ARIC Manuscript Proposal #3918

PC Reviewed: 9/14/2021 **Status: A** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title: Clonal hematopoiesis of indeterminate potential, circulating proteome, and coronary artery disease

b. Abbreviated Title (Length 26 characters): CHIP, proteome, and CAD

2. Writing Group: Zhi Yu, Sumeet Khetarpal, Michael Honigberg, Akhil Pampana, Amy Lin, Daniel Katz, Usman Tahir, Joshua Weinstock, Siddhartha Jaiswal, Thomas Austin, Peter Libby, Bing Yu, Josef Coresh, Vasan Ramachandran, Adrienne L. Cupples, Gina Peloso, Joshua Bis, Bruce Psaty, Adolfo Correa, Leslie Lange, Wendy Post, Jerome Rotter, Stephen Rich, Russell Tracy, Kent Taylor, Tom Blackwell, James Wilson, Vijay Sankaran, Alexander Bick, Christie M. Ballantyne, Benjamin Ebert, Robert Gerszten, Pradeep Natarajan (last author): others welcome

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

First author: **Zhi Yu, MB, PhD**
Address: Program in Medical and Population Genetics, Broad Institute of Harvard and MIT
75 Ames St,
Cambridge, MA 02142

Phone: 617-583-3861 Fax:
E-mail: zyu@broadinstitute.org

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: **Christie M. Ballantyne, MD**
Address: Department of Medicine, Baylor College of Medicine
One Baylor Plaza, MS BCM285, Suite 524D
Houston, TX 77030

Phone: (713) 798-5034 Fax:
E-mail: cmb@bcm.edu

3. Timeline: 12 months; manuscript submission in summer 2022

4. Rationale:

clonal hematopoiesis of indeterminate potential (CHIP), a pre-malignant state characterized by accumulated somatic mutations (e.g., *DNMT3A, TET2, ASXL1*) that predispose to clonal expansion of hematopoietic stem cells (HSCs), has been established as a causal risk factor for coronary artery disease (CAD) in both human and mice models[1-6]. However, the underlying mechanisms remains largely unknown. The plasma proteome consists of thousands of secreted proteins that involve in the physiological and pathological functions; thus, it can be a reservoir of biomarkers for capturing the biological processes between CHIP and CAD[7,8].

5. Main Hypothesis/Study Questions:

We propose to leverage the vast genomic and proteomic data resources in ARIC as well as 4 other cohorts in TOPMED, Jackson Heart Study (JHS), Multi-Ethnic Study of Atherosclerosis (MESA), Cardiovascular Health Study (CHS), Framingham Heart Study (FHS), to study the interplay between CHIP, plasma proteome, and CAD.

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).

Inclusion criteria:
(i) Have SOMAscan protein measurements available from plasma sample collected at visit 2 or visit 3
(ii) Have CHIP variables called from whole exome sequencing data collected at visit 2 or visit 3

Exclusion criteria:
(i) Non-white and non-black participants.
(ii) Non-white participants in Washington County and Minnesota.

Exposure variables:
CHIP variables including presence of CHIP, presence of CHIP with variant allele fraction (VAF) > 10%, in *DMNT3A* somatic mutation carriers with CHIP, in *TET2* carriers with CHIP, and in *ASXL1* carriers with CHIP

Outcome variables:
(i) ~5000 proteins per person measured using plasma collected at visit 2 using SOMAmer-based capture array.
(ii) Prevalent and incident CAD

Analytic Plan:
Aim & analysis 1: Examine the associations between CHIP and serum proteomics levels. CHIP has been measured and grouped through whole exome sequencing or whole genome sequencing (WES or WGS) and levels of ~1,300 (~5,000 in ARIC) serum proteins have been measured using the SOMAscan platform[9] in JHS, MESA, CHS, FHS, and ARIC (N=1821, 947, 936, 1723, and ~8200 respectively). We plan to use mixed models to evaluate the association between CHIP

and proteomics using random intercepts and random slopes to account for the variations in CHIP and proteomics associations across cohorts. CHIP will be modeled in multiple forms: presence of CHIP, presence of CHIP with variant allele fraction (VAF) > 10%, in *DMNT3A* somatic mutation carriers with CHIP, and in *TET2* carriers with CHIP. We will stratify by race and adjust for age, sex, batch, first 10 principal components (PCs) of ancestry, as well as probabilistic estimation of expression residuals (PEER) to account for potential latent factors. Association testing will be performed in each cohort separately and then meta-analyzed. We tentatively plan to conduct independent replication in WHI (N=1,200), where proteomics was measured using the Olink platform[10]. Given the differences between the Somalogic and Olink platforms, this replication may be subject to changes.

Aim & analysis 2: Assess for evidence of causal relationships between CHIP and proteomic markers of CHIP identified in aim 1 and evaluate the mediation effects of proteomic markers for the CHIP-associated CAD risk. We will include both prevalent and incident CAD cases for primary analysis in consideration of power and focus on incident cases in secondary analysis given its high clinical importance. We plan to use bi-directional Mendelian randomization to dig deeper into the causality and use mediation analysis to quantify how much of the CHIP-CAD risk is mediated through proteomics markers of CHIP. For proteomic markers of CHIP that demonstrated causal relations and/or mediation effects, we will follow up with pathway analysis as well as functional validation in aim 3.

Future work: Functionally validate the findings in CHIP-induced mice models through engrafting *TET2*, *DNMT3A*, or other *CHIP* mutations into bone marrow[4]. Those studies are separate from TOPMed data and analyses.

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? __X__ Yes ____ 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"? __X__ 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)?

None

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* 2017.27_)
"Proteomic longitudinal ARIC study: SOMAscan of multiple visits"

___ 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 https://www2.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.

We will make sure to complete in one to three years.

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. Xie, M., *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med* **20**, 1472-1478 (2014).
2. Genovese, G., *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* **371**, 2477-2487 (2014).
3. Jaiswal, S., *et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine* **371**, 2488-2498 (2014).
4. Jaiswal, S., *et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *New England Journal of Medicine* **377**, 111-121 (2017).

5. Fuster, J.J., *et al.* Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science* **355**, 842-847 (2017).
6. Bick, A.G., *et al.* Genetic Interleukin 6 Signaling Deficiency Attenuates Cardiovascular Risk in Clonal Hematopoiesis. *Circulation* **141**, 124-131 (2020).
7. Stastna, M. & Van Eyk, J.E. Secreted proteins as a fundamental source for biomarker discovery. *Proteomics* **12**, 722-735 (2012).
8. Uhlén, M., *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
9. Lollo, B., Steele, F. & Gold, L. Beyond antibodies: new affinity reagents to unlock the proteome. *Proteomics* **14**, 638-644 (2014).
10. Assarsson, E., *et al.* Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One* **9**, e95192 (2014).