ARIC Manuscript Proposal #3834

PC Reviewed: 4/13/21 **Status: _____** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title: Proteomic Signatures of Diabetes-Related Cardiac Dysfunction in the ARIC Study

b. Abbreviated Title (Length 26 characters): **Proteomics of diabetes**

2. Writing Group:

Writing group members: Justin Echouffo-Tcheugui, Jingsha Chen, Chiadi Ndumele, Kunihiro Matsushita, Mary Rooney, Christie Ballantyne, Ron Hoogeveen, Eric Boerwinkle, Bing Yu, Amil Shah, Joe Coresh, Liz Selvin

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. _(pending)_ **[please confirm with your initials electronically or in writing]**

First author: **Justin Echouffo Tcheugui**
Address: 5501 Hopkins Bayview Circle
Baltimore, MD 21224
E-mail: jechouf1@jhmi.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).
Name: Elizabeth Selvin
Address: 2024 E. Monument St. Suite 2-600
Baltimore, MD 21287
E-mail: eselvin@jhu.edu

3. Timeline: We aim to complete the manuscript within 1 year from the time of approval.

4. Rationale:

Diabetes and heart failure (HF) are highly prevalent[1,2] and strongly interrelated.[3,4] Community-based studies have shown that diabetes is a potent risk factor for HF, raising the risk by 2 to 4-fold.[3,4] Animal models have suggested mechanisms linking diabetes to cardiac dysfunction (e.g., glucotoxicity, lipotoxicity, oxidative stress, advanced tissue glycation, hyperinsulinemia), but the pathways leading from diabetes to HF in humans remain poorly characterized.[5–7] Progress in refining our understanding of diabetes-related cardiac dysfunction has been hampered by the lack of specific biological signatures.

Deep profiling using plasma proteomics can help to identify novel and clinically relevant biomarkers that reflect newly discovered disease pathways. Animal-based proteomic studies suggest a variety of molecular pathways linking diabetes and cardiac dysfunction; for example, the aberrant protein O-GlcNAcylation has been identified as a unique and important contributor to adverse remodeling in the diabetic heart.[8–11] Other pathways potentially involved in diabetic heart disease are those related to AMP activated protein kinase (AMPK), peroxisome

proliferator-activated receptor (PPAR), sodium glucose transporter -2 (SGLT2), protein kinase C (PKC), mitogen-activated protein kinase (MAPK), Nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB), nuclear factor erythroid 2–related factor 2 (Nrf2), cyclic adenosine 5′-monophosphate-responsive element modulator (CREM), microRNAs (miRNA), and exosomes.[6] There is a paucity of human and population-based data linking proteomic markers to diabetes-related cardiac dysfunction. Recent data from a population-based proteomic study suggest that there are distinct pathological pathways linking diabetes to HF.[12] However, this prior study had a small sample size (n=1,572), lacked racial/ethnic diversity (only whites were included), relied on a proteomic platform that was limited in the scope (only 92 proteins measured), and did not explore subtypes of HF.[12] Another small study (n=253) that mainly included whites, and used SomaScan assays, identified 10 proteins with significantly different expression in patients with diabetes and HF with preserved ejection fraction (HFpEF), with one of the key proteins being apolipoprotein M.[13]

Using SOMAScan data (~ 5000 proteins) at Visits 3 and 5 (as Visit 2 when available) in the ARIC study, we will examine the specific proteomic signatures associated with 1) prevalent diabetes among individuals without HF, and 2) with incident HF (including overall HF, HFpEF, and HF with reduced ejection fraction [HFrEF]) among individuals with diabetes. This will allow the identification of proteomic signatures that reflect mechanistic pathways specifically associated with diabetes-related cardiac dysfunction.

5. Main Hypothesis/Study Questions:

Main hypothesis: Much of the excess risk of HF in diabetes is explained by alterations in the plasma proteome.

Aim 1: What are the proteomic signatures of prevalent diabetes? In other words, what are the cross-sectional differences in plasma proteome between participants with and without diabetes? The primary analysis will be done in the absence of HF to inform examination of HF incidence.

