Manuscript Proposal #3739

PC Reviewed: 3/14/23	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Proteomic age acceleration and cancer incidence: The Atherosclerosis Risk in Communities Study

b. Abbreviated Title (Length 26 characters): Proteomic age acceleration and cancer risk

2. Writing Group:

Writing group members: Shuo Wang, Anna Prizment, Elizabeth Platz, Sithara Vivek, Bharat Thyagarajan, Anne Blaes, Jim Pankow, Weihua Guan (authorship order TBD and other ARIC researchers are welcome to participate).

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

First author: **Shuo Wang**
Address: 420 Delaware Street SE, MMC 480
Minneapolis, MN 55455

Phone: 404-660-4379 Fax:
E-mail: wang8310@umn.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: **Anna Prizment**
Address: 420 Delaware Street SE, MMC 480
Minneapolis, MN 55455

Phone: 612-301-1860 Fax:
E-mail: prizm001@umn.edu

3. Timeline: Analyses will begin within one year after approval.

4. Rationale:

In this study, we propose to examine the association between proteomic age acceleration and risk of cancer overall and by subtypes. We have an ancillary proposal (AS2021.06) entitled "Proteomic aging clock and cancer: Atherosclerosis Risk in Community study" approved in ARIC.

Aging plays a critical role in cancer, as indicated by an increasing incidence of most cancer types with age.[1,2] This could be explained by the accumulation of random genetic mutations and epigenetic alterations, weakened immune system and malfunction of repair processes in aging individuals.[3-6] Also, growing evidence shows that individuals with cancer age faster than those without cancer (so called accelerated aging).[7] However, it is not clear how much rate of aging differs between those with and without cancer, and whether this difference starts even before the cancer is diagnosed. These issues may be examined using aging clocks, which are a set of molecules capable of predicting individual's age. Aging clocks can provide specific information about how old an individual is biologically, independent of the chronological age.[8] Examining aging clocks in cancer is especially important because most cancer cases develop over a long period of time and have a long subclinical period, and 30%-50% of all cancer cases are preventable;[9] thus, aging clocks could provide a new method for identifying people at high risk of cancer that need more frequent screening or treatment with anti-aging agents, such as metformin and senolytics.

Several aging clocks have been created, and the most acknowledged among them is epigenetic clock,[10,11] a set of DNA methylation-based biomarkers in blood or tissue. Epigenetic clocks have been shown to be correlated with chronological age in humans and may predict several health outcomes, including cardiovascular disease, cancer, and mortality,[12-14] although the underlying mechanism of the age-associated DNA methylation remains unclear.[15]

Several studies examined the epigenetic clock in relation to the incidence of cancer, including overall cancer and cancers of lung, breast, colorectum and pancreas.[14,16-20] The Normative Aging Study (N = 442) found that an incorporated Hannum and Horvath age acceleration was associated with overall cancer (38 prostate, 50 skin, and 44 other) [HR (95% CI)= 1.06 (1.02,1.10)].[20] The Women's Health Initiative study (N = 2029) found a one-year increase in Horvarth[11] and PhenoAge[14] clocks were statistically significantly associated with a 50% ($p = 3.4\times10^{-3}$) and 5% ($p = 0.031$) increase in risk of lung cancer, respectively,[14,16] while in the Sister Study (N = 2764), Hannum,[10] Horvath, and PhenoAge age accelerations were statistically significantly associated with increased breast cancer risk with the strongest association observed for PhenoAge: HR (95% CI) = 1.15 (1.07, 1.23).[18] However, previous findings were inconsistent regarding colorectal cancer (CRC) and pancreatic cancer risk when using different epigenetic clocks.[17,19,21] For instance, EPIC-Italy study (N = 845) found an increased CRC risk in male associated with Horvath (P = 0.042) and FHL2 (P = 0.036) clocks, but not Hannum, Weidner, or ELOV2.[19] A pooled analysis of Nurses' Health study, Physician's health study, and the Health Professionals Follow-up Study (N=824) found positive dose-response trends of Hannum and PhenoAge age accelerations with pancreatic cancer risk with a stronger association observed for Hannum: Q4 vs. Q1: OR (95% CI) = 1.73 (1.11, 2.71), however, for Horvath age acceleration, the highest OR was found in the third quartile compared to the lowest quartile.[17] In summary, epigenetic clock showed associations with different types of cancer but the magnitude of associations depend on the cancer type and the clock used in the study. It is possible that

inconsistencies may be partially explained by the use of different type of clock and the small samples size of the studies.

