1.a. **Full Title:** Association between functional MICA polymorphisms, soluble MICA levels, colorectal cancer incidence and mortality: Results from the Atherosclerosis Risk in Communities study

b. **Abbreviated Title:** MICA polymorphisms, s-MICA levels, CRC incidence and mortality

* We received ARIC approval for ancillary study **2019.15.** This is the first manuscript proposal from that ancillary study.

2. **Writing Group:**

ARIC co-authors: Guillaume Onyeaghala, Chinenye Ugoji, Anna Prizment, Nathan Pankratz, Heather Nelson, Bharat Thyagarajan, Weihong Tang, Christian R. Gomez, Corinne Joshu, Elizabeth Platz, Joseph Coresh, David Couper. Other ARIC researchers are welcome.

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

**First author:** Guillaume Onyeaghala

Address: 1300 South 2nd St. Suite 300
Minneapolis, MN 55454
Phone: 507-401-0437
E-mail: onyea005@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: 612-625-9604
E-mail: prizm001@umn.edu

3. **Timeline:** About a year from approval date
4. **Rationale:**

**Background**

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer death among men and females in the United States\(^1\), and additional research is needed to decrease the burden of this disease. An important mechanism in cancer development and progression is tumor immunosurveillance, when the host immune system identifies and clears tumor cells\(^2\)–\(^4\). In this manuscript proposal, we will focus on the major histocompatibility complex (MHC) class 1 chain-related protein A (MICA), a transmembrane protein that serves as a major ligand for natural killer (NK) cells.

NK cells can target and eliminate cancer cells by binding to the MICA protein on the cancer cell through the immune cell’s NKG2D (natural-killer group 2, member D) receptors\(^5\)–\(^6\). Tumor cells can avoid the immune response by shedding MICA proteins from their cell surface in soluble form (called s-MICA)\(^7\), and this shedding is mediated by disintegrin and metalloproteinases (ADAM) 10 and 17\(^8\)–\(^9\). S-MICA has been reported to further cause downregulation of NKG2D on NK cells, and in turn lead to severe impairment of the antitumor immune response\(^10\)–\(^11\) (Figure 1).

![Figure 1. Inhibition of cancer cell immunosurveillance by s-MICA secretion](image-url)
Consistent with the mechanism above, elevated blood levels of s-MICA have been detected in sera of patients with leukemia as well as liver, oral, breast, prostate, and pancreatic cancers. In addition, elevated s-MICA levels have been significantly correlated with cancer stage and metastasis for liver and pancreatic cancer. In agreement with other studies, our team showed that pancreatic cancer cases have a higher median level of s-MICA than controls (58.48 vs 43.07 pg/ml, p<0.05).

To date, few studies have examined the role of MICA in CRC development and progression. Three small studies of CRC patients showed that higher MICA expression in tumor tissue is associated with better survival of CRC patients, while higher s-MICA levels in the blood are associated with poorer survival.

Given the important role of MICA in immune activation and surveillance against tumorigenesis, it is necessary to study associations between germline MICA polymorphisms and susceptibility to CRC. The gene encoding MICA, which is located in the major histocompatibility complex (MHC) genomic region on chromosome 6, is highly polymorphic, and to date, over 100 MICA alleles have been identified. Previous studies showed that the prevalence of MICA polymorphisms differ in patients with hepatocellular, cervical, and pancreatic cancers compared to those without cancer. To our knowledge, three case control studies have examined MICA polymorphisms in relation to CRC incidence and outcomes: one study reported a higher prevalence of the Val allele (major allele) for the Val129Met polymorphism among those with CRC compared to those without, and two additional studies found that both the Val allele (major allele) for the Val129Met polymorphism and MICA short tandem repeat (STR) polymorphisms A4 and A9 were associated with greater CRC metastases and worse CRC outcomes.
Polymorphisms in the *MICA* gene may influence (a) MICA’s binding to the NKG2D receptor, or (b) the ability of a cancer cell to shed the MICA protein into blood. In this proposal, we focus on two known functional *MICA* polymorphisms, which represent the above two major types of changes that may weaken the affinity of MICA-NK binding and increase s-MICA levels, thus diminishing immunosurveillance, which facilitates tumor progression. Among all functional genetic changes observed in the *MICA* gene, we are especially interested in the *MICA* A5.1 frameshift mutation, located in the gene region coding for the transmembrane domain of the MICA protein. This domain is encoded by alleles that have a variable number of STRs consisting of four to 10 GCT repeats, designated as A4, A5, A5.1, A6, A7, A8, A9, A10 respectively with A5 being the major allele\(^{36,42}\). The A5.1 allele contains an extra guanine (G) insertion after two GCT triplets, which causes a frameshift mutation resulting in a premature stop codon and a MICA protein that is shorter than its normal counterpart \(^{36,42}\). This short protein is more easily cleaved from the cell surface, leading to elevated levels of s-MICA. Our team previously showed that the presence of the *MICA* STR A5.1 mutation is associated with an increased risk of pancreatic cancer in a case-control study (OR=2.00, 95% CI: 1.06-3.79)\(^{38}\).

