gepi.21752.

## Genome-Wide Association Study of a Heart Failure-Related Metabolomic Profile among African Americans in the Atherosclerosis Risk in Communities (ARIC) Study

Bing Yu, MD, MS<sup>1</sup>, Yan Zheng, MD, MPH<sup>1</sup>, Danny Alexander, PhD<sup>2</sup>, Teri A. Manolio, MD, PhD<sup>3</sup>, Alvaro Alonso, MD, PhD<sup>4</sup>, Jennifer A. Nettleton, PhD<sup>1</sup>, and Eric Boerwinkle, PhD<sup>1,5</sup>

<sup>1</sup>Division of Epidemiology, Human Genetics and Environmental Sciences, University of Texas Health Science Center at Houston, Houston, TX

<sup>2</sup>Metabolon, Inc., Durham, NC

<sup>3</sup>Division of Genomic Medicine, National Human Genome Research Institute, Bethesda, MD

<sup>4</sup>Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN

<sup>5</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX

### Abstract

Both the prevalence and incidence of heart failure (HF) are increasing, especially among African-Americans, but no large-scale, genome-wide association study (GWAS) of HF-related metabolites have been reported. We sought to identify novel genetic variants that are associated with metabolites previously reported to relate to HF incidence. GWASs of three metabolites identified previously as risk factors for incident HF (pyroglutamine, dihydroxy docosatrienoic acid and X-11787, being either hydroxy-leucine or hydroxy-isoleucine) were performed in 1260 African-Americans free of HF at the baseline examination of the Atherosclerosis Risk in Communities (ARIC) study. A significant association on chromosome 5q33 (rs10463316, MAF = 0.358, p-value =  $1.92 \times 10^{-10}$ ) was identified for pyroglutamine. One region on chromosome 2p13 contained a nonsynonymous substitution in N-acetyltransferase 8 (*NAT8*) was associated with X-11787 (rs13538, MAF = 0.481, p-value =  $1.71 \times 10^{-23}$ ). The smallest p-value for dihydroxy docosatrienoic acid was rs4006531 on chromosome 8q24 (MAF = 0.400, p-value =  $6.98 \times 10^{-7}$ ). None of the above SNPs were individually associated with incident HF, but a genetic risk score (GRS) created by summing the most significant risk alleles from each metabolite detected 11% greater risk of HF per allele. In summary, we identified three loci associated with previously reported HF-related metabolites. Further use of metabolomics technology will facilitate replication of these findings in independent samples.

---

Address correspondence to: Eric Boerwinkle, PhD, Human Genetic Center, University of Texas School of Public Health, 1200 Herman Pressler E-447, Houston, TX 77030, Phone: 713.500.9816, Fax: 713.500.0900, Eric.Boerwinkle@uth.tmc.edu.

#### Disclosures

None.

## Keywords

metabolomics; genome-wide association; African-Americans; heart failure

Heart failure (HF) is a leading cause of hospitalization among the elderly, and its incidence and prevalence continue to rise[Haldeman, Croft, Giles, & Rashidee, 1999; Kannel, 2000; Rich, 1997; Roger et al., 2011]. The pathogenesis of HF includes genetic and environmental factors, and understanding the role of genetic variations and their interactions with the environment may improve our understanding of HF onset and progression. Genome-wide association studies (GWAS) have proven to be a powerful tool for identifying genes and genomic regions having common sequence variation affecting a trait of interest, such as HF[Hirschhorn & Daly, 2005].

Metabolomics is the high-throughput study of the small molecule end products of a variety of chemical processes in a biologic system[Nicholson, Lindon, & Holmes, 1999]. Metabolites are the ultimate downstream product of gene function and environmental exposures[van der Greef, Stroobant, & van der Heijden, 2004], and thus, may enhance our understanding of the role of both genes and the environment in the etiology of HF. A number of studies have reported differences in the metabolome between cases and controls for a number of common chronic diseases, including cardiovascular disease and HF[Dunn et al., 2007; Griffin, Atherton, Shockcor, & Atzori, 2011; Tuunanen, Ukkonen, & Knuuti, 2008; T. J. Wang et al., 2011; Z. Wang et al., 2011]. Recent studies combining genetics and metabolomics have provided novel functional insights related to several chronic diseases, including cardiovascular disease and type 2 diabetes[Suhre et al., 2011]. To date, no GWAS has been performed to interrogate the collective roles of genetics and the metabolome in the onset of incident HF. Zheng et al.[Zheng et al., 2013] identified three metabolites that were significantly associated with incident HF among African-Americans, including pyroglutamine, dihydroxy docosatrienoic acid, and the unnamed compound ( X-11787), which was revealed to be an isoform of either hydroxy-leucine or hydroxy-isoleucine. Identifying genetic factors that influence the levels of these novel metabolites may provide insights into their possible identity and function.

