

**SF 424 R&R and PHS-398**  
**Specific Table of Contents**

SF 424 R&R Cover Page.....	1
Research & Related Other Project Information.....	2
Project Summary/Abstract (Description).....	3
Project Narrative .....	4
Facilities & Other Resources.....	5
Equipment.....	8
Vertebrate Animals Section.....	9
PHS398 Cover Page Supplement.....	10
PHS 398 Research Plan.....	11
Specific Aims .....	12
Research Strategy.....	13
PHS Human Subjects & Clinical Trial Information.....	19
Justification That No Human Subjects are Involved.....	20
Resource Sharing Plan .....	21

PI: <b>Prizment, Anna</b>	Title: Immune-regulating MHC class I-like proteins and colorectal cancer risk	
	FOA: PAR18-021 Clinical Trial:Optional	
	FOA Title: NCI Small Grants Program for Cancer Research (NCI Omnibus R03 Clinical Trial Optional)	
	Organization: UNIVERSITY OF MINNESOTA	
	Department: Department of Medicine	
<i>Senior/Key Personnel:</i>	<i>Organization:</i>	<i>Role Category:</i>
Anna Prizment Ph.D	University of Minnesota	PD/PI
Nathan Pankratz Ph.D	University of Minnesota	Co-Investigator
Heather Nelson Ph.D	University of Minnesota	Co-Investigator

## RESEARCH & RELATED OTHER PROJECT INFORMATION

OMB Number: 4040-0001  
Expiration Date: 10/31/2019

<p><b>1. Are Human Subjects Involved?*</b>   <input type="radio"/> Yes   <input checked="" type="radio"/> No</p> <p>1.a. If YES to Human Subjects</p> <p>Is the Project Exempt from Federal regulations?   <input type="radio"/> Yes   <input type="radio"/> No</p> <p>If YES, check appropriate exemption number:   1 __ 2 __ 3 __ 4 __ 5 __ 6 __ 7 __ 8lf</p> <p>NO, is the IRB review Pending?   <input type="radio"/> Yes   <input type="radio"/> No</p> <p>IRB Approval Date:</p> <p>Human Subject Assurance Number</p>
<p><b>2. Are Vertebrate Animals Used?*</b>   <input type="radio"/> Yes   <input checked="" type="radio"/> No</p> <p>2.a. If YES to Vertebrate Animals</p> <p>Is the IACUC review Pending?   <input type="radio"/> Yes   <input type="radio"/> No</p> <p>IACUC Approval Date:</p> <p>Animal Welfare Assurance Number</p>
<p><b>3. Is proprietary/privileged information included in the application?*</b>   <input type="radio"/> Yes   <input checked="" type="radio"/> No</p>
<p><b>4.a. Does this project have an actual or potential impact - positive or negative - on the environment?*</b>   <input type="radio"/> Yes   <input checked="" type="radio"/> No</p> <p>4.b. If yes, please explain:</p> <p>4.c. If this project has an actual or potential impact on the environment, has an exemption been authorized or an environmental assessment (EA) or environmental impact statement (EIS) been performed?   <input type="radio"/> Yes   <input type="radio"/> No</p> <p>4.d. If yes, please explain:</p>
<p><b>5. Is the research performance site designated, or eligible to be designated, as a historic place?*</b>   <input type="radio"/> Yes   <input checked="" type="radio"/> No</p> <p>5.a. If yes, please explain:</p>
<p><b>6. Does this project involve activities outside the United States or partnership with international collaborators?*</b>   <input type="radio"/> Yes   <input checked="" type="radio"/> No</p> <p>6.a. If yes, identify countries:</p> <p>6.b. Optional Explanation:</p>

## **PROJECT SUMMARY**

Immune biomarkers may be useful for early detection and prevention of colorectal cancer. However, very little is known about the potent immune pathway involving major histocompatibility complex (MHC) class I-like proteins in colorectal cancer. These proteins, expressed by gastrointestinal cells, serve as major ligands for natural killer cells. However, they may be cleaved from the cell surface and released into blood in a soluble form – the mechanism widely used by tumors to escape from immune surveillance even at early stages of colorectal cancer development. In addition, MHC class I-like genes are highly polymorphic with functional mutations changing the secretion of these proteins into bloodstream and altering the binding between tumor and immune cells. These biological pathways provide opportunities for blood-based colorectal cancer screening and development of immune therapies aimed at the neutralization of MHC class I-like proteins in blood, although the utility of these proteins as clinical biomarkers in colorectal cancer remain to be determined. Our main hypothesis is that MHC class I-like proteins contribute to the risk of colorectal cancer development. To test this novel hypothesis, we propose to conduct a secondary data analysis in a prospective population-based study – Atherosclerosis Risk in the Community cohort using existing information on six plasma MHC class I-like proteins measured three times over 20 years of follow-up, genotyping data in MHC class I-like genes and data about CRC and its risk factors. The use of existing data will allow for quick and cost efficient testing of our hypothesis. Understanding the role of MHC class I-like proteins in colorectal cancer is clinically important because inhibition of these proteins in blood would provide a novel opportunity for controlling colorectal tumor development and early detection.

## PROJECT NARRATIVE

This proposal is *relevant to public health* because understanding the role of immune-regulating proteins may inform novel strategies for early detection and prevention of colorectal cancer.

## FACILITIES & OTHER RESOURCES

### UNIVERSITY OF MINNESOTA

Founded in 1851 as a public land-grant university, the University of Minnesota is the state's only research university and a major center of education, creative scholarship, research, and service. The University of Minnesota is one of the largest universities in the U.S. with more than 66,000 students and some 4000 full-time faculty, making it one of the nation's largest campuses with a rich interdisciplinary research and teaching agenda.