Recently, "proteomic aging clock", which combines a set of proteomic-based aging-related biomarkers, has been proposed to be a potential biological age estimator. The proteomic-based biomarker may be a promising alternative biomarker because it, as intermediate phenotype, can reveal direct information on biological pathways that are involved in many of the physiological and pathological manifestations of aging.[22,23] Using plasma to build a clock is advantageous because plasma proteins are easily measurable in a non-invasive way. Previous laboratory studies found that exposing young mice to plasma from old mice was sufficient to accelerate brain aging in young mice,[24] and exposing aged mice to plasma from young mice could improve their cognitive function.[25] The findings from those studies support the notion that "plasma proteome harbors key regulators of aging."[26] To our knowledge, there have been no studies that examined proteomic aging clock in relation to cancer.

In this proposed study, we will construct a new proteomic aging clock using ~5000 blood proteins that measured by SomaScan. These proteins have been measured at 1990-1992 (Visit 2, N=12,589), 1993-1995 (Visit 3, N=11,340) and 2011-2013 (Visit 5, N=6538). A previous ARIC study has shown an excellent precision in the SomaScan assay.[27] The ARIC study provides a unique opportunity to examine associations between proteomic-based aging biomarkers and cancer risk because it has already measured SomaScan biomarkers at several time points. This will allow us to use different approaches to build proteomic aging clocks based on the cross-sectional association with age (a standard approach) and based on the change in aging biomarkers levels within the individuals (more novel approach). The availability of well-characterized repeated data on main risk factors for cancer such as body mass index (BMI) and smoking will allow us to examine the impact of aging clock beyond the existing risk factors.

This proposal is the first proposal in the series of proposals about proteomic age acceleration and cancer. Our next proposal will examine the role of age acceleration after cancer diagnosis and compare the association between proteomic age acceleration and mortality among individuals with and without cancer.

5. Main Hypothesis/Study Questions:

The goal of this study is to determine the contribution of proteomic age acceleration to the risk of cancer in the ARIC study.

Hypothesis: A higher proteomic age acceleration is associated with an increased risk of overall cancer and specific cancer types.

Specific Aim 1. Create and validate proteomic aging clock in one-half of randomly selected cancer-free participants in ARIC.

SA1a. Create proteomic aging clocks.

We will create four different clocks based on four panels of proteins, respectively. The proteins in each panel will be chosen in the following way: proteins with largest intra-individual changes in blood levels between Visits 2 and 5 (Panel 1), proteins associated with chronological age in a cross-sectional analysis (Panel 2), or the two panels of age-related proteins proposed in the latest systematic review (Panels 3 and 4). Proteomic aging clock will be created as one value for each individual.

SA1b. Validate each of the proteomic aging clocks created in SA1a.

Specific Aim 2. Determine the association between the pre-diagnostic proteomic age acceleration and the risk of overall cancer and specific cancer types

SA2a. Determine the association of the pre-diagnostic proteomic age acceleration with the risk of overall cancer and most common cancers (lung, breast, colorectal, and prostate cancers)

SA2b. Determine the association between the pre-diagnostic proteomic age acceleration and the risk of obesity-related cancer (oropharynx cancer, esophagus cancer, lung and bronchus cancer, post-menopausal breast cancer, liver cancer, gallbladder cancer, pancreas cancer, kidney cancer, stomach cancer, colorectal cancer, endometrial cancer, and ovarian cancer) and smoking-related cancer (lung/bronchus, bladder, kidney, head and neck, stomach, pancreatic, liver and other urinary cancers).[28]