## Methods

### Study populations

The ARIC study is a prospective cohort study designed to ascertain the etiology and predictors of cardiovascular disease (CVD), which enrolled 15,792 middle-aged adults from four U.S. communities (Forsyth County, NC; Jackson, MS; suburbs of Minneapolis, MN; and Washington County, MD) between 1987–1989. A detailed description of the ARIC study design and methods was published elsewhere[The ARIC investigators, 1989]. Metabolomic profiles were measured in African-Americans randomly selected from the Jackson, MS field center, who were included in this genome-wide association study of HF-related metabolites (N = 1,260). African Americans from both Jackson, MS and Forsyth County, NC free of prevalent HF were analyzed for the associations between incident HF and top ranking genome-wide significant variants (N = 2,225). Participants were excluded if

they had prevalent HF, if they were first degree relatives of someone else in the study, if there were sample handling errors or discrepancies in race or sex between reported data and genotype data, or if they did not give consent for use of DNA information.

### Assessment of Metabolites

Metabolite profiling was completed in June 2010 using fasting serum samples which had been stored at  $-80^{\circ}$  since collection at the baseline examination in 1986–7. In total, detection and quantification of 602 metabolites was completed by Metabolon Inc. (Durham, USA) using an untargeted, gas chromatography-mass spectrometry and liquid chromatography-mass spectrometry (GC-MS and LC-MS)-based quantification protocol, the details of which were described elsewhere [Zheng et al., 2013].

### Follow-up and HF Ascertainment

Incident HF was defined during follow-up by either (1) a first hospitalization which included an International Classification of Diseases, 9<sup>th</sup> revision, discharge code of 428 (428.0 to 428.9) in any position for those without a prior HF hospitalization, or (2) a death certificate with a 428 (HF) or International Classification of Diseases, 10<sup>th</sup> revision, code I50 (HF) in any position [Loehr, Rosamond, Chang, Folsom, & Chambless, 2008; White et al., 1996]. Individuals were followed up for HF events from enrollment (baseline) until death or December 31, 2008, and those who were lost to follow-up were censored at the date of last contact.

### Genotyping and Imputation

In the ARIC study, autosomal single-nucleotide polymorphisms (SNPs) were genotyped on the Affymetrix 6.0 chip and were imputed to  $\approx 2.5$  million SNPs based on a panel of cosmopolitan reference haplotypes from HapMap CEU and YRI. MACH v1.0 was used to do imputation and allele dosage information was summarized in the imputation results. SNPs were excluded if they had no chromosomal location, were monomorphic, had a call rate  $< 95\%$ , or had a Hardy-Weinberg equilibrium  $p$ -value  $< 10^{-6}$ . For each SNP, the ratio of the observed versus expected variance of the dosage were served as the measure of imputation quality.

### Statistical Analyses

The three incident HF-related metabolites identified by Zheng et al. [Zheng et al., 2013] were treated as continuous variables, and covariates in the model were measured at the time when the serum samples were obtained in which metabolite levels were later measured. Linear regression analyses were applied to each metabolite respectively, adjusting for age, sex and the first 10 principal components derived from the principal components analysis to account for population stratification [Price et al., 2006]. Individuals with metabolite levels that were below the detectable limit of the assay were assigned the lowest detected value for that metabolite in all samples, and all metabolite values were natural log-transformed prior to analysis. SNP effects were estimated under an additive genetic model. SNPs with minor allele frequency (MAF)  $< 5\%$  were excluded. Quantile-quantile (QQ) plots were generated for each analysis to illustrate the distribution of the observed and expected  $p$ -values for all

eligible SNPs, and regional plots showing linkage disequilibrium (LD) and the location of nearby genes (if any) were generated for the top ranking SNPs for each metabolite. Genome-wide significance was defined as a p-value  $< 5 \times 10^{-8}$ , and a p-value  $< 1 \times 10^{-5}$  was considered suggestive evidence for association. If more than one significant or suggestive SNP clustered at a locus, the SNP with the smallest p-value was reported as the sentinel marker. The analyses were performed by ProbABEL[Aulchenko, Struchalin, & van Duijn, 2010] and R ([www.r-project.org](http://www.r-project.org)).

To evaluate the cumulative effect of the identified genetic variants, a genetic risk score (GRS) was constructed by summing the number of risk raising alleles (0/1/2) for the top ranking genome-wide significant SNP of each metabolite. The proportional hazards assumption was examined and not rejected according to our assessment using the methods developed by Grambsch and Therneau[Grambsch & Therneau, 1994] and these analyses were performed using R ([www.r-project.org](http://www.r-project.org)).