#### **Academic Health Center (AHC)**

The AHC includes schools of medicine, public health, nursing, pharmacy, dentistry, and veterinary medicine. Accordingly, all the University and AHC resources and infrastructure important for supporting research efforts are available for this project. They include general, biomedical, statistical, and mathematics libraries, and computer and bioinformatics facilities. The Health Sciences library collection includes over 4000 current serials subscriptions, many of which are available online.

As a major research university, the University of Minnesota has a positive environment encouraging and facilitating faculty and students to engage in active teaching/learning and research, including in statistics, biostatistics, genetic epidemiology, human genetics and bioinformatics. Weekly seminars in the School of Statistics and Division of Biostatistics host nationally and internationally known speakers. The Academic Health Center (including the Schools of Public Health, Medicine, Pharmacy, Nursing, Dentistry, and Veterinary Medicine) offers rich opportunities for biomedical applications of developed statistical theory and methodology.

#### **Masonic Cancer Center**

The Masonic Cancer Center, designated by the National Cancer Institute as a comprehensive center, serves as the focal point for cancer research at the University. About 500 members apply their expertise to the broad problem of cancer with research in cancer causes, prevention, treatment, outcomes, and survivorship.

The Masonic Cancer Center supports cancer research through:

*Research programs:* Masonic Cancer Center research is focused in research programs that bring together scientists from different disciplines to discover processes that affect cancer. These programs are organized around scientific themes that reflect advances in cancer research and provide opportunities for interactions across the cancer community to solve organ-specific clinical questions. These programs include:

- Carcinogenesis and Chemoprevention
- Cell Signaling
- Genetic Mechanisms of Cancer
- Immunology
- Screening, Prevention, Etiology and Cancer Survivorship
- Transplant Biology and Therapy
- Tumor Microenvironment

*Translational Working Groups:* The Translational Working Groups (TWGs) are a mechanism to foster interprogrammatic and translational research to promote new discoveries. TWGs organize MCC members from various research programs into groups designed to address site-specific cancers, foster collaborations, and provide the necessary scientific and clinical expertise to improve outcomes for the specific disease sites. **Dr. Prizment** actively participates in the TWG on colorectal cancer

*Shared resources:* A key goal of the Masonic Cancer Center is to provide access to technologies, services and scientific consultation that facilitate interaction and enhance scientific productivity. The shared resources highlighted below provide stability, reliability, cost-effectiveness, and quality control that would be difficult to achieve otherwise. The Masonic Cancer Center's includes support for many of these resources.

- Analytical Biochemistry
- Biostatistics and Bioinformatics
- Clinical Pharmacology
- Comparative Pathology

Cytogenomics  
Flow Cytometry  
Genome Engineering  
Mouse Genetics Laboratory  
Oncology Medical Informatics & Services  
Translational Cell Therapy  
X-RAD Biological Irradiator

**University of Minnesota Medical School  
Department of Medicine  
Division of Hematology, Oncology and Transplantation**

**Dr. Prizment** is an Associate Professor at the Division of Hematology, Oncology and Transplantation (H.O.T). The Division of H.O.T provides an excellent environment for Dr. Prizment to conduct the research activities proposed in her grant application. Investigators work in teams where senior faculty members assist junior investigators in their progress.

Dr. Prizment, as well as the other investigators, has full access to the resources in the H.O.T Division including office space and computer services. In addition, the Division has developed strong collaborative interactions within various University facilities, including the CTSI, the Advanced Research and Diagnostic Laboratory and the Genomics Center, that support study design, data collection, and laboratory analysis. Most cancer research in Division is conducted in collaboration with the NCI-designated Comprehensive Masonic Cancer Center, located on the same campus. All members of the proposed research team are active members of this Masonic Cancer Center.

Scientific Environment: As a major research university, the University of Minnesota has a positive environment encouraging and facilitating faculty/students to engage in active teaching/learning and research, including statistics, biostatistics, genetic epidemiology, bioinformatics and human genetics. Dr. Prizment takes full advantage of the various seminar series on campus to network with peers in the field and also participates in weekly seminars in the H.O.T Division and Masonic Cancer Center, which host nationally and internationally known speakers. She also participates in monthly meetings of the Translational Working Groups on colorectal cancer in the Masonic Cancer Center.

**Department of Laboratory Medicine and Pathology  
Division of Computational Pathology**

**Office space**

Dr. Pankratz is director of the Division of Computational Pathology and a member of the Division of Molecular Pathology and Genomics who conducts research at the intersection of genetics, epidemiology, statistics, and computer science in the context of population studies. Dr. Pankratz, has been allotted a well-equipped, newly-renovated set of offices of approximately 400 square feet to house the PI and seven additional employees. Each person has a desktop machine with four dual-core processors with 32 GB of RAM that is well-networked to the university system, to email, to the Internet, and to the servers at the Minnesota Supercomputing Institute that will be used for analyses. The office is part of a larger complex that also contains other basic office equipment, such as a copy machine and a fax machine.

The Department of Laboratory Medicine and Pathology is supported by the Academic Health Center Information Systems (AHC-IS) with a comprehensive and unified set of computing resources, including hardware, software, and a full staff of systems and operations programmers. Services provided include specialized data acquisition (from laboratory devices and survey instruments), data management, statistical analysis, exploratory and presentation graphics, word processing, project management, electronic mail, and data archives. In addition, AHC-IS runs three Data Centers (Minneapolis & St. Paul campuses, and WBOB) housing servers and offering 24/7 support, including around the clock security intrusion monitoring and scheduled nightly off-site data backup.

