ARIC Manuscript Proposal #3676

PC Reviewed: 8/11/20	**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 measured levels in ARIC study with prostate cancer risk and prognosis

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 Joshu
Nilanjan Chatterjee

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
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
Fax: 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 from which we identified 31 promising protein markers showing a significant association with prostate cancer risk for their genetically predicted levels (Wu et al, Cancer Research, 2019; PMID: 31337649). 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 prostate cancer risk, aggressiveness, and survival. We propose to further investigate these biomarker candidates in ARIC for their directly measured levels and to evaluate their potential utility for prostate cancer risk/prognosis 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 prostate cancer risk and prognosis for their directly measured levels, and could be used to improve risk/prognosis assessment of prostate 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 prostate cancer risk, aggressiveness, and survival identified in our completed and ongoing studies using a design of genetic instruments, we will investigate associations of their directly measured levels with prostate cancer risk, aggressiveness, and survival by using existing data collected in ARIC.

Below is an example table showing 22 novel protein-prostate cancer risk associations for proteins whose encoding genes are located at genomic loci at least 500kb away from any GWAS-identified prostate cancer risk variants (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	Index SNP(s)[a]	Distance of gene to the index SNP (kb)	Instrument variants	Type of pQTL	OR[b]	95% CI[b]	*P* value	FDR *P* value[c]	*P* value after adjusting for risk SNP[d]
ATF6A	Cyclic AMP-dependent transcription factor ATF-6 alpha	*ATF6*	1q23.3	rs4845695	6,824	rs8111, rs61738953	trans, trans	0.90	0.86-0.95	1.31×10^{-4}	9.18×10^{-3}	1.31×10^{-4}
NCF-2	Neutrophil cytosol factor 2	*NCF2*	1q25.3	rs199774366	20,932	rs4632248, rs28929474	trans, trans	0.95	0.92-0.97	9.93×10^{-5}	7.29×10^{-3}	NA*
Laminin	Laminin	*LAMC1*	1q25.3	rs199774366	21,377	rs62199218, rs4129858	trans, cis	0.93	0.89-0.97	4.16×10^{-4}	0.03	NA*
RM33	39S ribosomal protein L33_mitochondrial	*MRPL33*	2p23.2	rs13385191	7,106	rs28929474	trans	0.93	0.90-0.96	9.61×10^{-6}	7.43×10^{-4}	9.42×10^{-6}
LRRN1	Leucine-rich repeat neuronal protein 1	*LRRN1*	3p26.2	rs2660753	83,221	rs429358, rs6801789	trans, cis	0.97	0.95-0.99	7.21×10^{-4}	0.04	7.21×10^{-4}
TACT	T-cell surface protein tactile	*CD96*	3q13.13-3q13.2	rs7611694	1,891	rs3132451	trans	1.22	1.16-1.29	1.02×10^{-12}	3.75×10^{-10}	1.02×10^{-12}
IL-21	Interleukin-21	*IL21*	4q27	rs34480284	17,469	rs12368181, rs3129897	trans, trans	1.11	1.06-1.16	7.77×10^{-6}	7.43×10^{-4}	NA*
PDE4D	cAMP-specific 3_5-cyclic phosphodiesterase 4D	*PDE4D*	5q11.2-5q12.1	rs1482679	13,879	rs3132451	trans	1.17	1.12-1.22	1.02×10^{-12}	3.75×10^{-10}	1.02×10^{-12}
GNMT	Glycine N-methyltransferase	*GNMT*	6p21.1	rs4711748	763	rs57736976	cis	0.93	0.89-0.97	6.80×10^{-4}	0.04	2.78×10^{-4}
PIM1	Serine/threonine-protein kinase pim-1	*PIM1*	6p21.2	rs9469899	2,345	rs28929474	trans	0.88	0.83-0.93	9.61×10^{-6}	7.43×10^{-4}	9.42×10^{-6}
WISP-3	WNT1-inducible-signaling pathway protein 3	*WISP3*	6q21	rs2273669	3,090	rs28929474	trans	0.83	0.77-0.90	9.61×10^{-6}	7.43×10^{-4}	9.42×10^{-6}
TPST1	Protein-tyrosine sulfotransferase 1	*TPST1*	7q11.21	rs56232506	18,233	rs313829	cis	1.14	1.06-1.22	5.23×10^{-4}	0.03	5.43×10^{-4}
ARFP2	Arfaptin-2	*ARFIP2*	11p15.4	rs61890184	1,045	rs28929474	trans	1.23	1.12-1.35	9.61×10^{-6}	7.43×10^{-4}	9.42×10^{-6}
GRIA4	Glutamate receptor 4	*GRIA4*	11q22.3	rs1800057	2,291	rs3132451	trans	1.17	1.12-1.22	1.02×10^{-12}	3.75×10^{-10}	1.02×10^{-12}
KLRF1	Killer cell lectin-like receptor subfamily F member 1	*KLRF1*	12p13.31	rs2066827	2,873	rs11708955, rs62143194	trans, trans	1.13	1.05-1.20	5.74×10^{-4}	0.03	5.74×10^{-4}

