1.a. Full Title: Studying the Clinical Conditions, Lifestyle Factors, and Laboratory Markers Associated with Incidence and Expansion of Clonal Hematopoiesis of Indeterminate Potential (CHIP): The Atherosclerosis Risk in Communities (ARIC) Study

b. Abbreviated Title (Length 26 characters): The Clinical, Lifestyle, and Laboratory Risk Factors of CHIP

2. Writing Group:
   Writing group members:

   Seyedmohammad Saadatagah*
   Mesbah Uddin*
   Elizabeth Selvin
   Ron Hoogeveen
   Kunihiro Matsushita
   Roberta Florido
   Vijay Nambi
   Bing Yu
   Wenshung Sun
   Zhi Yu
   Pradeep Natarajan**
   Christie Ballantyne**

   Others welcome

   * Co-first author, **co-last authors

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

First authors: Seyedmohammad Saadatagah

Phone: 5072540219
E-mail: Seyedmohammad.saadatagah@bcm.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: Christie Ballantyne
3. **Timeline:**

The data needed for this analysis will be available in 6 months; we plan to submit for publication within 1.5 years.

4. **Rationale:**

As we age, genetic mutations happen in our cells, causing changes that were not present in our germline DNA, a phenomenon named somatic mosaicism. Like other stem cells, hematopoietic stem cells (HSC) have an increased chance of developing mosaicism. (1) A clone with survival benefits can expand and populate the peripheral blood (i.e., clonal hematopoiesis, CH). (2) The subset of CH with a driver mutation in one of the genes implicated in hematologic malignancies with variant allele frequency (VAF) of at least 2% in the absence of known hematologic malignancy or other clonal disorders is called clonal hematopoiesis of indeterminate potential (CHIP). (3)

Multiple studies confirmed CHIP as an age-related phenomenon that is rare before age 40 and increases to 5%-10% at age 70. (4-6) CHIP is found to be associated with a 1.4-fold increase in all-cause mortality, which is not due to hematologic malignancies. (5, 6). Multiple studies confirmed an increased odds of coronary artery disease, (6-10) stroke, (6) and heart failure (11, 12) in patients with CHIP. To confirm the causality and elaborate on the pathophysiology of the observed association between CHIP and cardiovascular diseases, investigators used mice models and reported accelerated atherosclerosis (7, 13) and worsening heart function (14-17) in mice engrafted with bone marrow carrying the CHIP-associated mutations.

Although CHIP has been accepted as a risk factor for cardiovascular diseases, little is known about factors associated with CHIP incidence and expansion. Age is the most crucial risk factor for CHIP, (4-6) and studies have suggested the increased odds of CHIP due to smoking, (18-20) previous chemo-radiotherapy, (19, 21) chronic infection, (22, 23) and HIV. (24, 25) In addition, recent studies also have suggested that the inflammatory state due to atherosclerosis increases the risk of CHIP (16, 26) suggesting a bidirectional association of CHIP and cardiovascular diseases. However, these all came from cross-sectional studies, and there is a lack of longitudinal evidence on the risk factors of CHIP incidence and its natural history, such as factors contributing to CHIP expansion.

Using the ARIC datasets, we aim to study the clinical conditions (e.g., myocardial infarction, heart failure, obesity, diabetes, and hypertension), lifestyle factors (e.g., smoking, diet, alcohol, BMI, and physical activity), and laboratory markers (e.g., lipid profile including LDL-C, TG, small dense LDL-C, Lp(a), and hsCRP), associated with
the incidence and expansion of CHIP. The ARIC study is ideal for this analysis given the comprehensive assessment of cardiovascular risk factors and having exome sequencing samples on different visits to ascertain CHIP.

5. **Main Hypothesis/Study Questions:**

**Aim 1:** Study the association of clinical conditions, lifestyle factors, and laboratory markers with incidence and expansion of CHIP (the selection of specific visits below are due to data availability of CHIP)

- **Hypothesis 1:** Cardiometabolic risk factors (diabetes, and hypertension), the presence of ASCVD (stroke, MI, CAD, PAD) and heart failure at V2 are associated with increased risk of incident CHIP at V5

- **Hypothesis 2:** Cardiometabolic risk factors (diabetes, and hypertension), the presence of ASCVD (stroke, MI, CAD, PAD) and heart failure at V2 are associated with an increased rate of CHIP expansion at V5

- **Hypothesis 3:** Unfavorable lifestyle factors at V2 are associated with an increased risk of incident CHIP at V5

- **Hypothesis 4:** Unfavorable lifestyle factors at V2 are associated with an increased rate of CHIP expansion at V5

- **Hypothesis 5:** Laboratory markers related to cardiometabolic risk, inflammation and subclinical cardiac injury at V2 are associated with an increased risk of incident CHIP at V5

- **Hypothesis 6:** Laboratory markers related to cardiometabolic risk, inflammation and subclinical cardiac injury at V2 are associated with an increased rate of CHIP expansion at V5