**Minnesota Supercomputing Institute**

Investigators at the University of Minnesota have access to outstanding computing facilities and resources at the Minnesota Supercomputer Institute (MSI). This includes state-of-the-art supercomputers, such as Mesabi, an HP Linux cluster with 741 nodes of various configurations (with memory ranging from 64 GB to 1 TB of RAM) with a total of 17,784 compute cores provided by Intel Haswell E5-2680v3 processors, providing 711 Tflop/s of peak performance. The system has >3.5 PB (petabytes) of high-performance, high capacity primary storage, backed by QDR InfiniBand interconnect, as well as 1.4 PB of secondary storage. MSI also maintains computing laboratories that provide workstations and unique visualization capabilities for researchers. The Institute's experienced technical staff is comprised of systems administrators and user consultants (many with PhDs). They offer an impressive array of in-house technical expertise and services including advanced software designers with proven expertise in Java, C, C++, Python, Ruby on Rails, PHP, SQL, .Net, R and Perl as well as Oracle, MySQL, and PostgreSQL object relational data management systems. Assistance provided by the user consultants includes data analysis, visualization, tutorials and user training, code development for supercomputers, and expert help with a broad range of software packages, including many of the bioinformatics and statistical packages used by this team. A detailed list of the computing resources available through the Supercomputing Institute can be found on the Institute's website, [www.msi.umn.edu](http://www.msi.umn.edu).

### **School of Public Health (SPH)**

Heather Nelson is an Associate Professor in the Division of Epidemiology and Community Health, part of the School of Public Health. The University of Minnesota School of Public Health is among the top schools of public health in the nation. For the last four out of five years, the School has attracted more National Institutes of Health research funding than any other public university school of public health. Our innovative educational programs feature 11 different majors and dozens of opportunities for joint degrees and concentrations. The Division of Epidemiology and Community Health offers a unique multidisciplinary environment for training in epidemiology and has a long history of and commitment to training scientists, both nationally and internationally.

### **Clinical and Translational Science Institute (CTSI)**

The CTSI at the University of Minnesota offers comprehensive research support for clinical investigators, from concept through publication. The mission of the CTSI is to accelerate discoveries that will impact human health at the level of individuals and populations. The CTSI supports all stages of the research process, providing research and regulatory support, specialized facilities, and analytical services. The CTSI provides comprehensive clinical research facilities and support across specialty areas, tailored to the needs of the investigator that allow the investigator to effectively and efficiently leverage resources without unnecessary duplication. Strong relationships across the University's Academic Health Center provide access to extensive expertise and collaborative opportunities. Bruce Blazar, MD directs the CTSI and is the principal investigator on the CTSA grant.

Biomedical Informatics (BMI): BMI provides the expertise, tools, and resources for health research. The BMI function of CTSI is driving the integration of clinical data across the University of Minnesota and Fairview, and giving researchers a one-stop-shop for the data, tools, and information technologies they need to accelerate their research. The Informatics Consulting Service (ICS) enables clinical-translational researchers to improve their effectiveness, efficiency and impact by providing them with access to informatics tools, resources and expertise. ICS provides researchers with access to the clinical data repository of +2 million patients for research purposes, including the self-service tool i2b2, for discovering if there is a cohort of patients in the clinical data repository that meets the criteria of interest.

### **University of Minnesota Bio-Medical Library**

The University of Minnesota Bio-Medical Library's Reference and Instructional Services department engages faculty, students and staff in the learning, research and service missions of the University of Minnesota by providing broad instructional, reference, research and mediated search services. Our staff provides in-person, telephone, and digital reference assistance, conduct classes, actively partner with AHC professional schools and centers to further their respective missions, and manage the print and electronic collections that serve the AHC, the University and the State. The Bio-Medical Library is conveniently located in the Health Sciences complex. It contains over 460,000 volumes and subscribes to 3300 print periodicals and more than 3000 electronic periodicals. Via its website, it offers faculty and students computer access to searchable databases (e.g., PubMed) and provides access to online journals.



## EQUIPMENT

### Pankratz Laboratory

No major existing equipment is required for this project besides the PI's and programmers' desktops and the access to the **Minnesota Supercomputing Institute**, which has ample shared storage (>2 petabytes) that we can use (on an as needed temporary basis) to perform the proposed work.

### Minnesota Supercomputing Institute (MSI)

Mesabi:

Compute infrastructure: A total of 702 nodes of various configurations with a total of 16,848 compute cores provided by Intel Haswell E5-2680v3 processors. This system provides 674 Tflop/s of peak performance. 40 of these nodes include 2 NVidia Tesla K20X GPUs. The GPU subsystem provides 105 Tflop/s of additional peak performance.

Memory: 616 nodes have 64 GB of RAM, 24 nodes feature 256 GB, and 16 nodes have 1 TB of RAM each. The 40 GPU nodes have 128 GB of RAM each. Hence the total memory of the system is 67 TB.

SSD input/output nodes: 32 nodes have 480 GB solid-state drives (SSDs) for ultra-high performance

Input/Output. The total system SSD capacity is 15 TB.

Interconnect: Infiniband FDR (56 Gbit/s) throughout. The system is configured in islands of 192 nodes. Within the islands there is full bisection, using EDR switches (80 Gbit/s). Storage: 1 PB of high-performance storage hardware from Panasas in conjunction with this system.

Itasca: Itasca is an HP Linux cluster with 1,091 HP ProLiant BL280c G6 blade servers, each with two quad-core 2.8 GHz Intel Xeon X5560 "Nehalem EP" processors sharing 24 GiB of system memory, plus 51 HP ProLiant BL460c G8 blade servers, each with two 8-core 2.6 GHz E5-2670 Xeon "Sandy Bridge EP" processors with 64, 128 or 256 GiB of memory, with a 40-gigabit QDR InfiniBand (IB) interconnect. In total, Itasca consists of 9712 compute cores and 25 TiB of main memory.

Cascade: Cascade is a mixed GPU and PHI cluster. There are eight compute nodes with four Tesla GPUs per node and four nodes with dual Kepler GPUs. In addition, there are three nodes which feature a PHI co-processor.