Aim 2: What are the proteomic risk factors for HF incidence among those with diabetes?

Aim 3: How much of the excess risk of HF associated with diabetes is explained by alterations in the plasma proteome?

Aim 4: Are the findings (Aims 2 and 3) similar for HFpEF and HFrEF?

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Design:

Cross-sectional design for Aim 1, and prospective design for Aims 2-4.

Midlife (Visit 3 and 2 when available) and late life (Visit 5) analyses will be conducted separately. The data will be divided into a 2/3 discovery and 1/3 validation sample.

For the analysis, to adjust for multiple comparisons, we will use a Bonferroni-corrected *P*-value as the significance threshold in a conservative analysis. FDR ($q<0.05$) will be used to examine a broader list of proteins for pathway analyses.

Exclusions:

In addition to the Standard ARIC race-center exclusions, we will exclude individuals with missing relevant SOMAscan data or failed QC, as well as those with missing covariate data. For prospective analyses, exclude prevalent HF at baseline (Visit 3 in the main analysis and Visit 5 in the late life analysis)

Exposures:
SOMAScan proteins (~5,000 proteins measured using SOMAScan aptamer-based proteomics platform). This well be guided by the QC document following the current recommended data cleaning, focus on log protein levels, and use flag 2 to exclude unreliable data (flag2=0 for N = 4877 aptamers excluding flag 1 cat1-3: Fc mouse/Contaminants, or var<0.01 or soma qc cvba>50% at v3 or v5) – this is the suggested list of aptamers to analyze).

Covariates:
Age, sex, race-center, current smoking, body mass index, total cholesterol, HDL-cholesterol, systolic blood pressure, hypertension medication use, degree of glycemic control (HbA_{1C}), fasting blood glucose, use of diabetes medications, estimated glomerular filtration rate (eGFR), and coronary heart disease.

Baseline:
Midlife (Visit 3 and when available Visit 2) and late-life (Visit 5) will be analyzed separately.

Outcomes:
Our outcomes include:
- For Aim 1: prevalent diabetes at baseline
- For Aims 2-4: Incident HF after baseline (including HFpEF and HFrEF)

Statistical Analysis:
We will log-transform protein levels similar to the dementia proteomics analysis.

Aim 1: We will use logistic regression with prevalence of diabetes at baseline as the outcome. The primary analysis will focus on individuals without heart failure since that will be the risk group of interest in the subsequent aims. We will first examine the proteins one at a time adjusted for:

- Model 1 : unadjusted
  Model 2: adjusted for age, sex, race-center
  Model 3: model 2 + eGFRcr-cys
  Model 4: model 3 + family history of diabetes, current smoking, body mass index, total cholesterol, HDL cholesterol, systolic blood pressure, hypertension medication use, physical activity and cardiovascular disease (other than HF).
  Model 5: model 4 + hyperglycemia measures (HbA_{1C} or fasting blood glucose)

Aims 2: Incidence of HF (including HFpEF and HFrEF) will be analyzed using Cox regression models. The proteomic risk factors analyzed will be those identified during the explorations to address Aim 1.
In all the analyses, we will assess protein associations using sequential modelling as follows:

- Model 1 : unadjusted
- Model 2 : adjusted for age, sex and race/center

- Model 3: model 2 + eGFRcr-cys
- Model 4: model 3 + smoking , body mass index, HbA_{1C}, use of diabetes medications, systolic blood pressure, use of blood pressure medications, total-cholesterol, HDL-cholesterol, and coronary heart disease (time-varying in the prospective analyses)
- Model 5: model 4 + targeted assays of NT-pro-BNP and hs-TnT

Aim 3: We will compare the significant proteomic "hits" for aims 1 (diabetes) and 2 (HF risk within diabetes) using a Venn diagram. We will examine the extent to which the association of prevalent diabetes at baseline is attenuated after adjustment for proteomic variables associated with HF risk within diabetes. We may consider a formal mediation analysis but measurement error and multiple comparisons could limit the inferences. We will also develop models for aim 3 in "one stage" (rather than building on aims 1 and 2) using elastic net to allow for a unified unbiased framework. We will test the validity of the results using 10-fold cross validation internally and seek external validation.