<table>
<tr><th colspan="2">Specific Aim 1</th><th>Specific Aim 2</th></tr>
<tr><td colspan="2">Study population</td><td>Study population</td></tr>
<tr><td>Stage A: Selection of training set</td><td>We will randomly select half of cancer-free participants at visit 5 and the same group of participants at Visit 2. The group of cancer-free participants from visit 2 will be used as a training set and the same group of cancer-free participants from visit 5 will be used as validation set.*</td><td rowspan="3">The remaining cohort (after taking out half of cancer-free participants (the training set)) at Visit 2, and who have the measurement of proteins at Visit 2 using the SomaScan assay.</td></tr>
<tr><td>Stage B: Creation of proteomic aging clock creation</td><td>Same group of people as in the training set.</td></tr>
<tr><td>Stage C: Validation of proteomic aging clock validation</td><td>The randomly selected half of cancer-free participants from visit 5 at stage A will be used as validation set.</td></tr>
</table>

*Of note, we assume proteomic aging clock in cancer-free participants will be the same as their chronological age. We randomly choose the cancer-free participants at Visit 5 first and then go back to choose the same group of participants visit 2 because this can make sure participants in both training set and validation set are cancer-free. We can also randomly choose one-half (or two-thirds) of all ARIC cancer-free participants, and used this group as the training set. The rest cancer-free participants will be used as the validation set.

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: The analytic sample will include participants who were free of cancer at baseline (Visit 2), who gave consent to participate in non-CVD research, and who have the measurement of proteins at Visit 2 using the SomaScan assay (or who have the measurement of DNA methylation for the analysis of epigenetic clocks).

Exposure: Circulating levels of proteins will be extracted from the SomaScan assay (v.4) (SomaLogic company).[29,30] Using this assay, the ARIC study has recently measured more than 5000 plasma proteins in frozen plasma samples collected at Visit 2 (1990-1992, N=12,589), Visit 3 (1993-1995, N=11,340) and Visit 5 (2011-2013, N=6538) (the Bland-Altman coefficient of variation (CVBA) for split samples were 6% at Visit 2, 12% at Visit 3, and 7% at Visit 5). The proteins constituting the proteomic aging clock will be assessed at baseline (Visit 2) and updated at (Visit 3 and Visit 5, where appropriate). Of note, for those with cancer, only proteins measured before cancer diagnosis will be examined. In all the analysis, we will use the log transformed values of proteins.

ARIC has measured DNA methylation in 2,879 Black participants and 1,100 White participants using biospecimens collected at Visit 2 or Visit 3. Epigenetic clocks, e.g., Hannum clock, Horvath clock, DNAm PhenoAge, and GrimAge, will be calculated at the time of blood collection, i.e., at Visit 2 or Visit 3. ARIC have been working on measuring DNA methylation in the full cohort.

Outcome: Cancer incidence was ascertained through 2015 using state Cancer registries in Minnesota, North Carolina, Maryland, and Mississippi, and supplemented by abstraction of medical records and hospital discharge summaries.[31] We will follow participants from Visit 2. A total of 4,407 incident cancer cases were ascertained over a maximum follow-up of 25.9 years (until 2015).

Other covariates of interest: Demographic and clinical variables of interest such as chronological age, sex, race, education, BMI, collected at baseline, and updated as needed during follow-up will be extracted. Cancer risk factors, such as smoking status, pack-years of smoking, alcohol use/intake, physical activity, and diabetes, will also be extracted at Visit 2 and other Visits. Information about estimated glomerular filtration rate (eGFR), aspirin use, and hormone replacement therapy will also be extracted.

Statistical analysis

Specific Aim 1. Create and validate the proteomic aging clock in cancer-free participants in ARIC.

We propose four panels of proteins constituting the clocks. We will create four different proteomic aging clocks based on four protein panels, respectively.

Creating and validating proteomic clock will consist of three stages:

- **Stage A**. develop the panel (for panels 1 and 2)
- **Stage B**. create the proteomic aging clock (which will have one value for each individual)
- **Stage C**. validate the proteomic aging clock

Panel 1: We will develop this panel by determining proteins with largest intra-individual changes in blood levels over ~20 years of follow-up among ARIC cancer-free participants (between Visits 2 and 5).