Red Nodes: Red Nodes is an MSI Hadoop cluster to support Big Data analytics. It is composed of 50 nodes each with "Sandy Bridge EP" E5-2620 processors with six 2 GHz cores, QDR infiniband, eight GB of memory, and a 500 GB hard drive. MSI has configured 40 of these compute nodes into two Hadoop clusters, each with 20 nodes.

1. Vertebrate Animals Section

Are vertebrate animals euthanized?       Yes       No

If "Yes" to euthanasia

Is the method consistent with American Veterinary Medical Association (AVMA) guidelines?

Yes       No

If "No" to AVMA guidelines, describe method and provide scientific justification

.....

2. \*Program Income Section

\*Is program income anticipated during the periods for which the grant support is requested?

Yes       No

If you checked "yes" above (indicating that program income is anticipated), then use the format below to reflect the amount and source(s). Otherwise, leave this section blank.

\*Budget Period    \*Anticipated Amount (\$)    \*Source(s)

## PHS 398 COVER PAGE SUPPLEMENT

### 3. Human Embryonic Stem Cells Section

\*Does the proposed project involve human embryonic stem cells?       Yes       No

If the proposed project involves human embryonic stem cells, list below the registration number of the specific cell line(s) from the following list: [http://grants.nih.gov/stem\\_cells/registry/current.htm](http://grants.nih.gov/stem_cells/registry/current.htm). Or, if a specific stem cell line cannot be referenced at this time, check the box indicating that one from the registry will be used:

Specific stem cell line cannot be referenced at this time. One from the registry will be used.

Cell Line(s) (Example: 0004):

### 4. Inventions and Patents Section (Renewal applications)

\*Inventions and Patents:       Yes       No

If the answer is "Yes" then please answer the following:

\*Previously Reported:       Yes       No

### 5. Change of Investigator/Change of Institution Section

Change of Project Director/Principal Investigator

Name of former Project Director/Principal Investigator

Prefix:

\*First Name:

Middle Name:

\*Last Name:

Suffix:

Change of Grantee Institution

\*Name of former institution:

<b>Introduction</b> 1. Introduction to Application (for Resubmission and Revision applications)	
<b>Research Plan Section</b>	
2. Specific Aims	Prizment_SA.pdf
3. Research Strategy*	Prizment_Research_0626.pdf
4. Progress Report Publication List	
<b>Other Research Plan Section</b>	
5. Vertebrate Animals	
6. Select Agent Research	
7. Multiple PD/PI Leadership Plan	
8. Consortium/Contractual Arrangements	
9. Letters of Support	LettersG.pdf
10. Resource Sharing Plan(s)	Prizment_ResourceSharing.pdf
11. Authentication of Key Biological and/or Chemical Resources	
<b>Appendix</b>	
12. Appendix	

## SPECIFIC AIMS

Our long-term goal is to understand how the immune system can be harnessed to reduce the incidence and fatality of colorectal cancer (CRC) – the 2<sup>nd</sup> leading cause of cancer deaths in the U.S. **The objective of this study** is to determine the role of MHC class I-like proteins (MICA, MICB, ULBP1, ULBP2, ULBP3, and ULBP4) in CRC development. These proteins serve as ligands for the immune cells (NK and CD8+) and play a mechanistic role in immune response. In healthy individuals, the proteins are primarily expressed in epithelial gastrointestinal cells, and overexpressed on tumor cells starting at the early stage of cancer development. The secretion of these proteins into blood hinders the binding between immune and tumor cells, precluding tumor elimination<sup>1,2</sup>. Soluble MICA (s-MICA) and other MHC class I-like proteins are detectable in blood even in healthy individuals. We<sup>3,4</sup> and others<sup>5,6</sup>, observed higher levels of s-MICA in those with several cancer types versus those without. We hypothesized that elevated levels of s-MICA and other MHC class I-like proteins could be used to detect early cancer. In addition, using the funding from an American Cancer Society award<sup>3,4</sup>, we showed that the presence of certain mutations in the MICA gene, microsatellite repeats and SNPs, are associated with higher plasma s-MICA levels. Among cancer-free participants, the presence of a single and two copies of A5.1 variant in the transmembrane region of MICA protein was associated with a 24% and 38% increase in s-MICA levels, respectively, suggesting a dose response<sup>3</sup>. Hence, the presence of this mutation results in increased MICA secretion into bloodstream and thus may impede elimination of tumor cells.

Few studies examined the MHC class I-like proteins in CRC<sup>7-10</sup>. Those studies focused on the proteins' role in CRC progression, and reported an increased mortality of CRC patients with high s-MICA<sup>10</sup>. To our knowledge, no studies have prospectively examined the contribution of these proteins to the risk of CRC development, although the presence of functional mutation (MICA 129-Met/Val) was shown to predispose to CRC<sup>9</sup>. Thus, **our main hypothesis** is that higher plasma levels of MHC class I-like proteins are associated with increased risk of subsequent CRC. We will conduct a secondary data analysis in the Atherosclerosis Risk of Community (ARIC) study – the prospective biracial cohort of ~13,000 men and women that has collected detailed information on cancer (360 CRC cases diagnosed in 1990-2015). The ARIC study is uniquely suited to examine this hypothesis, because six MHC class I-like proteins (MICA, MICB and ULBP1, ULBP2, ULBP3, and ULBP4) have been measured or being measured by SOMAScan – a novel multi-protein panel – in frozen plasma samples collected three times during follow-up. Additionally, ARIC has collected genotyping data based on deep whole genome sequencing (WGS, x30-40) and genome-wide association studies (GWAS) in the majority of participants. Our preliminary analysis showed that several SNPs in the MICA gene are associated with CRC risk in the ARIC cohort. Now, we are conducting an analysis of SNPs and microsatellite repeats in several MHC-class I-like genes in relation to CRC. Finally, ARIC data include a wealth of information on CRC risk factors, stage at diagnosis, and first course of treatment that will allow accounting for potential confounders and effect modifiers. The specific aims for this study are as follows:

**Specific aim 1.** Determine whether the risk of CRC development is associated with pre-diagnostic plasma levels of MHC class I-like proteins (including MICA, MICB, ULBP1, ULBP2, ULBP3, and ULBP4)

- a) at baseline of this study (1990-92);
- b) over time (changes in levels between visits but before diagnosis)

*Hypothesis: Higher levels and greater changes in these protein levels are associated with higher risk of subsequent CRC development.*

**Specific aim 2.** Identify demographic and lifestyle factors (e.g., age, race, sex, smoking, BMI) and genetic variants in cancer-free ARIC participants that contribute to levels of the plasma MHC-class I proteins (described in SA1)

- a) at baseline of this study (1990-92);
- b) over time (changes in levels between visits).

*Hypothesis: Specific functional mutations (e.g. MICA A5, A5.1, Val123Met) in these genes and other risk factors for CRC such as older age are associated with higher levels of MHC class I-like proteins at baseline and accelerated changes in the levels over time.*

**Secondary aim.** Explore an association between MHC class I-like genetic variants associated with CRC, the levels of the corresponding proteins and CRC risk using Mendelian randomization analysis. This method will allow us to assess a causal effect of plasma MHC class I-like proteins on CRC risk.

We expect this study to elucidate the role of these immune proteins in tumor surveillance and evasion mechanisms. Identifying the impact of circulating MHC class I-like proteins levels on CRC risk could help identify those who will benefit from frequent CRC screening or from modulating MICA secretion<sup>11-13</sup>.

## RESEARCH STRATEGY

### A. SIGNIFICANCE

**Scientific premise and rigor of prior research.** Emerging evidence shows that CRC tumors may be recognized by the immune system, and the cancer development may be inhibited by enhanced immune response<sup>14</sup>. The critical unanswered questions in the immunosurveillance of CRC are whether inter-individual differences in host immune defense could predict future CRC and whether it is possible to harness immune system to inhibit CRC development. One of the most important interactions between immunity and tumors involve MHC class I-like proteins that are over-expressed on tumors, recognized by immune cells, and detectable in blood even in healthy individuals. Thus, these proteins may hold promise for prevention and early detection of CRC, especially among those at high CRC risk, if coupled with established screening methods or other biomarkers.

MHC class I-like proteins include several proteins – MICA, MICB and ULBP1, ULBP2, ULBP3, all of which bind to activating receptor (called NKG2D) on immune cells, mainly NK cells, and most likely have similar role in cancer. Thus, to discuss the role of these proteins in this proposal, we will use MICA, the best studied protein in this family, as an example.

*In vitro* and *in vivo* studies showed that MICA expression on tumor cells and binding of MICA to the NKG2D receptor promotes the tumor lysis by NK and T cells<sup>1</sup> resulting in lower stage at the cancer diagnosis and better prognosis for cancer patients<sup>6,15</sup> including those with CRC<sup>7,10</sup>. However, tumor cells evade this immune response through MICA shedding via proteolytic cleavage and its release into circulation. The shedding of MICA into the blood leads to the surface depletion of MICA, downregulation of NKG2D and weakening of MICA-NKG2D binding<sup>15-17</sup>, all of which decrease the ability of immune cells to recognize and kill tumor cells (Fig. 1). Consistent with this mechanism, higher blood levels of soluble MICA (s-MICA) have been correlated with advanced cancer stage at diagnosis, metastasis, and worse prognosis of CRC, pancreatic, liver, oral and several other cancers<sup>2,6,10,15</sup>. Clinical studies have shown that it is possible to prevent tumor formation by suppressing s-MICA accumulation<sup>18</sup>. Moreover, approaches are being developed that inhibit MICA shedding via targeting proteases responsible for MICA cleavage<sup>1,17,19</sup> or reduce s-MICA levels using anti-MICA antibodies<sup>20</sup>.

Despite the important immune-regulating role of MHC class I-like proteins in cancer, to our knowledge, no studies prospectively examined the contribution of these proteins to the risk of subsequent CRC in initially cancer-free people. Given that MICA shedding is a critical mechanism by which colorectal tumors escape from immunosurveillance<sup>1,17</sup> and that targeting this mechanism has shown promise in immunological treatment, the role of the s-MICA and other MHC I class-like proteins needs to be comprehensively elucidated in CRC development.

**The mechanisms underlying the interaction between MHC class I-like ligands and cancer** are partially explained by the variants in the relevant genes (Fig. 1). MICA and MICB as well as ULBP molecules are located within/near the human leukocyte antigen (HLA/MHC) locus and are highly polymorphic, and allelic variation can alter their expression levels or their affinity for NKG2D (reviewed in <sup>11,21,22</sup>). As a result, NKG2D-L polymorphisms may strongly affect immune-cell binding and elimination of tumor cells or other stressed targets<sup>23</sup>. To date, 107 MICA and 47 MICB alleles have been described (updated allele numbers can be found at <http://hla.alleles.org/nomenclature/index.html>). There are two major types of functional polymorphisms in these genes including (1) SNPs located mainly in extracellular domains, through which these ligands bind to NKG2D and (2) microsatellite repeats (also called short tandem repeats) located in the transmembrane domain (e.g. exon 5 in MICA has 8 alleles: A4, A5, A5.1, A6, A7, A8, A9, and A10). Both of these mutation types may weaken the affinity of MICA-NK binding and increase s-MICA levels and thus diminish the cytotoxic response of immune cells and promote tumor progression<sup>1,24</sup>. For instance, rs1051792 (Val129Met) dimorphism affects NKG2D receptor affinity: MICA-129Met variant has an affinity 10- to 50-fold higher than MICA-129Val,<sup>9</sup> while transmembrane microsatellite repeats such as A5.1 (the frequency is 30% in the population) affects the secretion of s-MICA into blood, as have been shown in our and other studies<sup>3</sup>, and therefore interferes with immune recognition by decreasing the number of available ligands for detection and by hindering the access of