## Results

A total of 1,260 African-Americans were involved in the genome-wide association analyses, and their baseline characteristics are shown in Table 1. In this study, pyroglutamine and dihydroxy docosatrienoic acid had at least one locus that reached genome-wide significance (p-value  $< 5 \times 10^{-8}$ ). The Manhattan plots, QQ plots and detailed information about statistically significant and suggestive loci are provided in the Supplement Table 1–3 and Figure 1–2.

For pyroglutamine, the most significant SNP that exceeded the genome-wide significant threshold was rs10463316. (MAF = 0.358, p-value =  $1.92 \times 10^{-10}$ ) (Figure 1A). rs10463316 is an intergenic SNP located on chromosome 5q33, 18.93 kb from the *SLC36A2* gene (solute carrier family 36, member 2).

One region encompassing two genes on chromosome 2p13 was significantly associated with X-11787 levels. The top ranking SNP is intergenic (rs6546857, MAF = 0.477, p-value =  $9.58 \times 10^{-24}$ ) 0.91 kb from *ALMS1* (Alstrom syndrome 1). The other gene in this same region is *NAT8* (N-acetyltransferase 8). The sentinel SNP in *NAT8* is a missense variant rs13538 (MAF = 0.481, p-value =  $1.71 \times 10^{-23}$ ), which causes a serine to phenylalanine substitution (F143S) within the acetyltransferase domain of N-acetyltransferase 8. These two SNPs (rs6546857 and rs13538) are in strong linkage disequilibrium (LD) with  $r^2 = 0.8$  (Figure 1B). After conditioning on the missense variant, no other SNP in this region was shown a statistically significant relation to dihydroxy docosatrienoic acid levels (data not shown).

The SNP with smallest p-value for dihydroxy docosatrienoic acid was rs4006531 on chromosome 8q24 (MAF = 0.400, p-value =  $6.98 \times 10^{-7}$ ), 90.48 kb from a hypothetical gene *LOC10013023*. No SNP reached the genome-wide significant threshold for this unnamed metabolite.

The top ranking SNP for each metabolite was chosen to construct a GRS in 2,225 African-Americans free of HF at the baseline examination who had been monitored for the onset of

HF for up to 22 years (396 HF events). Baseline characteristics of these individuals are shown in Table 1. Even though none of these SNPs was significantly related to incident HF individually, the association between the cumulative GRS and incident HF was statistically significant in a Cox proportional hazards model after adjusting for traditional risk factors. (HR = 1.11, 95% CI: 1.02–1.22, p-value = 0.019, Table 2). When a model is fit, that includes the traditional risk factors, the three metabolites and the GRS, then the GRS is no longer statistically significant (data not shown), reinforcing that the metabolites are likely a mediator of these genetic effects.

## Discussion

This study utilized an untargeted metabolomics approach combined with genome-wide association screening in a large, well-defined sample of African-Americans to identify two genetic loci that influence two HF-related metabolic traits. The gene most likely influencing pyroglutamine levels is *SLC36A2*, which is an electrogenic amino acid symporter for amino acids with small side chains, especially glutamine [Bode, 2001; Boll, Foltz, Rubio-Aliaga, Kottra, & Daniel, 2002]. Pyroglutamine is a cyclic derivative of glutamine and it is plausible, but not proven, that variants in *SLC36A2* may affect pyroglutamine transport. A recent family-based study reported interaction between *SLC36A2* and *SLC6A20*, a proline imino transporter, on the onset of iminoglycinuria [Broer et al., 2008]. Our results provide new insights into *SLC36A2* function at a population level.

We previously explored the identity of metabolite X-11787 which has a likely chemical formula of  $C_6H_{13}NO_3$ , consistent with the chemical structures of either hydroxy-leucine or hydroxy-isoleucine [Zheng et al., 2013]. Hydroxy-leucines are oxidized from leucine or hydroperoxy-leucines and provide useful *in vivo* markers of protein oxidation [Fu & Dean, 1997]. Protein oxidation is implicated in aging and oxidative stress [Berlett & Stadtman, 1997], which is associated with a number of diseases, such as atherosclerosis, hypertension, diabetes, and chronic kidney disease [de Champlain et al., 2004; Maritim, Sanders, & Watkins, 2003; Singh & Jialal, 2006; Small, Coombes, Bennett, Johnson, & Gobe, 2012]. The genetic locus influencing X-11787 contains two genes on chromosome 2p13, *ALMS1* and *NAT8*. Mutations in *ALMS1* are reported to cause Alström syndrome, a rare autosomal recessive disease sharing several features with the metabolic syndrome, namely obesity, hyperinsulinemia, and hypertriglyceridemia [Joy et al., 2007], *NAT8* participates in the development and maintenance of normal kidney and liver function [Ozaki, Fujiwara, Nakamura, & Takahashi, 1998]. Recent GWAS report that this locus is associated with kidney function and chronic kidney disease [Chambers et al., 2010; Kottgen et al., 2010]. Given *NAT8*'s function and the fact that X-11787 has an independent association with incident HF after adjusting for BMI, glucose levels, lipid levels and kidney function, we hypothesize that *NAT8* influences an underlying, but as yet unknown, process that affects both HF and kidney disease.