To do this, we will randomly choose one-half of all ARIC cancer-free participants at Visit 5 and then choose the same group of participants from Visit 2. The group of cancer-free people at visit 2 will be used as a training set for selecting the proteins and creating the clock, and the same group of people at Visit 5 will be used as validation set.

Of note, we choose cancer-free participants to develop a proteomic clock because our goal is to identify a clock that will tell us specifically about aging as a risk factors of cancer. We may conduct a sensitivity analysis by building a clock in healthy people and compare with the results with the clocks developed by cancer-free participants. People may be considered healthy if they are absent of any chronic disease (with the exception of controlled hypertension) and cognitive or functional impairment.[22] We randomly choose the cancer-free participants at Visit 5 first and then go back to choose the same group of participants visit 2 because this can make sure participants in both training set and validation set are cancer-free. The remaining cohort will be used in specific aim 2.

Stage A. Using the training set, we will calculate the mean difference in levels of each protein between Visit 2 and Visit 5 (~ 20 years apart) and the standard deviation for each protein level at visit 2. Then, we will create a standardized ratio for each protein value by dividing the mean difference by its standard deviation. Finally, we will rank the proteins based on their standardized values. The top 1250 proteins (around 25% of the ~5000 proteins) with the largest intra-individual changes in blood levels over time will be included in **Panel 1**.

Of note, if we will also be able to look at proteins that have largest inter-individual change over time, we might include the top 1250 proteins with largest inter-individual change to Panel 1 if these proteins are not found to be with largest intra-individual changes.

Stage B. In the training set, proteins from this panel will be used to fit a regression model to predict proteomic aging clock using multiple linear regression (MLR), Klemera and Doubal method (K+D method),[32] or other penalized regressions, e.g., elastic net.[22] Of note, one regression will be used for all proteins. When fitting the regression in cancer-free people, we will assume chronological age is a surrogate marker of the proteomic aging clock. We will use the regression co-efficient for each protein as its weight to build the formula to predict proteomic aging clock.

For instance, if we use **MLR** to fit the regression, the formula will be:

$$Chronological\ age = \beta_0 + \sum_{j=1}^{1250} \beta_j Protein_j$$

Here, $protein_j$ is the level of j^{th} protein from the panel. **Panel 1** includes the top 1250 proteins with the largest intra-individual changes in blood levels, therefore, 1250 proteins will be included in this MLR.

If we use **K+D method**, [32]

$$Chronological\ age = \frac{\sum_{j=1}^{m}\left(Protein_j - q_j\right)\frac{k_j}{s_j^2} + \frac{C}{s^2_{Chronological\ age}}}{\sum_{j=1}^{m}(\frac{k_j}{s_j})^2 + \frac{1}{s_B^2}}$$

The parameters in the equation will be calculated following the instructions in Klemera and Doubal 2006.[32]

If we use **elastic net,** the formula will be:

$$Chronological\ age = \beta_0 + \sum_{j=1}^{m}\beta_j Protein_j$$

Compared to the MLR model, the "m" proteins will be selected using the elastic net, which penalizes the coefficients using L_1 and L_2 norms.

Stage C. In the validation set, we will predict proteomic aging clock for the same set of individuals at visit 5, multiplying their protein levels at visit 5 and the coefficients estimated in the training set.

We will estimate the prediction accuracy of each proteomic aging clock using Pearson and Spearman correlations (between proteomic aging clock and chronological age) as well as by calculating the mean or median absolute error, i.e., the absolute difference between the proteomic aging clock and chronological age, in the validation set. A proteomic age estimator will be considered useful if its correlation with chronological age exceeds 0.7 in the validation set.

Panel 2: We will develop this panel by looking at associations between chronological age and protein levels among ARIC cancer-free participants.

Stage A. We will determine proteins associated with age using a cross-sectional analysis in one half of ARIC cancer-free participants as discussed in Tanaka 2018.[22] In the training set, we will apply linear regression for each protein separately to assess the association of chronological age at Visit 2, after adjustment for sex, race-center, and the most important factors in aging, i.e., BMI and smoking status (see equation below). A false discovery rate (FDR) corrected threshold of $P < 0.05$ will be considered significant. The significant proteins based on FDR p-value will be included in this panel.