The second important functional mutation *MICA*-129, also known as Val129Met (SNP rs1051792) is a valine to methionine missense mutation at codon 129 in exon 3 of the *MICA* gene. The Met allele has a greater affinity of *MICA* for the NKG2D receptor\(^{43}\), while the presence of the Val allele has been associated with increased odds of CRC in case-control studies in both German (European) and Han Chinese populations\(^{39-41,44}\).

We conducted a preliminary analysis of *MICA* SNPs in a subset of 224 CRC cases among 8,663 white ARIC participants (followed up to 2012) and found that several functional *MICA* SNPs, which decrease MICA-NK binding affinity, tend to be associated with increased CRC risk
(Table 1). Of note, the association between *MICA* STR A5.1 and CRC risk was not part of our preliminary analysis.

Thus, the **objective of this manuscript proposal** is to elucidate the association between main functional germline mutations in the *MICA* gene, plasma s-MICA levels and CRC incidence and mortality in the population-based Atherosclerosis Risk in the Community (ARIC) study of 15,792 white and black men and women aged 45-64 years at baseline with follow-up through 2015. This manuscript proposal is the first in a series of proposals from our funded R03 and University of Minnesota Masonic Cancer internal grants. The ARIC ancillary proposal was approved in 2019.

### Table 1. Preliminary analysis of *MICA* SNPs and CRC Risk in ARIC.

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Location on/near MICA gene</th>
<th>Reference Allele</th>
<th>Allele Frequency (%)</th>
<th>Comparison Allele</th>
<th>Allele Frequency (%)</th>
<th>Result of the mutation</th>
<th>Risk (95% Confidence intervals (CI))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Val129Met (rs1051792)</td>
<td>exon 3</td>
<td>A</td>
<td>29.18</td>
<td>G</td>
<td>70.82</td>
<td>↓ MICA's binding affinity</td>
<td>1.19 (0.96, 1.48)</td>
</tr>
<tr>
<td>Lys96Glu (rs1051704)</td>
<td>exon 3</td>
<td>A</td>
<td>29.19</td>
<td>G</td>
<td>70.81</td>
<td>↓ MICA's binding affinity</td>
<td>1.19 (0.96, 1.48)</td>
</tr>
<tr>
<td>rs131896</td>
<td>exon 4</td>
<td>A</td>
<td>23.55</td>
<td>G</td>
<td>76.45</td>
<td>No direct effect on MICA's binding affinity</td>
<td>0.97 (0.78, 1.21)</td>
</tr>
<tr>
<td>rs1051795</td>
<td>exon 4</td>
<td>A</td>
<td>44.68</td>
<td>G</td>
<td>55.32</td>
<td>No direct effect on MICA's binding affinity</td>
<td>1.10 (0.91, 1.33)</td>
</tr>
<tr>
<td>rs2116448</td>
<td>Intron Variant</td>
<td>C</td>
<td>63.59</td>
<td>T</td>
<td>36.41</td>
<td>No direct effect on MICA's binding affinity</td>
<td>1.12 (0.93, 1.35)</td>
</tr>
<tr>
<td>rs763288</td>
<td>Intron Variant</td>
<td>A</td>
<td>5.53</td>
<td>G</td>
<td>94.47</td>
<td>No direct effect on MICA's binding affinity</td>
<td>1.08 (0.74, 1.55)</td>
</tr>
<tr>
<td>rs2395029</td>
<td>HLA complex P5 Gene near the MICA gene cluster</td>
<td>G</td>
<td>2.71</td>
<td>T</td>
<td>97.29</td>
<td>No direct effect on MICA's binding affinity</td>
<td>2.45 (1.17, 5.10)</td>
</tr>
</tbody>
</table>

### 5. Main Hypothesis/Study Questions:

**Aim 1.** Determine whether functional *MICA* polymorphisms (particularly, *MICA* STR A5.1 and Val at Position 129 (rs1051792, Val129Met) and higher baseline s-MICA levels are associated with increased CRC risk in the ARIC cohort.

**Hypothesis:** Having the *MICA* STR A5.1 and Val at Position 129 (rs1051792, Val129Met), and higher plasma s-MICA levels are associated with increased CRC risk and mortality.