GPC6	Glypican-6	*GPC6*	13q31.3-13q32.1	rs9600079	20,151	rs28929474	trans	0.81	0.73-0.89	9.61×10^{-6}	7.43×10^{-4}	9.42×10^{-6}
TM149	IGF-like family receptor 1	*IGFLR1*	19q13.12	rs8102476	2,502	rs12459634	cis	1.06	1.02-1.09	7.31×10^{-4}	0.04	4.68×10^{-3}
TIP39	Tuberoinfundibular peptide of 39 residues	*PTH2*	19q13.33	rs2659124	1,428	rs375375234	trans	1.22	1.13-1.32	3.06×10^{-7}	4.99×10^{-5}	2.96×10^{-7}
ZN175	Zinc finger protein 175	*ZNF175*	19q13.41	rs2735839	710	rs28929474	trans	0.91	0.87-0.95	9.61×10^{-6}	7.43×10^{-4}	9.42×10^{-6}
NKp46	Natural cytotoxicity triggering receptor 1	*NCR1*	19q13.42	rs103294	620	rs2278428	cis	1.16	1.06-1.26	9.91×10^{-4}	0.05	9.65×10^{-4}
SNAB	Beta-soluble NSF attachment protein	*NAPB*	20p11.21	rs11480453	7,945	rs429358, rs7658970	trans, trans	0.91	0.86-0.96	9.77×10^{-4}	0.05	9.77×10^{-4}
ZHX3	Zinc fingers and homeoboxes protein 3	*ZHX3*	20q12	rs11480453	8,460	rs1694123	trans	0.79	0.71-0.88	9.38×10^{-6}	7.43×10^{-4}	9.67×10^{-6}

Inclusion/exclusion criteria:
We will include men 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 prostate cancer cases file[1] and will include prostate cancer risk (total prostate cancer), aggressiveness (lethal prostate cancer); fatal prostate cancer; prostate cancer age at diagnosis <=55; prostate cancer age at diagnosis > 55 years; and among men with a prostate cancer diagnosis, survival (prostate-cancer-specific survival)

Other variables:
Participant id, case status (incident, lethal, fatal, case-fatality), field center, age at diagnosis (for cases), family history of prostate cancer, PSA levels at diagnosis (we may use the V2/V3 PSA RFU as a rough measure of earlier PSA level), Gleason Score at diagnosis, stage and grade 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, 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 prostate cancer risk, aggressiveness, and survival in the overall study, as well as according to race (White [European ancestry] vs Black [African ancestry]). We will use Cox proportional hazard models. We will adjust for potential confounding variables [2]. Potential non-linear relationships will also be assessed [3]. We will assess 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).

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 work [4], 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.

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 __×__ 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.cscc.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 prostate cancer, although other investigators are planning an agnostic approach to identify proteins associated with prostate cancer risk, lethal disease, fatal disease, and case-fatality

These studies include the SomaScan data and include prostate 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.cscc.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.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. Joshu 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.