ARIC Manuscript Proposal #

PC Reviewed: ___/___/18  Status: _____  Priority: ____
SC Reviewed: _________  Status: _____  Priority: ____

1.a. Full Title: Prospective study of circulating galectin-3 and total and site-specific cancer incidence and mortality: the Atherosclerosis Risk in Communities (ARIC) Study

b. Abbreviated Title (Length 26 characters): galectin-3 and cancer incidence and mortality

2. Writing Group:
Writing group members: Michael Marrone, Anna Prizment, Aaron Folsom, Christie Ballantyne, Elizabeth Platz, Corinne Joshu, Other interested ARIC Cancer Working Group members

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

First author: Michael Marrone
Address: 615 N. Wolfe St. Rm E6133

E-mail: mmarron6@jhu.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 Platz
Address: 615 N. Wolfe St. Rm E6132

Phone: 410 614 9674  Fax: 410 614 2632
E-mail: eplatz1@jhu.edu

3. Timeline: The proposed project is an analysis of existing data. We anticipate the analysis of the existing data will take 12 months from the time of manuscript approval.

4. Rationale:
Galectin-3 belongs to a family of soluble β-galactoside-binding lectins, distinguished by a carbohydrate recognition domain (CRD) allowing galectin-3 to bind with glycoproteins and glycolipids. Galectin-3 is secreted by multiple cell types including adipocytes and immune cells, and is involved in multiple physiologic processes including fibrosis, cell adhesion and proliferation, inflammation, and the immune response. It appears galectin-3 can have opposing effects and has been observed to have both pro- and anti-inflammatory effects, and both pro-apoptotic and anti-apoptotic effects. The contrasting effects of galectin-3 in essential physiologic processes associated with common chronic diseases is believed to be dependent on
both the particular glycoprotein or glycolipid receptors engaged and the particular cell types.\textsuperscript{1-3} Once secreted, galectin-3 is believed to typically remain in the vicinity of the tissue by binding to glycoproteins and glycolipids on the cell surface, or to glycan ligands in the extracellular matrix surrounding the cell.\textsuperscript{1} This observation suggests that plasma galectin-3 originates from peripheral blood leukocytes rather than from local tissue production, and could thereby be considered a biomarker of inflammation and immune function in persons not know to have cancer. Other sources of circulating galectin-3 include cancer cells, which are believed to secrete galectin-3 to facilitate adhesion of circulating tumors cells to distant endothelial cells and contributes to an immune-escape mechanism through deactivation of T cell infiltrates and up regulation of regulatory T cells.\textsuperscript{3} Concentration of extracellular gal-3 was increased in participants with breast, colorectal, lung, or ovarian cancer, melanoma, and non-Hodgkin’s lymphoma compared to healthy participants.\textsuperscript{4} Furthermore, gal-3 concentrations were higher in participants with metastatic cancer compare to individuals with localized disease.\textsuperscript{4} However, the exact mechanism through which galectin-3 enters circulation is not well understood.\textsuperscript{1}

Several studies have reported evidence suggesting galectin-3 plays a key role in regulating immune cell function by facilitating cross-linking with glycan ligands and glycoproteins on the surface of immune cells.\textsuperscript{1,5} The resulting scaffolding of molecules on the immune cell surface, referred to as the glycan lattice, regulates the mobility and retention of surface proteins and their interaction with other surface proteins and intracellular signaling pathways.\textsuperscript{1,5} Studies have shown galectin-3 mediates immune cell activation and function including the production of pro-inflammatory cytokines.\textsuperscript{6} Thus, galectin-3 may play a role in tumorigenesis by facilitating a pro-inflammatory microenvironment.

Similarly, galectin-3 may modulate the retention of cell surface transporter proteins and receptors to modulate metabolic function relevant to cancer development.\textsuperscript{7} Recent investigations showed galectin-3 promotes insulin resistance and glucose intolerance, and inhibition of galectin-3 has also been shown to improve insulin sensitivity.\textsuperscript{2,8} Diabetes and related metabolic perturbations (e.g., hyperglycemia) have been shown to be associated with an increased risk of total cancer incidence and total and site-specific cancer mortality in ARIC.\textsuperscript{9} Insulin and insulin-like growth factors are also thought to contribute to cancer development.\textsuperscript{10} Together, the converging evidence suggests galectin-3 regulates multiple physiologic processes that may contribute to cancer development through the interaction between the galectin-3, the glycan lattice and cell-surface proteins and receptors.