**Aim 2.** Determine whether functional *MICA* SNPs (particularly, *MICA* STR A5.1 and Val at position 129 (rs1051792, Val129Met) and higher pre-diagnostic s-MICA levels are associated with increased CRC-specific mortality in the ARIC cohort.
**Hypothesis:** Having the MICA STR A5.1 and Val at Position 129 (rs1051792, Val129Met), and higher plasma s-MICA levels are associated with higher CRC mortality.

In an exploratory analysis we will also examine overall and CRC-specific mortality among those with CRC.

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:**

For the analysis of CRC incidence and mortality: The analytic sample will include participants who were free of cancer at baseline and who gave consent to participate in non-CVD research and genetic studies.

**Exposures:**

A genome-wide association study (GWAS) in the ARIC study was previously performed using the Affymetrix Genome-Wide Human SNP Array 6.0 and imputed to the 1000 Genomes Phase 3 reference panel. Recently, the co-investigator on this project, Dr. Nathan Pankratz, has imputed the genetic data to TOPMed. The data from these two imputation panels will be used to study one of the main functional polymorphisms of interest (rs1051792 for Val129Met), and additional MICA SNPs previously shown to be associated with different cancer types (rs1051794, rs1131896, rs1063635, rs2516448, rs3763288, rs2395029) (Table 1).

Whole genome sequencing in ~12,000 participants in the ARIC studies has been conducted by Trans-Omics for Precision Medicine Program (for ~8,000 of all samples by TOPMed; NHLBI) and by Centers for Common Disease Genomics (for ~4000 samples, NHGRI). These data will be used to study the other main functional polymorphisms of interest.
(MICA A5.1) and other STR(triple) polymorphisms in the MICA gene (MICA STRs A4–A10 alleles), which have also been associated with cancer risk.

Circulating levels of s-MICA will be extracted from the SOMAscan assay (v.4) (SomaLogic company)\(^{45,46}\). Using this assay, the ARIC study has recently measured more than 5000 plasma proteins in frozen plasma samples collected at Visit 3 (1993-1995, N=11,340) and Visit 5 (2011-2013, N=6538) (Median CV range across ~5 proteins measured from ARIC visit 3 samples: 6.9 – 8.3\%). Currently, these proteins are being measured using SomaScan at Visit 2 (1990-1992, N=12,589). For all s-MICA analyses in this manuscript proposal, s-MICA values will be log2 transformed to normalize its distribution, in accordance to ARIC guidelines.

**Outcomes:**

CRC incidence and mortality was ascertained from 1987 through 2015 using state cancer registries in Minnesota, North Carolina, Maryland, and Mississippi, and supplemented by abstraction of medical records and hospital discharge summaries\(^{47}\). A total of 435 incident CRC cases and 152 CRC deaths were ascertained over a maximum follow-up of 29.1 years.

**Statistical Analysis**

**s-MICA levels:**

We will examine the distribution of s-MICA levels at different visits and their change between visits (visit 3 and visit 5, 18 years, N=5530) in those with and without CRC. Of note, we will also examine MICA levels at visit 2 when the data becomes available. The log 2 transformed level of s-MICA will be presented as a continuous variable and will be compared across participants’ characteristics at baseline, including age, sex, race, center, BMI, height,
waist circumference, diabetes status, smoking, HRT use (women), aspirin use, periodontal
disease, CRP, B2M, and eGFR.

**MICA genetic variants and circulating s-MICA levels:**

We will examine the prevalence all *MICA* genetic variants (*MICA* STR A5.1 and
Val129Met) in whites and African Americans. We will test associations between *MICA* variants
and circulating s-MICA levels in whites and African Americans using generalized linear
regression to assess mean log transformed s-MICA levels across *MICA* polymorphisms
(stratified by race) at visit 3 and visit 5. Of note, we will also run this analysis for s-MICA levels
at Visit 2 when data becomes available.

In addition, we will use the same approach to study other *MICA* SNPs and STRs (e.g.
*MICA* STRs A4 and A9).

**MICA SNPs and CRC incidence and mortality:**

To examine the association between *MICA* genetic variants and CRC incidence (Aim 1)
and CRC mortality (Aim 2), we will use Cox proportional hazard models and report hazard ratios
(HRs) and their 95% CIs. SNPs will be presented using genetic additive models. *MICA* STRs
derived from ARIC WGS data will be presented as both genetic additive models and genetic
dominant models. Analyses will be adjusted for potential confounders such as age, sex, and
center. If the associations show similar trends in Whites and African Americans, the populations
will be combined. If the estimates are different, the results will be presented separately, or the
data will be meta-analyzed. All genetic analyses will be also adjusted for principal components
of genetic ancestry in whites and African Americans (2 for individuals of European Ancestry and
4 for African Americans) by entering them as continuous covariates in the model. In an
additional exploratory analysis, we will also examine all-cause and CRC-specific mortality after adjusting for stage at diagnosis. In this analysis, follow-up will be estimated from date of diagnosis to death or end of 12/31/2015

Although we do not anticipate confounding in the studies of germline variations, we will test whether the association changes after adjusting for BMI, diabetes status, alcohol consumption, and smoking for all the models above. In addition, we will evaluate the interaction between the SNPs that decrease MICA binding affinity (e.g. Val129Met allele) and SNPs that increase s-MICA levels to evaluate the joint effects of these mutations on CRC incidence and mortality. We will aim to replicate these associations in the UK Biobank. Dr. Pankratz has already has received access to the UK Biobank and identified CRC cases.