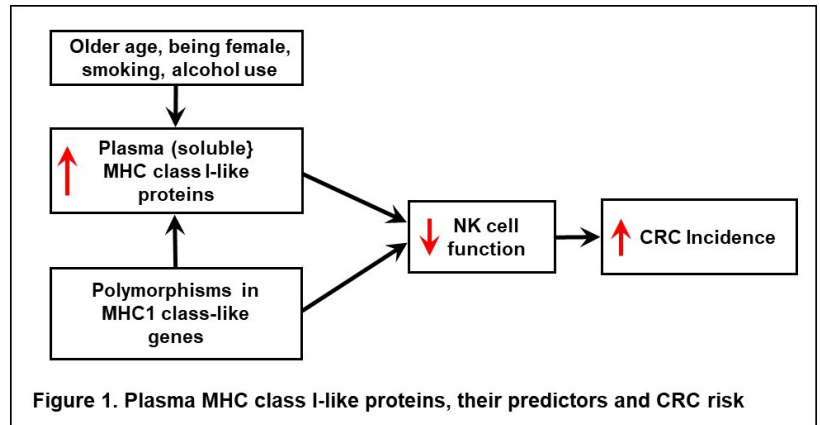


Figure 1. Plasma MHC class I-like proteins, their predictors and CRC risk

immune cells to the tumor. In our previous study, that mechanism most likely explained an observed association between the presence of MICA A5.1 variants and increased pancreatic cancer risk<sup>3</sup>.

In preparation for this study, we conducted a preliminary analysis of MICA SNPs in CRC in the ARIC study and found several polymorphisms to be associated with increased CRC risk (Table 1). In addition, we showed that several microsatellite repeats are associated with increased s-MICA levels in cancer-free participants in the pancreatic case-control study (Table 2)<sup>3</sup>.

**The critical and unique features** of our proposal include comprehensive examining of plasma MHC class I-like proteins in a novel and cost-efficient way, since all of these proteins have been already measured (or are being measured) three times as part of highly sensitive proteomic analysis by SomaLogic (<https://somalogic.com/technology/our-platform/>) in the ARIC study. The availability of these measures along with already conducted GWAS and deep WGS in the same study and information on CRC and its risk factors will enable us to study the mechanisms underlying the effect of MHC class I-like proteins in CRC.

The insights gained by this study will provide a scientific foundation for a larger NIH grant application that will replicate this study and test whether these protein biomarkers improve the prediction of CRC. Although s-MICA has been studied as a potential clinical biomarker in liver and pancreatic cancer<sup>25,26</sup>, we realize that these proteins are unlikely to be used for CRC screening by themselves. However, they may be potentially used to complement colonoscopy in the individuals at high CRC risk or in combination with other stool biomarkers, since, contrary to many other biomarkers, these proteins are produced at a very early stage of tumor development. Another distinguishing feature of these biomarkers is their direct functional role in immune response in cancer that may be utilized in the future for immunological targeting of high-risk CRC adenoma. Establishing novel biomarkers is essential now when CRC is becoming a disease of younger adults and new guidelines for prevention and screening should be developed.

## B. INNOVATION

1. Testing an association between the plasma levels of MHC class I-like proteins and the risk of subsequent CRC development is a novel hypothesis that has not been researched before. We will also examine the associations specifically in African-Americans. To our knowledge, MHC class I-like proteins have not been studied in this population that have higher CRC incidence.
2. Another essential feature of our study is the use of the novel, highly sensitive multi-protein assay – SOMAscan (by SomaLogic) to measure all six MHC-I class-like proteins in the same samples, which will enable an integrated analysis of these biomarkers.
3. Our study will provide a new comprehensive approach to examining the role of MHC class I-like proteins in cancer, since we will examine the MHC-I class-like proteins together with the corresponding genetic mutations and CRC, which will allow us to explore a causative role of these proteins in CRC.

This study is expected to establish a novel link between the immunosurveillance and CRC risk. The pathway leading to changes in expression and secretion of NKG2D ligands may be a productive target for design of therapeutic agents to enhance the immunogenicity of tumor cells at early stage of their development.

SNPs	Location on/near MICA gene	Reference/other allele (frequency of ref. allele)	RR (95% CI)
rs1051792 (Val123Met)	exon 3	A/G (29%)	1.19 (0.96; 1.48)
rs1051794	exon 3	A/G (29%)	1.19 (0.96; 1.48)
rs1131896	exon 4	A/G (24%)	0.97 (0.78; 1.21)
rs1063635	exon 4	A/T (45%)	1.10 (0.91; 1.33)
rs2516448	adjacent, in LD with A5.1	C/T (64%)	1.12 (0.93; 1.35)
rs3763288	promoter	A/G (6%)	1.08 (0.74; 1.55)
rs2395029	near the MICA gene (HLA complex P5 gene)	G/T (3%)	2.45 (1.17; 5.10)

MICA STR genotypes	s-MICA levels (Geometric Mean)	P-value
MICA A4 X/X X/A4 or A4/A4	67.96 (66.51 - 69.45) 30.11 (28.35 - 31.98)	<0.0001
MICA A5 X/X X/A5 or A5/A5	65.70 (64.26 - 67.17) 54.78 (52.78 - 56.85)	<0.0001
MICA A6 X/X X/A6 or A6/A6	61.00 (59.59 - 62.45) 70.45 (68.41 - 72.55)	<0.0001
MICA A9 X/X X/A9 or A9/A9	73.78 (71.86 - 75.75) 35.09 (32.70 - 37.66)	<0.0001
MICA A5.1 X/X X/A5.1 or A5.1/A5.1	49.42 (47.44 - 51.48) 67.13 (65.64 - 68.65)	<0.0001

## C. APPROACH

Our overall strategy is to conduct a secondary data analysis within the existing ARIC cohort that will examine the role of plasma MHC class I-like proteins in relation to CRC development and establish the determinants of their levels.