Aim 4: We will repeat the analysis in aims 2 and 3 for HFpEF and HFrEF separately and compare hits. We will also formally test for interaction in the protein associations keeping in mind that power could be limited.

For all analyses:
We will conduct cross-sectional analyses at both Visits 3 and Visit 5 (and Visit 2 when available), and prospective analyses using Visits 3 and 5 (and Visit 2 when available) as baselines. The visit 3 analyses will be the primary analysis since they will have greater power and longer follow-up.

Replication within ARIC: In the replication analysis, we will examine which Bonferonni significant "hits" in the development (2/3) sample were also significant in the (1/3) validation sample at $p<0.05/\#hits$ tested. We will scrutinize models within the 2/3 development sample using 10-fold cross-validation before taking them to the 1/3 independent validation sample. We hypothesize that risk factors may vary across the lifecourse, consequently, we will conduct the replication analysis separately for midlife and late-life. We will also compare the overlap (using a Venn diagram) of proteins from midlife and late life HF analyses.

Additional analyses: We will collaborate with other groups to seek external replication as well as add a systems biology approach to enhance our understanding of causal mechanisms.

Mendelian Randomization: We will conduct bi-directional Mendelian randomization analyses using SNPs identified from published GWAS to validate potential causal pathways and their direction (causes of diabetes-related cardiac dysfunction will have protein SNPs that also lead to diabetes-related cardiac dysfunction risk vs. consequences of diabetes-related cardiac dysfunction whereby the related GWAS SNPs will lead to alterations in proteins), as performed in similar ARIC analyses for other conditions (e.g., dementia and kidney disease).

Early vs. late life in ARIC: As the result of the prolonged ARIC follow-up (~30 years from midlife) the risk assessment would benefit from an inclusion of short - and long-term risk factors,

which may not be similar, as exposures vary over time. Consequently, we will assess whether the strength of associations between proteomic risk factor and diabetes-related HF vary by decade of follow-up. We hypothesize that the vast majority of proteomic risk factors in the first decade will also be associated with outcomes in subsequent decades but to a lower extent. As a result, the focus of our predictive models will be on the 10-year incidence of HF.

Prediction: We will explore the incremental value of identified "hits" in predicting HF. We will develop predictive models in the development sample and test their discrimination and calibration in the diabetes validation sample. We will examine 10-year risk of HF among individuals with diabetes as the primary outcome. We use our fully adjusted model (aforementioned Model 3) as the base model to be improved (testing change in C-statistic and categorical NRI).

Limitations:
Soma Scan protein quantification has variable correlation with other methods of protein quantification. Generally, about one third of the time correlations are high, one third medium and one third very low. Inferences should be made keeping in mind this complexity of quantification and the proteins themselves.
The lack of external validation of our results in a different cohort is a limitation of our study. We will attempt to conduct an external validation of our analyses to other cohorts with SOMAScan data.
The assessment of causal pathways using the mendelian randomization approach will be somewhat limited by the paucity of GWAS data specifically on variants associated with diabetes –related cardiac dysfunction.

7.a. Will the data be used for non-CVD analysis in this manuscript? ____ Yes __x__ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = "CVD Research" for non-DNA analysis, and for DNA analysis RES_DNA = "CVD Research" would be used? ____ Yes ____ No
(This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? ____ Yes __x__ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = "No use/storage DNA"? ____ Yes ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status.
ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/ARIC/search.php

___x___ Yes _____ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

No prior proposal on the proteomics of diabetes-related HF. We discussed the work with leaders of proteomic analyses of diabetes (which will focus on incidence) and HF which will focus on the total ARIC sample, not explaining diabetes risk (and will likely publish before this paper). Relevant leaders are included in this proposal to optimize collaboration.