Equation for the linear regression:

$$Protein_i\ Level = \beta_0 + \beta_1 Chronological\ age$$

Here, $protein_i$ represents the i^{th} protein in the ~5000 proteins.

Stage B. The clock will be created as described in approach 1.
Stage C. The clock will be validated as described in approach 1.

Of note, Proteins in Panels 1 and 2 may overlap a lot (both capturing aspects of chronological age, but Panel 1 should also capture proteins that change in a more extreme way that just age alone).

Panel 3 and Panel 4: we will use two panels that were shown to be associated with aging in previous systematic review and validating in our population.

Stage A. We will use each panel of proteins proposed in the latest systematic review of proteomic-based aging research.[8] In the systematic review, Johnson et al. proposed two versions of proteomic aging clock: the 23-protein panel **(Panel 3)** and the 83-protein panel **(Panel 4)**. The 23-protein panel and 83-protein panel included plasma proteins that were reported to be significantly associated with chronological age in 4+ and 3+ different studies, respectively. They evaluated the prediction accuracy of the two panels using Pearson and Spearman correlations between proteomic aging clock and chronological age, as well as by calculating the mean absolute delta age in the INTERVAL cohort, which comprised of 3,301 healthy individuals aged 18-76 years with a median age of 45 (Q1 = 31, Q3 = 55).[33] They measured the proteins in those two panels using the SomaScan assay. In their validation set (1123 participants from the INTERVAL cohort), the 23-protein and 83-protein panel had a Pearson correlation of 0.66 and 0.87 with chronological age, respectively, Spearman correlation of 0.69 and 0.88, respectively, and mean absolute delta age of 8.17 and 4.88 in year respectively.[8]

Stage B. The clock will be created as described in approach 1.
Stage C. The clock will be validated as described in approach 1.

To examine association with cancer, we may choose the most robust proteomic aging clock based on the correlations and mean or median absolute error or apply two best proteomic aging clocks in the analysis of cancer risk. We will work with a statistician to decide on the best way for creating the clock.

Since this area of research is novel and developing, we may also include senescence biomarkers (if they present in ARIC) that have been established as aging biomarkers in the individual studies of cancer, such as breast cancer,[34] that have been already conducted or will be conducted during working on the proposed analysis.

Of note, a total of 32 research publications were included the latest systematic review. All the 32 studies measured fewer proteins (< 5000) compared to SomaScan in ARIC. Therefore, we expect that Panels 3 and 4 may not provide more robust results than Panels 1 and 2. We may include established senescence biomarkers associated with aging (if they present in ARIC) to Panels 3 and 4 to determine if these existing panels can be improved.

Specific Aim 2. Determine the association between the pre-diagnostic proteomic age acceleration and the risk of overall cancer and specific cancer types.

In this analysis, we will include the remaining participants at Visit 2 that are not selected in the training set.

Proteomic age acceleration will be calculated as the residual after regressing each participant's proteomic aging clock on their chronological age. A positive value of age acceleration indicates that the proteomic age is higher than the chronological age.[35]

Cox proportional hazards regression will be used to calculate hazard ratios (HRs) and 95% confidence intervals (CIs) for cancer risk associated with proteomic age acceleration. The proportional hazards assumption will be tested by including an interaction term between proteomic age acceleration and follow-up time in the Cox model (or using graphical methods). Person-years will be estimated from the start of follow-up at Visit 2 until the date of cancer diagnosis, death, or the end of follow-up, whichever occurred first. The proteomic age acceleration will be examined as a continuous variable, but will be transformed or categorized into quartiles if a nonlinear relationship with cancer risk is observed. To account for the change of proteomic age acceleration during follow-up, we will model proteomic age acceleration as a time-dependent variable, which is defined as a variable whose value for a given subject may change over time. We will adjust for age, sex, race-center, eGFR, and liver function in the analysis. We will also adjust for BMI, smoking status, pack-years of smoking, diabetes, aspirin use, and hormone replacement therapy, to determine whether the accelerated aging clocks are simply capturing cancer risk factors. We may also adjust socioeconomic status (SES), because those of lower SES may have less access to care/less access to medications to treat/cure diseases, so the aging clocks that may end up differing.