### Overview of the study design

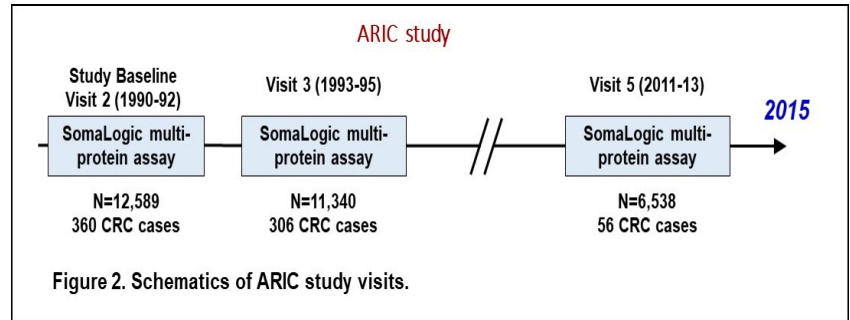
This longitudinal analysis will take advantage of the wealth of data collected in the ARIC study – a longstanding prospective cohort of 15,792 White and Black men and women (45 - 64 years old at Visit 1 in 1987– 1989) recruited from 4 US communities: Forsyth County, North Carolina; Jackson, Mississippi; suburban Minneapolis, Minnesota; and Washington County, Maryland. Each visit included interviews, laboratory measurements, and clinic examinations (<https://sites.csc.unc.edu/aric/>).<sup>27</sup> Participants were asked to report their demographic characteristics, education, lifestyle behaviors, and medical history at each follow-up. Trained personnel collected anthropometric measures and blood (serum and plasma) samples that were frozen, stored under controlled conditions, and have been never thawed. Incident cancers were ascertained from 1987 through 2015 by linkage with state cancer registries in Maryland, Minnesota, Mississippi, and North Carolina, and by additional active follow-up of the cohort, including cases diagnosed before these cancer registries were established<sup>27</sup>.

Genotyping was performed at the Broad Institute of MIT and Harvard as part of the Gene Environment Association Studies initiative (GENEVA, <http://www.genevastudy.org>) funded by the trans-NIH Genes, Environment, and Health Initiative (GEI). Data cleaning and harmonization were done at the GEI-funded GENEVA Coordinating Center at the University of Washington. Whole genome sequencing in ~12,000 participants in the ARIC studies have been conducted by Trans-Omics for Precision Medicine Program (TOPMed in NHLBI, about half of all samples have been studied) and by Centers for Common Disease Genomics (NHGRI, the other half).

ARIC has recently established an agreement with SomaLogic company to measure 4,931 plasma proteins using SOMAscan (v.4) in the ARIC cohort with stored blood samples and full consent at Visits 2, 3, and 5 (Fig. 2). These proteomic data are available to ARIC investigators at no cost, making the proposed research and discovery highly cost-effective.

**SOMAscan assays and quality control.** The SOMAscan is a highly multiplexed aptamer-based proteomic assay<sup>28</sup>. It utilizes chemically modified, single-stranded DNA about 40 nucleotides long (i.e. SOMAmers) as specific, high affinity binding reagents to protein targets. The proteins are measured based on the quantity of protein-bound SOMAmers detected as relative fluorescent units<sup>28,29</sup>. This assay captures proteins across a wide range of abundance and extends the lower limit of detection beyond traditional assays<sup>29</sup>. The SOMAscan assay has excellent reproducibility (median coefficient of variation [CV] of 4% to 8%)<sup>30–32</sup>, and is robust to short-term sampling procedures (e.g. from centrifugation to freezing)<sup>33</sup>.

The SOMAscan v 4.0 used in the ARIC includes 4,931 human proteins. First, the protein measures undergo the SOMAscan standardization and normalization process<sup>34</sup>. Briefly, hybridization control normalization is being applied to each sample based on a set of hybridization control sequences to correct for systematic biases during hybridization. Second, median signal normalization is applied to protein measures within a plate to remove sample or assay biases due to pipetting variation, variation in reagent concentrations, assay timing, and other sources of systematic variability within a single plate run. Finally, each plate contains calibrator samples for each SOMAmer reagent to correct for plate-to-plate variation based on established global reference standards. To date, for the SOMAscan v.4, the lifetime median total precision (i.e. CV%) over 220 runs/660 replicate QC samples for ~5000 SOMAmers is 6.2%. To detect potential batch effects, ARIC includes 26 pooled blinded plasma quality control samples in triplicate in all batches. Also, ARIC includes blind duplicate samples to evaluate reproducibility and drift in quality over time as it has collected ~5% blind split-specimen duplicates at the time of blood drawing. The proteins in frozen plasma samples from Visits 3 (1993-95) and 5 (2011-13) have been measured and proteomic analysis of frozen plasma samples from Visit 2 (1990-92) is being conducted now.





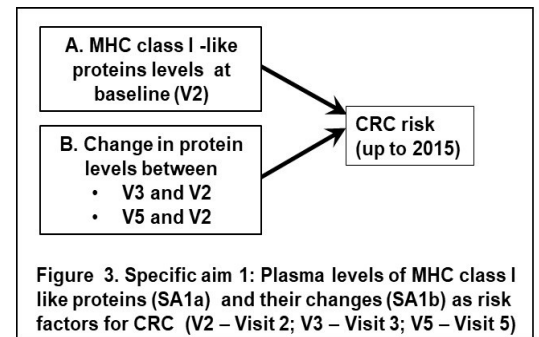
We have conducted a **pilot study** of SOMAscan v.3 assay consisting of 4,001 aptamers in 42 ARIC participants and obtained a median CV of 5%, Spearman  $r$  of 0.89, and intra-class  $r$  of 0.96 based on a blinded split sample analysis<sup>35</sup>. The levels of these proteins were stable over a period of 6 to 8 weeks with only one protein showing significant change.

