

ARIC Manuscript Proposal #3772

PC Reviewed: 2/9/21

Status: _____

Priority: 2

SC Reviewed: _____

Status: _____

Priority: _____

1.a. Full Title: Validation of candidate protein biomarkers identified in studies using genetic instruments for the directly aptamer-measured levels in ARIC study with pancreatic cancer

b. Abbreviated Title (Length 26 characters): Validation of protein markers

2. Writing Group:

Writing group members:

We will invite all interested ARIC investigators, including

Josef Coresh

Corinne Joshi

Nilanjan Chatterjee

Anna Prizment

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

First author: Lang Wu

Address:

Cancer Epidemiology Division, Population Sciences in the Pacific Program

University of Hawaii Cancer Center, University of Hawaii at Manoa

701 Ilalo Street, Building B, Room 520

Honolulu, HI 96813

Phone: [808-564-5965](tel:808-564-5965)

Fax: 808-586-2982

E-mail: lwu@cc.hawaii.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 A. Platz

Address:

Department of Epidemiology

Johns Hopkins Bloomberg School of Public Health

615 N. Wolfe St., Room E6132

Baltimore, MD 21205

Phone: [410.614.9674](tel:410.614.9674)

Fax: [410.614.2632](tel:410.614.2632)

E-mail: eplatz1@jhu.edu

3. Timeline: It is expected that manuscript will be shared with co-authors within 3 years of the study approval.

4. Rationale:

We recently conducted large scale proteome-wide association studies using genetic instruments of protein quantitative trait loci (pQTL) and genetic data from large consortia Pancreatic Cancer Cohort Consortium (PanScan) and Pancreatic Cancer Case-Control (PANC4) Consortium, from which we identified 38 promising protein markers showing a significant association with pancreatic cancer risk for their genetically predicted levels (Zhu et al, Cancer Epidemiol Biomarkers Prev, 2020; PMID: 32439797). We are in the process of conducting more comprehensive analyses leveraging protein genetic prediction models in blood and other tissues, for which we expect that we will be able to identify additional promising protein biomarkers related to pancreatic cancer risk. For example, based on preliminary analyses of PanScan/PanC4 data we were able to identify six additional proteins associated with pancreatic cancer risk which have not been identified in our work using pQTL as instruments. We propose to further investigate these biomarker candidates in ARIC for their directly measured levels and to evaluate their potential utility for pancreatic cancer risk assessment. We propose to leverage the comprehensive proteome data that are measured in ARIC for the proposed work.

5. Main Hypothesis/Study Questions:

Protein biomarker candidates identified in studies using genetic instruments are associated with pancreatic cancer risk for their directly measured levels, and could be used to improve risk assessment of pancreatic cancer.

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).

Study design:

For promising protein biomarker candidates associated with pancreatic cancer risk identified in our completed and ongoing studies using a design of genetic instruments, we will investigate associations of their directly measured levels with pancreatic cancer risk by using existing data collected in ARIC.

Below is an example table showing eight protein-pancreatic cancer risk associations that are independent of previously identified risk SNPs (using pQTL as instruments for this analysis). We will select top proteins showing a significant association after correcting for multiple comparisons for validating in ARIC.

Protein	Protein full name	Protein-encoding gene	Region for protein encoding gene	Instrument variants	Type of pQTL	OR ^a	Lower bound 95% CI ^a	Upper bound 95% CI ^a	P-value	FDR P-value ^b	P-value after adjusting for risk SNPs ^c
LMA2L	VIP36-like protein	<i>LMAN2L</i>	2q11.2	rs2271893	<i>cis</i>	1.39	1.15	1.68	6.47×10^{-4}	3.17×10^{-2}	7.72×10^{-4}
TM11D	Transmembrane protease serine 11D	<i>TMPRSS11D</i>	4q13.2	rs3197999	<i>trans</i>	1.17	1.06	1.29	1.11×10^{-3}	3.78×10^{-2}	2.44×10^{-3}
IP-10	C-X-C motif chemokine 10	<i>CXCL10</i>	4q21.1	rs11548618	<i>cis</i>	0.79	0.69	0.91	1.19×10^{-3}	3.93×10^{-2}	9.71×10^{-4}
ADH1B	Alcohol dehydrogenase 1B	<i>ADH1B</i>	4q23	rs13085791	<i>trans</i>	1.22	1.08	1.37	1.28×10^{-3}	4.14×10^{-2}	2.81×10^{-3}
STOM	Erythrocyte band 7 integral membrane protein	<i>STOM</i>	9q33.2	rs6770670	<i>trans</i>	1.19	1.07	1.33	1.05×10^{-3}	3.78×10^{-2}	2.27×10^{-3}
TENC1	Tensin-2	<i>TNS2</i>	12q13.13	rs3197999	<i>trans</i>	1.25	1.09	1.42	1.11×10^{-3}	3.78×10^{-2}	2.44×10^{-3}
DOCK9	Dedicator of cytokinesis protein 9	<i>DOCK9</i>	13q32.3	rs3197999	<i>trans</i>	1.32	1.12	1.56	1.11×10^{-3}	3.78×10^{-2}	2.44×10^{-3}
CRBB2	Beta-crystallin B2	<i>CRYBB2</i>	22q11.23	rs3197999	<i>trans</i>	1.52	1.18	1.95	1.11×10^{-3}	3.78×10^{-2}	2.44×10^{-3}