To date, only two studies have used a genome-wide association study approach to localize genes (or gene regions) associated with incident HF, one including both European Americans and African Americans and the other in European Americans only [Larson et al., 2007; Smith et al., 2010]. In this sample of African-Americans from the ARIC study, none

of the top genetic loci associated with HF were associated with any of the three metabolites reported previously to be related to incident HF [Zheng et al., 2013]. Therefore, the loci identified in earlier GWAS most likely have effects outside of biologic pathways represented in the serum metabolome measured here. The general concept for the occurrence of common diseases is considered as a result of complex interactions between multiple genetic and environmental predisposing factors [Hirschhorn & Daly, 2005]. Several recently studies have shown common genetic variants have joint effects on cardiovascular diseases, such as CHD [Anderson et al., 2010; Morrison et al., 2007], however such studies on HF is rare. Our study underscores the potential multifactorial nature of HF.

### Strengths and Limitations

This study, to our knowledge, is the first GWAS to reveal genetic risk variants for human metabolomic profile in African-Americans. In addition, it is the first study to estimate genetic effects on HF risk via disease related metabolites. Our study also has several limitations. First, there are no appropriate African-American sample sets with metabolomic profiles or enough incident HF events that can be used for replication. Second, although strong signals were detected at several genetic loci, the generalizability of these results has not been established. Through our internal quality control process, we limited our analyses to those metabolites with few missing values, and the remaining missing values were imputed with the lowest observed value. If the very low values were due to genetic variation, sensitivity of a metabolomic technology may impact the statistical power of these analyses.

### Conclusions

From a panel of three HF-related metabolites, we identified two loci that reached genome-wide significance for two of the metabolites. These findings contribute to the knowledge-base of HF physiology and to our understanding of the human metabolic profile. Further use of metabolomics technology should enable replication of these findings.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

We acknowledge the essential role of the Atherosclerosis Risk in Communities (ARIC) Study in developing and support for this article. The authors also thank the staff and participants of the ARIC study for their important contributions. The metabolomics research was sponsored by National Human Genome Research Institute (NHGRI). B.Y. and Y.Z. are supported in part by a training fellowship from Burroughs Wellcome Fund – The Houston Laboratory and Population Science Training Program in Gene-Environment Interaction (BWF Grant No. 1008200). J.A.N. is supported by a K01 from the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases (5K01DK082729-04).

#### Sources of Funding

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute grant U01HG004402; and National Institutes of Health contract HHSN268200625226C. Infrastructure was

partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

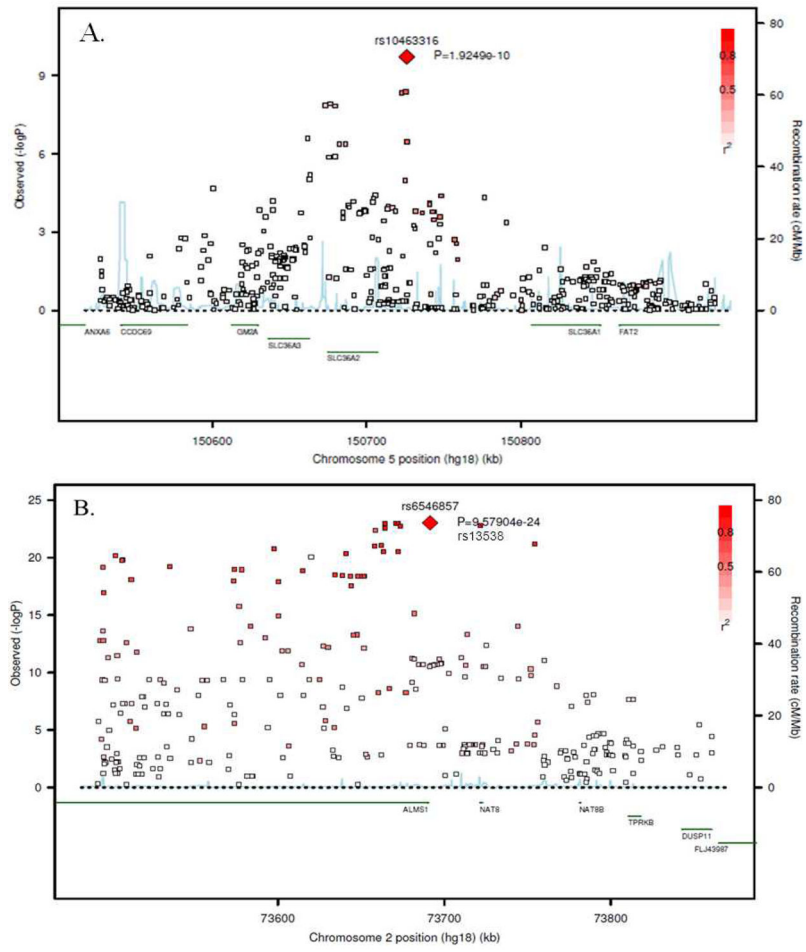
## References

- Anderson JL, Horne BD, Camp NJ, Muhlestein JB, Hopkins PN, Cannon-Albright LA, Mower CP, Park JJ, Clarke JL, Nicholas ZP, McKinney JT, Carlquist JF. Joint effects of common genetic variants from multiple genes and pathways on the risk of premature coronary artery disease. *Am Heart J*. 2010; 160(2):250–256. e253. [PubMed: 20691829]
- Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics*. 2010; 11:134. [PubMed: 20233392]
- Berlett BS, Stadtman ER. Protein oxidation in aging, disease, and oxidative stress. *J Biol Chem*. 1997; 272(33):20313–20316. [PubMed: 9252331]
- Bode BP. Recent molecular advances in mammalian glutamine transport. *J Nutr*. 2001; 131(9 Suppl): 2475S–2485S. discussion 2486S–2477S. [PubMed: 11533296]
- Boll M, Foltz M, Rubio-Aliaga I, Kottra G, Daniel H. Functional characterization of two novel mammalian electrogenic proton-dependent amino acid cotransporters. *J Biol Chem*. 2002; 277(25): 22966–22973. [PubMed: 11959859]
- Broer S, Bailey CG, Kowalczyk S, Ng C, Vanslambrouck JM, Rodgers H, Auray-Blais C, Cavanaugh JA, Broer A, Rasko JE. Iminoglycinuria and hyperglycinuria are discrete human phenotypes resulting from complex mutations in proline and glycine transporters. *J Clin Invest*. 2008; 118(12): 3881–3892. [PubMed: 19033659]
- Chambers JC, Zhang W, Lord GM, van der Harst P, Lawlor DA, Sehmi JS, Gale DP, Wass MN, Ahmadi KR, Bakker SJ, Beckmann J, Bilo HJ, Bochud M, Brown MJ, Caulfield MJ, Connell JM, Cook HT, Cotlarciuc I, Davey Smith G, de Silva R, Deng G, Devuyst O, Dikkeschei LD, Dimkovic N, Dockrell M, Dominiczak A, Ebrahim S, Eggermann T, Farrall M, Ferrucci L, Floege J, Forouhi NG, Gansevoort RT, Han X, Hedblad B, Homan van der Heide JJ, Hepkema BG, Hernandez-Fuentes M, Hyponen E, Johnson T, de Jong PE, Kleefstra N, Lagou V, Lapsley M, Li Y, Loos RJ, Luan J, Luttrupp K, Marechal C, Melander O, Munroe PB, Nordfors L, Parsa A, Peltonen L, Penninx BW, Perucha E, Pouta A, Prokopenko I, Roderick PJ, Ruokonen A, Samani NJ, Sanna S, Schalling M, Schlessinger D, Schlieper G, Seelen MA, Shuldiner AR, Sjogren M, Smit JH, Snieder H, Soranzo N, Spector TD, Stenvinkel P, Sternberg MJ, Swaminathan R, Tanaka T, Ubink-Veltmaat LJ, Uda M, Vollenweider P, Wallace C, Waterworth D, Zerres K, Waeber G, Wareham NJ, Maxwell PH, McCarthy MI, Jarvelin MR, Mooser V, Abecasis GR, Lightstone L, Scott J, Navis G, Elliott P, Kooner JS. Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet*. 2010; 42(5):373–375. [PubMed: 20383145]
- de Champlain J, Wu R, Girouard H, Karas M, AELM, Laplante MA, Wu L. Oxidative stress in hypertension. *Clin Exp Hypertens*. 2004; 26(7–8):593–601. [PubMed: 15702613]
- Dunn WB, Broadhurst DI, Deepak SM, Buch MH, McDowell G, Spasic I, Ellis DI, Brooks N, Kell DB, Neyses L. Serum metabolomics reveals many novel metabolic markers of heart failure, including pseudouridine and 2-oxoglutarate. *Metabolomics*. 2007; 3(4):413–426.
- 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. [PubMed: 9164839]
- Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994; 81:515–526.
- Griffin JL, Atherton H, Shockcor J, Atzori L. Metabolomics as a tool for cardiac research. *Nat Rev Cardiol*. 2011; 8(11):630–643. [PubMed: 21931361]
- Haldeman GA, Croft JB, Giles WH, Rashidee A. Hospitalization of patients with heart failure: National Hospital Discharge Survey, 1985 to 1995. *Am Heart J*. 1999; 137(2):352–360. [PubMed: 9924171]
- Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*. 2005; 6(2):95–108. [PubMed: 15716906]
- Joy T, Cao H, Black G, Malik R, Charlton-Menys V, Hegele RA, Durrington PN. Alstrom syndrome (OMIM 203800): a case report and literature review. *Orphanet J Rare Dis*. 2007; 2:49. [PubMed: 18154657]