**Aim 1:** Study the cross-sectional association of CHIP at V5 with clinical conditions, lifestyle factors, and laboratory markers

- **Hypothesis 1:** Presence of CHIP is associated with increased odds of ASCVD (stroke, MI, CAD, PAD) and heart failure

- **Hypothesis 2:** Presence of CHIP is associated with increased levels of inflammatory markers including IL6 and IL18

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:** Individuals with exome sequencing at ARIC both visit 2 and visit 5 (n = 4214)

**Exclusions:**
- No exome sequencing available for CHIP calling
- Sex-mismatched sequencing samples
- Those with CHIP at V2 will be excluded from incidence analysis

**Exposure:**
- Prevalent clinical conditions at visit 2
- Lifestyle factors at visit 2
- Laboratory markers levels at visit 2

**Outcomes:**
1. Primary outcome: Incident CHIP, defined as the presence of CHIP in visit 5 while it was absent in visit 2
2. Secondary outcome: CHIP expansion, defined as an increased variant allele frequency in visit 5 or a new mutation found in visit 5 in those who had CHIP at visit 2
4. Exploratory analysis: Cross-sectional study of CHIP at V5 with inflammatory markers and ASCVD

**Covariates:**

Demographic including age, sex, race/ethnicity; clinical conditions including myocardial infarction, stroke, coronary revascularization; history of other atherosclerotic cardiovascular diseases including peripheral vascular disease (PAD), malignancy, hypertension, diabetes mellitus, dyslipidemia; lifestyle factors including obesity, cigarette smoking, alcohol use, diet, physical activity, body mass index (BMI); medication history including chemo or radiation therapy lipid-lowering medication use, antihypertensive medication use, glucose-lowering medication use; and laboratory markers including complete blood count (CBC, measured at V1), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglyceride, small dense LDL-C (measured at V4), LDL-triglyceride (measured at V4), Lp(a) (measured at V4), hemoglobin A1c (HbA1c), fructosamine, glycated albumin, 1,5-anhydroglucitol, cystatin C, estimated glomerular filtration rate (eGFR), high-sensitivity C-reactive protein (hsCRP), high-sensitivity cardiac troponin T (hs-cTnT), and N-terminal pro-brain natriuretic peptide (NT-proBNP), β-2 microglobulin, alanine aminotransferase, aspartate aminotransferase, gamma-glutamyltransferase, fibroblast growth factor 23, galectin 3, and Vitamin D [inclusive of: 25 hydroxyvitamin D3, 25 hydroxyvitamin D3, vitamin D3 epimer (3-epi-25(OH)D3).

CHIP will be determined using exome sequencing (ES) of blood DNA using the GATK MuTect2 (27) somatic variant caller based on the 74 prespecified driver sequence.
variations in genes known to promote clonal expansion of hematopoietic stem cells. (6, 7, 28) A conventional variant allele frequency (VAF) of >2% will be used to identify CHIP, and those with VAF >10% will be considered large clones. CHIP calling will be conducted at BROAD institute.

The covariates that are not measured at the baseline, including those measure at V1 or V4 are additional exploratory analyses.

**Statistical Analyses:**

Continuous variables will be reported using mean (SD) or median (IQR) depending on normality of the data, while categorical variables will be expressed as count (percentage).

Those who were found to have CHIP at visit 2 will be excluded from the analysis of the incident CHIP and will be included in the CHIP expansion analysis. The incrementally adjusted multivariable logistic regression model will be used to calculate the odds ratio for the association between clinical conditions, lifestyle factors, and laboratory markers with outcomes. Covariates including clinical parameters will be measured at the baseline and change in covariate status over time will not be incorporated. All the analyses will be adjusted for age, sex, and ancestry.

Logistic regression models will be constructed to assess associations between covariates at visit 2 and incidence or expansion of CHIP at visit 5. In the discovery process, we will build univariate models and a full model with all the covariates reported to be associated with CHIP in the literature or found to have a p-value <0.1 in our univariate analysis. Adjusted P value will be used for multiple comparisons.

We will also explore the association between CHIP and inflammatory markers, including IL6 and IL18 measured at V5, to study how they interact to increase ASCVD risk.

**Sensitivity Analyses:**

- The concordance of the CHIP ascertainment method used in visits 2 and 5 will be confirmed using cross-referencing some samples.

- We are resequencing on the newer Novasequ platform with over 400 samples from V2 that have incident CHIP at V5 to make sure that technical improvements were not the major reason for detecting CHIP.

**Limitations:**

- Small sample size: analyses of events may be underpowered.
- There is the potential for residual confounding.

7.a. Will the data be used for non-CVD analysis in this manuscript? 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? NA
(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

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”? Confirmed

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.cscn.unc.edu/ARIC/search.php

Yes

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 relevant proposals have been identified.

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? Yes

11.b. If yes, is the proposal

   A. primarily the result of an ancillary study (list number:)

   B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* (#2009.24, #2009.16, #2009.17, #2017.20)

*ancillary studies are listed by number at http://www.cscn.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
References