**s-MICA levels and CRC incidence and mortality:**

To examine the association between circulating s-MICA levels and CRC incidence and mortality, we will limit the analyses to Visit 3 (306 CRC cases) and Visit 2 (360 CRC cases) when it becomes available, as we will be underpowered to study an association at Visit 5 (56 CRC cases). s-MICA levels will be presented as a continuous variable (log-transformed) and categorized as quartiles. We will use Cox proportional hazard models reporting hazard ratios (HRs) and their 95% CIs and controlling for potential confounders such as age, sex, race*center (5-level variable), BMI, height, waist circumference, diabetes status, smoking, HRT use (women), aspirin use, periodontal disease, eGFR and blood volume (if available). In addition, we will test whether controlling for other inflammatory markers [CRP, beta-2-microglobulin (B2M)] change the association, i.e., check if the s-MICA – CRC association is independent of CRP and B2M circulating levels. We will also check for interaction of age, sex, race, diabetes status, CRP and B2M levels with circulating s-MICA levels, as previous studies have indicated that these
variables can impact s-MICA levels. Finally, we will examine interaction with circulating proteins affecting the affinity of MICA with NK or MICA cleaving (such as the disintegrin and metalloproteinases (ADAM) 10 and 17\(^8,9\)).

In an exploratory analysis, we will examine all-cause and CRC-specific mortality among those diagnosed with CRC using similar models additionally adjusted for stage at diagnosis. For this analysis, we will include s-MICA levels before the diagnosis but closest to it.

The \( p \) value for statistical significance will be set at 0.05, and all statistical analyses will be conducted using SAS or STATA.

**Mendelian randomization:**

If we find an association between s-MICA and CRC risk, we will investigate a potential causative role of s-MICA in CRC development using the Mendelian randomization principle, to examine whether SNPs that increase s-MICA concentration are associated with incident CRC\(^{48-50}\).

Future studies of our group will use the same approach to study cancers previously shown to be associated with \( MICA \) polymorphisms (e.g., prostate cancer, all hematological cancers combined) and s-MICA levels in relation to \( MICA \) A5.1, \( MICA \) 129 (Val129Met), other \( MICA \) variants, and s-MICA levels.

**Power calculation:**

Assuming 435 CRC cases among 14,688 individuals in the ARIC study, we will have 80% power (\( \alpha=0.05 \), 2-sided) to detect an HR=1.37 or larger for CRC risk associated with minor allele frequencies of 30%. We observed a minor allele frequency of 28.4% for \( MICA \) STR A5.1 in our
previous study\textsuperscript{38} and a minor allele frequency of 29.2\% for Val129Met in our preliminary ARIC analysis.

Assuming 360 cases among 12,589 individuals in the ARIC study at visit 2, we will have 80\% power ($\alpha=0.05$, 2-sided) to detect a 1.56 effect size for extreme tertiles of s-MICA and 1.23 per 1 SD increase effect size in s-MICA presented as a continuous variable, based on estimates our previous study of s-MICA levels in a population based case control study.

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

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES$_{\text{OTH}}$ = “CVD Research” for non-DNA analysis, and for DNA analysis RES$_{\text{DNA}}$ = “CVD Research” would be used? Yes

(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$_{\text{DNA}}$ = “No use/storage DNA”? Yes

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

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)?

1429 Inflammatory and allergy markers as predictors of colorectal cancer risk (CRC): Atherosclerosis Risk in Communities (ARIC) study

1690 Association of polymorphisms in C-reactive protein and fibrinogen genes with cancer risk: Atherosclerosis Risk in Communities (ARIC) study
2187 Circulating beta-2 microglobulin (B2M) and cancer risk and mortality: Atherosclerosis Risk in Communities (ARIC) Study

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

__X__ A. primarily the result of an ancillary study

1995.04 Cancer Study

2011.07 Enhancing ARIC Infrastructure to Yield a New Cancer Epidemiology Cohort

2012.10 Whole-genome sequencing in ARIC

2017.27 Proteomic longitudinal ARIC study: SOMAscan of multiple visits

2019.15 MHC class I chain-related proteins, functional polymorphisms and colorectal cancer

*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.

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