- Kannel WB. Incidence and epidemiology of heart failure. *Heart Fail Rev.* 2000; 5(2):167–173. [PubMed: 16228142]
- Kottgen A, Pattaro C, Boger CA, Fuchsberger C, Olden M, Glazer NL, Parsa A, Gao X, Yang Q, Smith AV, O’Connell JR, Li M, Schmidt H, Tanaka T, Isaacs A, Ketkar S, Hwang SJ, Johnson AD, Dehghan A, Teumer A, Pare G, Atkinson EJ, Zeller T, Lohman K, Cornelis MC, Probst-Hensch NM, Kronenberg F, Tonjes A, Hayward C, Aspelund T, Eiriksdottir G, Launer LJ, Harris TB, Rumpersaud E, Mitchell BD, Arking DE, Boerwinkle E, Struchalin M, Cavalieri M, Singleton A, Giallauria F, Metter J, de Boer IH, Haritunians T, Lumley T, Siscovick D, Psaty BM, Zillikens MC, Oostra BA, Feitosa M, Province M, de Andrade M, Turner ST, Schillert A, Ziegler A, Wild PS, Schnabel RB, Wilde S, Munzel TF, Leak TS, Illig T, Klopp N, Meisinger C, Wichmann HE, Koenig W, Zgaga L, Zemunik T, Kolcic I, Minelli C, Hu FB, Johansson A, Igl W, Zaboli G, Wild SH, Wright AF, Campbell H, Ellinghaus D, Schreiber S, Aulchenko YS, Felix JF, Rivadeneira F, Uitterlinden AG, Hofman A, Imboden M, Nitsch D, Brandstatter A, Kollerits B, Kedenko L, Magi R, Stumvoll M, Kovacs P, Boban M, Campbell S, Endlich K, Volzke H, Kroemer HK, Nauck M, Volker U, Polasek O, Vitart V, Badola S, Parker AN, Ridker PM, Kardina SL, Blankenberg S, Liu Y, Curhan GC, Franke A, Roach T, Paulweber B, Prokopenko I, Wang W, Gudnason V, Shuldiner AR, Coresh J, Schmidt R, Ferrucci L, Shlipak MG, van Duijn CM, Borecki I, Kramer BK, Rudan I, Gyllenstein U, Wilson JF, Witteman JC, Pramstaller PP, Rettig R, Hastie N, Chasman DI, Kao WH, Heid IM, Fox CS. New loci associated with kidney function and chronic kidney disease. *Nat Genet.* 2010; 42(5):376–384. [PubMed: 20383146]
- Larson MG, Atwood LD, Benjamin EJ, Cupples LA, D’Agostino RB Sr, Fox CS, Govindaraju DR, Guo CY, Heard-Costa NL, Hwang SJ, Murabito JM, Newton-Cheh C, O’Donnell CJ, Seshadri S, Vasani RS, Wang TJ, Wolf PA, Levy D. Framingham Heart Study 100K project: genome-wide associations for cardiovascular disease outcomes. *BMC Med Genet.* 2007; 8(Suppl 1):S5. [PubMed: 17903304]
- 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(7):1016–1022. [PubMed: 18359324]
- Maritim AC, Sanders RA, Watkins JB 3rd. Diabetes, oxidative stress, and antioxidants: a review. *J Biochem Mol Toxicol.* 2003; 17(1):24–38. [PubMed: 12616644]
- Morrison AC, Bare LA, Chambless LE, Ellis SG, Malloy M, Kane JP, Pankow JS, Devlin JJ, Willerson JT, Boerwinkle E. Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. *Am J Epidemiol.* 2007; 166(1):28–35. [PubMed: 17443022]
- Nicholson JK, Lindon JC, Holmes E. ‘Metabonomics’: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica.* 1999; 29(11):1181–1189. [PubMed: 10598751]
- Ozaki K, Fujiwara T, Nakamura Y, Takahashi E. Isolation and mapping of a novel human kidney- and liver-specific gene homologous to the bacterial acetyltransferases. *J Hum Genet.* 1998; 43(4):255–258. [PubMed: 9852678]
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38(8):904–909. [PubMed: 16862161]
- Rich MW. Epidemiology, pathophysiology, and etiology of congestive heart failure in older adults. *J Am Geriatr Soc.* 1997; 45(8):968–974. [PubMed: 9256850]
- Roger VL, Go AS, Lloyd-Jones DM, Adams RJ, Berry JD, Brown TM, Carnethon MR, Dai S, de Simone G, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Greenlund KJ, Hailpern SM, Heit JA, Ho PM, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, McDermott MM, Meigs JB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Rosamond WD, Sorlie PD, Stafford RS, Turan TN, Turner MB, Wong ND, Wylie-Rosett J. Heart disease and stroke statistics--2011 update: a report from the American Heart Association. *Circulation.* 2011; 123(4):e18–e209. [PubMed: 21160056]
- Singh U, Jialal I. Oxidative stress and atherosclerosis. *Pathophysiology.* 2006; 13(3):129–142. [PubMed: 16757157]