To our knowledge, no prospective epidemiologic investigation has reported the association between galectin-3 and cancer incidence and mortality in people free of a cancer diagnosis at the time of galectin-3 measurement. In ARIC, circulating galectin-3 was measured at visit 4, along with the relevant anthropometric and metabolic factors, thus making ARIC an ideal cohort to study the association of galectin-3 across multiple common cancers. Given the potential role of galectin-3 as a regulator of multiple pathways contributing to cancer development including inflammation and metabolism, the objective of the current proposal is to evaluate the association between galectin-3 in individuals without a prior cancer diagnosis and total and site-specific cancer incidence and mortality.

5. **Main Hypothesis/Study Questions:**
We hypothesize higher concentrations of galectin-3 will be associated with increased risk of total and site-specific cancer incidence and will have a stronger association with total and site-specific cancer mortality. Because of the different points in the natural history of cancer development (incidence vs. mortality) and the hypothesized mechanism through which galectin-3 promotes cancer development, we do not expect the magnitude of association to be consistent for both cancer incidence and mortality. This difference in magnitude of association between HbA1c and cancer incidence and mortality was observed in prior analyses in ARIC with follow-up through 2006. Further, we hypothesize that a positive association is not explained by solid cancer diagnoses close in time to visit 4 (reverse causation). For hematopoietic cancers, we hypothesize that higher plasma galectin-3 concentration is associated with a higher cancer incidence and mortality, but that this association is explained by hematopoietic cancer diagnoses close in time to visit 4. We also propose to test whether systemic inflammation, diabetes, insulin resistance, or hyperglycemia are mediators in the path between galectin-3 and total and site-specific cancer incidence and mortality.

Study question 1: What is the association between galectin-3 and total and site-specific cancer incidence and mortality in participants free of a cancer diagnosis at visit 4?

Study question 2: Does systemic inflammation (e.g., CRP), diabetes, insulin resistance or hyperglycemia mediate the association between galectin-3 and total and site-specific cancer incidence and mortality?

Study question 3: Is the association between galectin-3 and total and site-specific cancer incidence and mortality modified by systemic inflammation (e.g., CRP), obesity, diabetes, insulin resistance, or hyperglycemia? In the Analysis section, we provide further details of the methods to investigate the potential role of the inflammatory and metabolic factors as effect modifiers and mediators in the complex biological systemic in which galectin-3 is hypothesized to contribute to cancer development.

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: Prospective analysis of galectin-3 in participants without a cancer diagnosis at baseline. Galectin-3 was measured in plasma collected at ARIC study visit 4 (1996-1998) as part of ancillary study #2013.12 (PI: Ballantyne, C). Visit 4 will serve as baseline for this analysis.

Exclusions: Participants with prevalent cancer at baseline (visit 4), without information on galectin-3, and those who did not consent to non-CVD research.

Exposure: Galectin-3 was measured in plasma samples collected at visit 4 using a chemiluminescent microparticle immunoassay on an Architect i 2000sr instrument (Abbott, Abbott Park, IL).
Outcome: We will utilize the 2012 adjudicated cancer case file with incident cancers and cancer deaths for bladder, breast, colorectal, liver, lung, pancreatic, and prostate cancer. Total cancer incidence is defined as the first primary cancer diagnosis in men and women free of any cancer diagnosis at baseline. Total cancer mortality is defined as the death from any cancer listed as the underlying cause of death on death certificates. Site-specific cancer (bladder, breast, colorectal, liver, lung, pancreatic, and prostate cancer) incidence and mortality will be evaluated for individual cancers, diabetes-associated cancers (e.g., liver, pancreas, endometrium, colorectal, breast and bladder)\(^\text{10}\) and obesity-associated cancers combined that have sufficient number of events. Due to the source of circulating galectin-3, circulating leukocytes, we hypothesize that galectin-3 will be associated with a higher risk of cancer incidence and mortality, but that this increase in risk is explained by hematopoietic cancer diagnoses close in time to visit 4 (e.g., reverse causation).