We will also compare whether or not the associations between proteomic aging clock and cancer risk are similar to the associations between epigenetic clocks and cancer risk. We will calculate epigenetic clocks, e.g., Hannum clock, Horvath clock, DNAm PhenoAge, and GrimAge, using the available DNA methylation data in ARIC. We will follow the same method described in the analysis of proteomic aging clock to calculate epigenetic age acceleration and examine the association between epigenetic age acceleration and cancer risk. ARIC has been working on measuring methylation data in the full cohort. We will calculate the epigenetic clocks in the full cohort when the data is available.

Power calculations

Specific Aim 1. The sample size that is needed for validating an age estimator depends on the accuracy of the age estimator. Based on Horvath et al.,[36] 62 test samples would need if a true correlation is 0.9. A sample of 205, 404, and 867 would need to test a true correlation of 0.80, 0.70, and 0.50, respectively. We expect to have 2896 cancer-free participants from Visit 2 used as the training set and 2896 cancer-free participants from Visit 5 used as the validation set.

Specific Aim 2. The power calculations with cancer incidence up to 2015 are presented in Tables 1 and 2 for **Specific Aim 2**. Categorizing into groups is a conservative approach to estimate power for a continuous variable. Thus, in *Specific Aims 1 and 2*, we will estimate power for the two situations below: a) dichotomize proteomic age acceleration (or epigenetic age acceleration)

at median (higher and lower groups), and b) categorize proteomic age acceleration (or epigenetic age acceleration) at quartiles (Q1, Q2, Q3, and Q4) based on its distribution at V2.

Table 1. Power calculation in ARIC participants (80% power, alpha = 0.05, two-sided) for the association between pre-diagnosis proteomic age acceleration and overall cancer risk, most common individual cancer risk, obesity-related cancer risk, and smoking-related cancer risk

Cancer type	Number of cancer cases	Number of all participants[a]	Minimal detectable risk for dichotomized exposure	Minimal detectable risk for quartiles
Any cancer	4,407	9,693	1.13	1.18
Breast cancer	621	9,693	1.37	1.57
Ovarian cancer	68	9,693	2.62	3.89
Endometrial cancer	115	9,693	2.09	2.83
Kidney cancer	145	9,693	1.93	2.53
CRC	411	9,693	1.48	1.74
Pancreatic cancer	136	9,693	1.97	2.62
Lung cancer	680	9,693	135	1.54
Obesity-related cancer[b]	1,605	9,693	1.22	1.32
Smoking-related cancer[c]	1,328	9,693	1.24	1.36

[a]The number of all participants at Visit 2 is 9,693 because we use 2896 cancer-free participants to develop the clock.
[b]The number of obesity-related cancer is calculated by combining the number breast (n = 621), colorectal (n = 411), endometrial (n = 115), ovarian (n = 68), kidney (n = 145), pancreatic (n = 136) cancers, liver cancer (n = 36), and lethal prostate cancer (approximately 100 cases).
[c]The number of smoking-related cancer is calculated by combining the number lung (n = 680), pancreatic (n = 136), esophageal (approximately 35 cases), bladder (n = 204), stomach (n = 67), liver (n = 36), cervix (n = 25), and kidney cancers.

Table 2. Power calculation in ARIC participants (80% power, alpha = 0.05, two-sided) for the association between pre-diagnosis epigenetic age acceleration and overall cancer risk[a]

Cancer type	Number of cancer cases	Number of all participants[b]	Minimal detectable risk for dichotomized exposure	Minimal detectable risk for quartiles
Any cancer	939	3,625	1.20	1.30

[a]The power calculation was based on the 3,625 pariticpants without prevalent cancer at the time of blood collection (Visit 2 or Visit 3) and with DNA methylation measured.

7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? ____ Yes _X__ 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.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)?