- Small DM, Coombes JS, Bennett N, Johnson DW, Gobe GC. Oxidative stress, anti-oxidant therapies and chronic kidney disease. *Nephrology (Carlton)*. 2012; 17(4):311–321. [PubMed: 22288610]
- Smith NL, Felix JF, Morrison AC, Demissie S, Glazer NL, Loehr LR, Cupples LA, Dehghan A, Lumley T, Rosamond WD, Lieb W, Rivadeneira F, Bis JC, Folsom AR, Benjamin E, Aulchenko YS, Haritunians T, Couper D, Murabito J, Wang YA, Stricker BH, Gottdiener JS, Chang PP, Wang TJ, Rice KM, Hofman A, Heckbert SR, Fox ER, O'Donnell CJ, Uitterlinden AG, Rotter JJ, Willerson JT, Levy D, van Duijn CM, Psaty BM, Witteman JC, Boerwinkle E, Vasani RS. 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. [PubMed: 20445134]
- Suhre K, Shin SY, Petersen AK, Mohny RP, Meredith D, Wagele B, Altmaier E, Deloukas P, Erdmann J, Grundberg E, Hammond CJ, de Angelis MH, Kastenmuller G, Kottgen A, Kronenberg F, Mangino M, Meisinger C, Meitinger T, Mewes HW, Milburn MV, Prehn C, Raffler J, Ried JS, Romisch-Margl W, Samani NJ, Small KS, Wichmann HE, Zhai G, Illig T, Spector TD, Adamski J, Soranzo N, Gieger C. Human metabolic individuality in biomedical and pharmaceutical research. *Nature*. 2011; 477(7362):54–60. [PubMed: 21886157]
- The ARIC investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*. 1989; 129(4):687–702. [PubMed: 2646917]
- Tuunanen H, Ukkonen H, Knuuti J. Myocardial fatty acid metabolism and cardiac performance in heart failure. *Curr Cardiol Rep*. 2008; 10(2):142–148. [PubMed: 18417015]
- van der Greef J, Stroobant P, van der Heijden R. The role of analytical sciences in medical systems biology. *Curr Opin Chem Biol*. 2004; 8(5):559–565. [PubMed: 15450501]
- Wang TJ, Larson MG, Vasani RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, O'Donnell CJ, Carr SA, Mootha VK, Florez JC, Souza A, Melander O, Clish CB, Gerszten RE. Metabolite profiles and the risk of developing diabetes. *Nat Med*. 2011; 17(4):448–453. [PubMed: 21423183]
- Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung YM, Wu Y, Schauer P, Smith JD, Allayee H, Tang WH, DiDonato JA, Lusis AJ, Hazen SL. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*. 2011; 472(7341):57–63. [PubMed: 21475195]
- White AD, Folsom AR, Chambless LE, Sharret AR, Yang K, Conwill D, Higgins M, Williams OD, Tyroler HA. Community surveillance of coronary heart disease in the Atherosclerosis Risk in Communities (ARIC) Study: methods and initial two years' experience. *J Clin Epidemiol*. 1996; 49(2):223–233. [PubMed: 8606324]
- Zheng Y, Yu B, Alexander D, Manolio TA, Aguilar D, Coresh J, Heiss G, Boerwinkle E, Nettleton JA. Associations between Metabolomic Compounds and Incident Heart Failure among African Americans: the Atherosclerosis Risk in Communities (ARIC) Study. *Am J Epidemiol*. 2013 In press.