Covariates: For total cancer, we will adjust for covariates common across cancer types including age, sex, race*field center, smoking status, education status, BMI, alcohol consumption, family history of cancer, aspirin and NSAID use, and obesity. We will assess if systemic inflammation (classified with C-reactive protein), diagnosed diabetes, insulin resistance (classified using the homeostatic model assessment [HOMA-IR], or hyperglycemia (glucose and HbA1c) are mediators of the association between galectin-3 and total and site-specific cancer incidence and mortality. A recent GWAS reported significant associations between several SNPs and circulating galectin-3,\(^\text{11}\) suggesting genetic variation may influence the galectin-3 measurement in ARIC participants; thus we will stratify by this SNP. We will optimize adjusted models for site-specific cancers with site-specific covariates accordingly: breast cancer (HRT use, age at menarche, parity, age of first birth, total number of months breast feeding); colorectal cancer (red meat consumption); and liver cancer (cirrhosis which is in the path from galectin-3 to liver cancer, possibly).

Analysis: We will create quantiles of continuously measured galectin-3 based on the distribution in participants without cancer at the end of follow-up (December 2012). We will also classify participants using reference cutpoints (e.g., https://www.mayomedicallaboratories.com/test-catalog/Clinical+and+Interpretive/86202) to define low, intermediate, and high concentrations of galectin-3. We will use Cox proportional hazards regression to estimate cause-specific hazard ratios (HR) and 95% confidence intervals (CI) comparing galectin-3 and total and site-specific cancer incidence and mortality. For cancer incidence, participants will contribute person time from visit 4 until diagnosis of total or site-specific cancer, diagnosis of another cancer (in the site-specific cancer analysis), death from any cause, or December 31, 2012, whichever comes first. For cancer mortality, participants will contribute person time from visit 4 until death from total cancer or site-specific cancer, death from another cause, or December 31, 2012, whichever comes first. We will enter galectin-3 concentration as indicator variables for quantile (or clinically relevant categories) and will test for trend by entering median values of galectin-3 for each quantile as an ordinal variable in the model and examine the Wald p-value. We will analyze the association between continuous galectin-3 per 1-standard deviation change and total and site-specific cancer incidence and mortality. We will use restricted cubic splines to describe the shape of the association between galectin-3 and total and site-specific cancer incidence and mortality. We will use traditional methods to test whether inflammation, diabetes, insulin resistance or hyperglycemia are mediators of the association between galectin-3 and total and site-specific...
cancer incidence and mortality, by comparing the beta-coefficients for galectin-3 from the models with and without the potential mediators. If the hypothesized variables are mediators, the beta-coefficient for galectin-3 would be attenuated in a multivariable model including the mediator(s). Recognizing the complex biological system through which galectin-3 and the metabolic and inflammatory factors may contribute to carcinogenesis, we will explore more advanced methods for mediation to investigate potential exposure-mediator interaction to determine both the direct and indirect effects of galectin-3. However, we do not have sufficient power or repeated measures of galectin-3 to conduct more complex mediation analyses to rule out time-varying confounding.

Stratification variables: We will stratify total cancer and site-specific cancer analyses by gender and race. Stratification for site-specific cancers will be dependent on a sufficient number of events. We will also group diabetes and obesity-related cancers in a separate subgroup analysis. We will use the likelihood ratio test to test for interaction using cross-product interaction terms with continuously measured galectin-3 and CRP (continuous), BMI (continuous), waist circumference (continuous), HOMA-IR (continuous), and hyperglycemia (continuous). We will create joint-categories of CRP, obesity, insulin resistance, and hyperglycemia to classify participants in the normal range on all factors, or elevated on one or more factor and test for interaction between the joint categories and galectin-3.

A recent GWAS reported significant associations between several SNPs and circulating galectin-3, suggesting genetic variation may influence the galectin-3 measurement in ARIC participants. In prior analyses, rs4644 was reported to explain only 0.4% of variation in galectin-3 in ARIC. Therefore, we will stratify by rs4644 in our analyses to determine if genetic variation influencing galectin-3 measurement is differentially associated with total and site-specific cancers.