Inclusion/exclusion criteria:

We will include subjects who are cancer free at Visit 2 (when available) or Visit 3, as the biospecimens collected at Visit 2 (pending) or Visit 3 underwent proteome measurement. We will exclude subjects with cancer before Visit 2 (or Visit 3).

Outcome and variables of interest:

Study outcomes will be obtained from the 2015 ARIC pancreatic cancer cases file¹.

Other variables:

Participant id, case status, field center, age at diagnosis (for cases), family history of pancreatic cancer, family history of any cancer, stage of disease, first course of treatment, race (Black or White), height, weight, body mass index, smoking status and pack-years, alcohol consumption, history of type 2 diabetes, history of chronic pancreatitis, physical activity levels [variable for meeting physical activity guidelines].

Data analysis:

A preliminary examination to investigate outliers and appropriate variable transformation will be performed, if needed. We will perform a data reduction step where all protein analytes that have limited variability will be removed from further consideration. We will test the associations of pancreatic cancer risk in the overall study, as well as according to race (White [European ancestry] vs Black [African ancestry]) in exploratory analyses. We will use Cox proportional hazard models. We will adjust for potential confounding variables². Potential non-linear relationships will also be assessed³. We will explore the potential effect of time interval since blood draw on the associations of interest (comparing time interval since blood draw to cancer diagnosis of within 5 years, between 5-10 years, and over 10 years). We will also explore the time-dependent analysis when there are protein levels measured at different time points.

Potential limitations

The protein levels in ARIC are measured using an aptamer based method, and it is not entirely clear whether protein levels generated from this measurement method are highly concordant with those from ELISA-based measurements. On the other hand, based on previous work⁴, an alternative assay, a proximity extension assay method (Olink Bioscience, Uppsala, Sweden) demonstrated strongly correlation between the SOMAscan and Olink platforms. The aptamer based method has the advantage of being able to include a large number of proteins in the study. Importantly, for the genetic instrument based analyses, we will use data of protein levels generated in SOMAscan platform, which is consistent with the measurement method in ARIC. So this should less likely be a significant issue for our study.

The number of pancreatic cancer cases is relatively small (estimated to be 160 incident primary cases). However, with the available sample size, for each quintile increment in protein level we will have sufficient power (>80%) to detect an OR of 1.17 for the overall analyses at an alpha=0.05, or 1.26 at an alpha=0.001. The power for stratified analyses could be low, and we will only assess them in exploratory analyses.

7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? ____ Yes ___×___ No

b. If Yes, is the author aware that the current derived consent file ICTDER05 must be used to exclude persons with a value RES_OTH and/or RES_DNA = “ARIC only” and/or “Not for Profit” ? ___ Yes ___ No

(The 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 current derived consent file ICTDER05 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.csc.c.unc.edu/aric/mantrack/maintain/search/dtSearch.html>

___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 overlap with other manuscript proposals on pancreatic cancer, although other investigators are planning an agnostic approach to identify proteins associated with cancer risk.

These studies include the SomaScan data and include pancreatic cancer among the outcomes:

#3482: Plasma Proteins and All-Cause Mortality in Cancer Survivors in ARIC

#3445: The Association of Systemic inflammation with Mortality Due to Non-Index Cancer in Older Adult Cancer Survivors.

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

___ **A. primarily the result of an ancillary study (list number* 2011.07; 1995.04)**

___ **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 <https://sites.csc.c.unc.edu/aric/approved-ancillary-studies>

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.csc.c.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. Joshi CE, Barber JR, Coresh J, et al. Enhancing the Infrastructure of the Atherosclerosis Risk in Communities (ARIC) Study for Cancer Epidemiology Research: ARIC Cancer. *Cancer Epidemiol Biomarkers Prev.* 2018;27(3):295-305.
2. Maldonado G, Greenland S. Simulation study of confounder-selection strategies. *Am J Epidemiol.* 1993;138(11):923-936.
3. Durrleman S, Simon R. Flexible regression models with cubic splines. *Statistics in medicine.* 1989;8(5):551-561.
4. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. *Nature.* 2018;558(7708):73-79.