**Figure 1.** Regional association plots of the top ranking genome-wide significant markers for two metabolites. (A) Pyroglutamine, (B) X-11787

**Table 1**

Characteristics of ARIC African Americans Participants in the Analyses

	AAs for 3 HF-related Metabolites GWAS	AAs for Genetic Risk Score Analysis	
		HF Free	Incident HF
Participants, n	1260	1829	396
Incident HF, n (per 1000 person-years)	166 (7.9)	-	396 (10.2)
Age, y	52.6 ± 5.6	52.8 ± 5.8	55.2 ± 5.6
Male, n (%)	439 (34.8)	693 (37.9)	161 (40.1)
Prevalent CHD, n (%)	37 (2.9)	38 (2.1)	31 (7.9)
Current smoker, n (%)	326 (25.8)	509 (27.8)	141 (35.6)
Physical activity	6.6 ± 1.4	6.7 ± 1.4	6.2 ± 1.4
Education levels, n (%)			
Grade school or 0 years education	215 (17.1)	287 (15.7)	102 (25.8)
High school, but no degree	250 (19.8)	361 (19.7)	102 (25.8)
High school graduate	240 (19.0)	405 (22.1)	76 (19.2)
Vocational school	65 (5.2)	125 (6.8)	15 (3.8)
College	219 (17.4)	349 (19.1)	67 (16.9)
Graduate or Professional school	182 (14.4)	302 (16.5)	34 (8.6)
BMI, kg/m <sup>2</sup>	29.5 ± 5.9	29.1 ± 5.6	31.6 ± 7.3
SBP, mmHg	127.7 ± 20.8	127.0 ± 19.8	135.3 ± 22.7
Anti-hypertensive medication use, n (%)	475 (37.7)	620 (33.9)	215 (54.3)
Fasting glucose, mmol/L	6.1 ± 2.3	6.1 ± 2.3	8.3 ± 5.0
Prevalent diabetes, n (%)	172 (13.7)	235 (12.8)	157 (39.6)
TC : HDL ratio	4.2 ± 1.5	4.2 ± 1.5	4.8 ± 1.7
Lipid-lowering medication use, n (%)	12 (0.8)	20 (1.1)	6 (1.5)
LVH in electrocardiogram, n (%)	49 (3.9)	63 (3.4)	43 (10.9)
eGFR, mL/min/1.73 m <sup>2</sup>	104.9 ± 17.1	104.8 ± 17.1	97.6 ± 23.8

AA indicates African Americans; HF, heart failure, CHD, coronary heart disease; BMI, body mass index; SBP, systolic blood pressure; TC:HDL ratio, serum total cholesterol to high-density-cholesterol lipoprotein ratio; LVH, left ventricular hypertrophy; and eGFR, estimated glomerular filtration rate.

For continuous variables, mean values ± standard errors are shown.

**Table 2**  
Associations of Incident Heart Failure with Genetic Risk Score SNPs and Genetic Risk Score

Metabolite	SNP	Chr	Position	Closest Gene	Risk Allele (Frequency)	Reference Allele (Frequency)	HR (95%CI)	P-value
pyroglutamine	rs10463316	5	150726227	18.92 kb from <i>SLC36A2</i>	A (0.642)	G (0.358)	1.07 (0.92, 1.24)	0.37
X-11787	rs6546857	2	73691463	0.91 kb from <i>ALMS1</i>	A (0.477)	G (0.523)	1.15 (0.99, 1.33)	0.06
X-11308	rs4006531	8	126936687	90.48 kb from <i>LOC100130231</i>	G (0.600)	A(0.400)	1.10 (0.94, 1.28)	0.24
<b>GRS</b>	-	-	-	-	-	-	<b>1.11 (1.02, 1.22)</b>	<b>0.019</b>

GRS indicates genetic risk score; each model adjusted for age, sex, study site, principal components, prevalent coronary heart disease (CHD), physical activity, education, current smoking status, body mass index (BMI), systolic blood pressure (SBP), anti-hypertensive medication use, fasting glucose, prevalent diabetes, serum total cholesterol to high-density-cholesterol lipoprotein (TC:HDL) ratio, lipid-lowering medication use, left ventricular hypertrophy (LVH) and estimated glomerular filtration rate (eGFR).