Galectin-3 is currently being investigated as a marker for cardiovascular disease risk. The association between galectin-3 and CVD could introduce detection bias from increased utilization of health care for CVD indications increasing the likelihood of cancer screening and detection. To reduce potential detection bias, we will conduct sensitivity analyses restricting to participants 1) with health insurance, 2) with private or pre-paid health insurance or Medicare, 3) who have been to the doctor at least once in the last five years.

The table below provides a range of minimum detectable HRs with 80% power across the number of participants at risk at baseline and event probabilities. With over 9,000 ARIC participants with galectin-3 measured at visit 4 and over 3,000 total incident cancer cases (26% Kaplan-Meier risk estimate over 25 years) and more than 1,500 total cancer deaths (13% Kaplan-Meier risk estimate over 25 years) in ARIC ascertained through 2012, we have the ability to detect HRs in magnitude similar to the HRs reported for another etiologic circulating biomarker of immune function (beta-2 microglobulin) for total cancer incidence (HR: 1.25; 95% CI: 1.06-1.47) and total cancer mortality (HR: 1.30; 95% CI: 1.02-1.64) previously reported in ARIC. The minimum detectable HRs in the table are also within the range of HRs comparing beta-2 microglobulin and colorectal cancer incidence (HR: 2.21; 95% CI: 1.32-3.70) lung cancer incidence (HR: 1.64; 95% CI: 1.03-2.61), lung cancer mortality (1.69; 95% CI: 1.10-2.59), and hematological cancers (HR: 2.44; 95% CI: 1.20-4.95). The minimum detectable HRs in the table are also within the range of the association comparing galectin-3 and chronic kidney
With 55% of female participants in ARIC, we expect to have sufficient number of participants to detect HRs total cancer incidence and mortality stratified by gender similar to those reported for beta-2 microglobulin. Given 27% ARIC participants are African American, we expect to have sufficient numbers to detect differences in the association between galectin-3 and total cancer incidence in African American participants.

Table. Minimal detectable hazard ratios with 80% power and a 2-sided test with alpha=0.05 across number of participants at risk and risk of cancer.

<table>
<thead>
<tr>
<th>No. at risk</th>
<th>Risk of cancer 0.20</th>
<th>Risk of cancer 0.10</th>
<th>Risk of cancer 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,000</td>
<td>1.33</td>
<td>1.49</td>
<td>1.75</td>
</tr>
<tr>
<td>4,000</td>
<td>1.22</td>
<td>1.32</td>
<td>1.49</td>
</tr>
<tr>
<td>6,000</td>
<td>1.18</td>
<td>1.26</td>
<td>1.38</td>
</tr>
<tr>
<td>8,000</td>
<td>1.15</td>
<td>1.22</td>
<td>1.32</td>
</tr>
<tr>
<td>10,000</td>
<td>1.13</td>
<td>1.19</td>
<td>1.28</td>
</tr>
</tbody>
</table>

Limitations: We propose using CRP in our models to adjust for systemic inflammation and to test for the interaction between galectin-3 and inflammation. As a general marker of inflammation, CRP may not be a direct measure of the interaction between galectin-3 and circulating immune cells limiting our ability to detect a statistical interaction. As mentioned above, beta-2 microglobulin is an immune-related biomarker associated with total cancer incidence and mortality in ARIC and will be used in addition to CRP to adjust for systemic inflammation. A final limitation is the small number of site-specific cancer deaths. We have pre-specified a subgroup analysis including obesity-related cancers to maximize power if there are insufficient number of events for site-specific analyses.

7.a. Will the data be used for non-CVD analysis in this manuscript? ___X__ Yes  ____ 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? ___X__ 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? ____ 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 file ICTDER03 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.unc.edu/ARIC/search.php

___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)?

   ARIC MS #2835
   ARIC MS #2805
   ARIC MS #2939
   ARIC MS #2882
   ARIC MS #2967

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* _2013.21________) 
   ___     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 http://www.csc.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 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.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