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? __x__ Yes ____ No

11.b. If yes, is the proposal

_x__ A. primarily the result of an ancillary study (list number* _ AS2017.27___)
___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____ _____)

*ancillary studies are listed by number at http://www.cscc.unc.edu/aric/forms/

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PubMed Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

References

1. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and Trends in Diabetes Among Adults in the United States, 1988-2012. *JAMA*. 2015;314:1021–1029.
2. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, et al. Heart Disease and Stroke Statistics—2021 Update: A Report From the American Heart Association. *Circulation*. 2021;143:e254-e743.
3. Aune D, Schlesinger S, Neuenschwander M, Feng T, Janszky I, Norat T, Riboli E. Diabetes mellitus, blood glucose and the risk of heart failure: A systematic review and meta-analysis of prospective studies. *Nutr Metab Cardiovasc Dis*. 2018;28:1081–1091.
4. Ohkuma T, Komorita Y, Peters SAE, Woodward M. Diabetes as a risk factor for heart failure in women and men: a systematic review and meta-analysis of 47 cohorts including 12 million individuals. *Diabetologia*. 2019;62:1550-1560.
5. Bugger H, Abel ED. Molecular mechanisms of diabetic cardiomyopathy. *Diabetologia*.

2014;57:660–71.
6. Jia G, Hill MA, Sowers JR. Diabetic cardiomyopathy: An update of mechanisms contributing to this clinical entity. *Circ. Res*. 2018;122(4):624-638.
7. Kenny HC, Abel ED. Heart Failure in Type 2 Diabetes Mellitus. *Circ Res*. 2019;124:121–141.
8. Banerjee PS, Ma J, Hart GW. Diabetes-associated dysregulation of *O* -GlcNAcylation in rat cardiac mitochondria. *Proc Natl Acad Sci*. 2015;112:6050-5.
9. Ma J, Banerjee P, Whelan SA, Liu T, Wei AC, Ramirez-Correa G, McComb ME, Costello CE, O'Rourke B, Murphy A, Hart GW. Comparative Proteomics Reveals Dysregulated Mitochondrial O-GlcNAcylation in Diabetic Hearts. *J Proteome Res*. 2016;203:282-9.
10. Ramirez-Correa GA, Ma J, Slawson C, Zeidan Q, Lugo-Fagundo NS, Xu M, Shen X, Gao WD, Caceres V, Chakir K, DeVine L, Cole RN, Marchionni L, Paolocci N, Hart GW, Murphy AM. Removal of abnormal myofilament O-GlcNAcylation restores Ca2+sensitivity in diabetic cardiac muscle. *Diabetes*. 2015;64:3573-87.
11. Frustaci A, Ciccosanti F, Chimenti C, Nardacci R, Corazzari M, Verardo R, Ippolito G, Petrosillo N, Fimia GM, Piacentini M. Histological and proteomic profile of diabetic versus non-diabetic dilated cardiomyopathy. *Int J Cardiol*. 2016;203:282-9.
12. Tromp J, Voors AA, Sharma A, Ferreira JP, Ouwerkerk W, Hillege HL, Gomez KA, Dickstein K, Anker SD, Metra M, Lang CC, Ng LL, van der Harst P, van Veldhuisen DJ, van der Meer P, Lam CSP, Zannad F, Sama IE. Distinct Pathological Pathways in Patients With Heart Failure and Diabetes. *JACC Heart Fail*. 2020;8(3):234-242.
13. Hanff TC, Cohen JB, Zhao L, Javaheri A, Zamani P, Prenner SB, Rietzschel E, Jia Y, Walsh A, Maranville J, Wang Z, Adam L, Ramirez-Valle F, Schafer P, Seiffert D, Gordon DA, Cvijic ME, Cappola TP, Chirinos JA. Quantitative Proteomic Analysis of Diabetes Mellitus in Heart Failure With Preserved Ejection Fraction. *JACC Basic to Transl Sci*. 2021;6(2):89-99.