MP3515 A comparison of the inflammatory proteome in cancer survivors and individuals with no cancer history.
MP3482 Plasma Proteins and All-Cause Mortality in Cancer Survivors in ARIC
(MP3515 is a subset of MP3482 for a doctoral dissertation)

MP3057 Repeatability and Longitudinal Variability of the Plasma Proteome.
MP3617 Association between functional MICA polymorphisms, soluble MICA levels, colorectal cancer incidence and mortality: Results from the Atherosclerosis Risk in Communities study.

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* _
AS1995.04 Cancer Study
AS2011.07 Enhancing ARIC Infrastructure to Yield a New Cancer Epidemiology Cohort
AS2017.27 Proteomic longitudinal ARIC study: SOMAscan of multiple visits
AS2019.15 MHC class I chain-related proteins, functional polymorphism and colorectal cancer
AS2020.32 Proteomic Aging Clock and Colorectal Cancer

_X__ 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. White MC, Holman DM, Boehm JE, Peipins LA, Grossman M, Henley SJ. Age and cancer risk: a potentially modifiable relationship. *Am J Prev Med*. Mar 2014;46(3 Suppl 1):S7-15. doi:10.1016/j.amepre.2013.10.029
2. Institute NC. Age and Cancer Risk. https://www.cancer.gov/about-cancer/causes-prevention/risk/age
3. Herceg Z, Hainaut P. Genetic and epigenetic alterations as biomarkers for cancer detection, diagnosis and prognosis. *Mol Oncol*. Jun 2007;1(1):26-41. doi:10.1016/j.molonc.2007.01.004
4. Takeshima H, Ushijima T. Accumulation of genetic and epigenetic alterations in normal cells and cancer risk. *NPJ Precis Oncol*. 2019;3:7. doi:10.1038/s41698-019-0079-0
5. Baylin SB, Jones PA. Epigenetic Determinants of Cancer. *Cold Spring Harb Perspect Biol*. Sep 2016;8(9)doi:10.1101/cshperspect.a019505
6. Shurin MR. Cancer as an immune-mediated disease. *Immunotargets Ther*. 2012;1:1-6. doi:10.2147/ITT.S29834
7. Cupit-Link MC, Kirkland JL, Ness KK, et al. Biology of premature ageing in survivors of cancer. *ESMO Open*. 2017;2(5):e000250. doi:10.1136/esmoopen-2017-000250
8. Johnson AA, Shokhirev MN, Wyss-Coray T, Lehallier B. Systematic review and analysis of human proteomics aging studies unveils a novel proteomic aging clock and identifies key processes that change with age. *Ageing Res Rev*. 07 2020;60:101070. doi:10.1016/j.arr.2020.101070
9. Organization WH. Cancer prevention. https://www.who.int/cancer/prevention/en/#:~:text=Between%2030%2D50%25%20of%20all,for%20the%20control%20of%20cancer.
10. Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. Jan 2013;49(2):359-367. doi:10.1016/j.molcel.2012.10.016
11. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):R115. doi:10.1186/gb-2013-14-10-r115
12. Chen BH, Marioni RE, Colicino E, et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)*. 09 2016;8(9):1844-1865. doi:10.18632/aging.101020
13. Perna L, Zhang Y, Mons U, Holleczek B, Saum KU, Brenner H. Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clin Epigenetics*. 2016;8:64. doi:10.1186/s13148-016-0228-z
14. Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*. 04 2018;10(4):573-591. doi:10.18632/aging.101414
15. Declerck K, Vanden Berghe W. Back to the future: Epigenetic clock plasticity towards healthy aging. *Mech Ageing Dev*. 09 2018;174:18-29. doi:10.1016/j.mad.2018.01.002
16. Levine ME, Hosgood HD, Chen B, Absher D, Assimes T, Horvath S. DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging (Albany NY)*. Sep 2015;7(9):690-700. doi:10.18632/aging.100809
17. Chung M, Ruan M, Zhao N, et al. DNA methylation aging clocks and pancreatic cancer risk: Pooled analysis of three prospective nested case-control studies. *medRxiv*. 2020;
18. Kresovich JK, Xu Z, O'Brien KM, Weinberg CR, Sandler DP, Taylor JA. Methylation-Based Biological Age and Breast Cancer Risk. *J Natl Cancer Inst*. 10 2019;111(10):1051-1058. doi:10.1093/jnci/djz020
19. Durso DF, Bacalini MG, Sala C, et al. Acceleration of leukocytes' epigenetic age as an early tumor and sex-specific marker of breast and colorectal cancer. *Oncotarget*. Apr 2017;8(14):23237-23245. doi:10.18632/oncotarget.15573
20. Zheng Y, Joyce BT, Colicino E, et al. Blood Epigenetic Age may Predict Cancer Incidence and Mortality. *EBioMedicine*. Mar 2016;5:68-73. doi:10.1016/j.ebiom.2016.02.008

21. Wang T, Maden SK, Luebeck GE, et al. Dysfunctional epigenetic aging of the normal colon and colorectal cancer risk. *Clin Epigenetics*. 01 2020;12(1):5. doi:10.1186/s13148-019-0801-3
22. Tanaka T, Biancotto A, Moaddel R, et al. Plasma proteomic signature of age in healthy humans. *Aging Cell*. Oct 2018;17(5):e12799. doi:10.1111/acel.12799
23. Zaghlool SB, Kühnel B, Elhadad MA, et al. Epigenetics meets proteomics in an epigenome-wide association study with circulating blood plasma protein traits. *Nat Commun*. 01 2020;11(1):15. doi:10.1038/s41467-019-13831-w
24. Villeda SA, Luo J, Mosher KI, et al. The ageing systemic milieu negatively regulates neurogenesis and cognitive function. *Nature*. Aug 2011;477(7362):90-4. doi:10.1038/nature10357
25. Villeda SA, Plambeck KE, Middeldorp J, et al. Young blood reverses age-related impairments in cognitive function and synaptic plasticity in mice. *Nat Med*. Jun 2014;20(6):659-63. doi:10.1038/nm.3569
26. Lehallier B, Gate D, Schaum N, et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat Med*. 12 2019;25(12):1843-1850. doi:10.1038/s41591-019-0673-2
27. Tin A, Yu B, Ma J, et al. Reproducibility and Variability of Protein Analytes Measured Using a Multiplexed Modified Aptamer Assay. *J Appl Lab Med*. 07 2019;4(1):30-39. doi:10.1373/jalm.2018.027086
28. Marrone MT, Mondul AM, Prizment AE, et al. Lipid-Lowering Drug Use and Cancer Incidence and Mortality in the ARIC Study. *JNCI Cancer Spectr*. Oct 2021;5(5)doi:10.1093/jncics/pkab080
29. Gold L, Ayers D, Bertino J, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One*. Dec 2010;5(12):e15004. doi:10.1371/journal.pone.0015004
30. Candia J, Cheung F, Kotliarov Y, et al. Assessment of Variability in the SOMAscan Assay. *Sci Rep*. 10 2017;7(1):14248. doi:10.1038/s41598-017-14755-5
31. 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*. 03 2018;27(3):295-305. doi:10.1158/1055-9965.EPI-17-0696
32. Klemera P, Doubal S. A new approach to the concept and computation of biological age. *Mech Ageing Dev*. Mar 2006;127(3):240-8. doi:10.1016/j.mad.2005.10.004
33. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. *Nature*. 06 2018;558(7708):73-79. doi:10.1038/s41586-018-0175-2
34. Sanoff HK, Deal AM, Krishnamurthy J, et al. Effect of cytotoxic chemotherapy on markers of molecular age in patients with breast cancer. *J Natl Cancer Inst*. Apr 2014;106(4):dju057. doi:10.1093/jnci/dju057
35. Bressler J, Marioni RE, Walker RM, et al. Epigenetic Age Acceleration and Cognitive Function in African American Adults in Midlife: The Atherosclerosis Risk in Communities Study. *J Gerontol A Biol Sci Med Sci*. Feb 2020;75(3):473-480. doi:10.1093/gerona/glz245
36. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet*. 06 2018;19(6):371-384. doi:10.1038/s41576-018-0004-3