### **Statistical analysis**

As a first step in studying these novel proteins, descriptive statistics will be conducted that will examine the distribution, correlations, and trajectories of MHC class I-like proteins (MICA, MICB, ULBP1, ULBP2, ULBP3, and ULBP4) over time collected at Visits 2, 3 and 5. Since these proteins were measured at Visit 2 (1990-92) for the first time, this visit will be considered as a baseline for this study. The analyses will be conducted using SAS and Stata (Stata Corporation, College Station, Texas)

**Specific aim 1a.** *Determine whether the risk of CRC development is associated with the plasma levels of MHC class I-like proteins at study baseline (1990-92) (Fig. 3A).*

For this analysis, participants who were diagnosed with cancer before the study baseline date will be excluded. Multivariable Cox proportional hazard regression model will be used to estimate hazard ratios and 95% confidence intervals (CI) for CRC risk. Person-years will be estimated from the start of follow-up until the date of cancer diagnosis, death or the end of follow-up, whichever occurs first. The proportional hazards assumption will be tested, and if violated, we will model non-proportional hazards or use generalized gamma regressions. The levels of biomarkers will be log-transformed to reduce the skewness in their distributions.



As our main analysis, we will conduct a factor analysis with a goal to decrease the number of variables and create a weighted score of the 6 biomarkers of interest. Based on the published data, the correlation between soluble MHC class I-like proteins in healthy individuals is expected to be ~0.4-0.5<sup>5</sup>. We will use the principal axis method to extract the factors (also called components) using log-transformed biomarkers. Given the limited number of correlated biomarkers ( $n=6$ ), the first factor will be retained (eigenvalue >1). The first factor represents a linear combination of the variables accounting for the maximum of the data variance<sup>36</sup>. A factor score based on this biomarker pattern will be calculated by summing concentrations of the biomarkers weighted by their factor loadings. The association between factor scores presented as continuous variable and categorized into quartiles and CRC risk will be examined.

All the analyses will be adjusted for baseline demographic factors (age, sex, race, center, and education) and other risk factors for CRC if they are also associated with plasma levels of MHC class I-like proteins. Potential confounders will include body mass index (BMI), physical activity, smoking status and pack-years, diabetes, use of alcohol, aspirin use, intake of fiber and red meat, and estimated glomerular filtration rate (eGFR), because ARIC preliminary analysis showed that this eGFR affects SOMAscan protein levels. A sensitivity analysis will be conducted that will exclude participants diagnosed within two or five years after blood collection to test whether the undiagnosed cancer at baseline affected the levels of biomarkers. Since in our previous analysis, we showed that the association between s-MICA levels and pancreatic cancer differed by sex with a stronger association in males<sup>4</sup>, we will examine an interaction with sex in relation to CRC risk. Also, we will test whether an association differs by race, site (colon vs rectal) and length of follow-up, categorized at median.

In an exploratory analyses, we will test the association between each individual biomarker (log-transformed) and CRC risk to determine proteins most strongly associated with CRC risk after adjusting for confounders described above. In addition, each biomarker will be presented as a categorical variable, a separate category for participants with undetectable values will be created, and its association with CRC risk will be tested. Finally, a sensitivity analysis will be conducted by comparing findings from the analysis including all the participants (with both non-missing and missing values) and the analysis including only those with non-missing values. If more than 10% of values are missing for any protein, we will impute missing values using inverse-probability weights.

**Power analysis for SA1a:** With 12,589 participants at the study baseline (without cancer) and 360 CRC cases, we will be able to detect a hazard ratio (HR) of 1.24 for one standard deviation of the factor score presented as continuous and HR of 1.58 presented as tertiles with a power of 80% (two-sided  $\alpha = 0.05$ )

**Specific aim 1b.** Determine whether the risk of CRC development is associated with the change in the plasma levels of MHC class I-like proteins (Fig. 3B).

We will test whether the change in the pre-diagnostic levels of biomarkers over time is associated with CRC risk, since it is unclear whether the baseline protein levels of protein or their change are most important for the development of cancer. Similar to SA1a, the factor analysis will be our main analysis; marginal structural Cox hazards model<sup>37</sup> will be used to estimate hazard ratios and 95% CI for CRC risk. We will account for time-dependent confounding due to BMI, smoking, and other variables (as discussed in SA1a) through inverse probability of treatment weighting<sup>37</sup>. Separate inverse probability of treatment weights will be generated to model a short-term change in time-dependent covariates between Visits 2 and 3. The follow-up for this analysis will start at Visit 3, and all the participants with cancer up to Visit 3 will be excluded. The exploratory analysis will examine a long-term change between Visits 2 and 5 with the start of follow-up for cancer at Visit 5. The interaction with several variables as discussed in SA1a will be tested. In addition, we will use marginal structural Cox hazards model to explore the changes in individual plasma biomarkers in relation to CRC risk.

**Power analysis for SA1b:** With 11,340 cancer-free participants at Visit 3 and 306 CRC cases, we will have 80% power to detect a hazard ratio (HR) of 1.26 associated with one standard deviation in the change (between Visits 3 and 2) of factor score presented as continuous variable and HR of 1.65 for the factor score presented as quartiles.

**Specific aim 2a.** Identify environmental (demographic and lifestyle) and genetic characteristics in cancer-free ARIC participants that contribute to the levels of the plasma MHC-class I proteins at the study baseline (1990-92) (Fig. 4a).

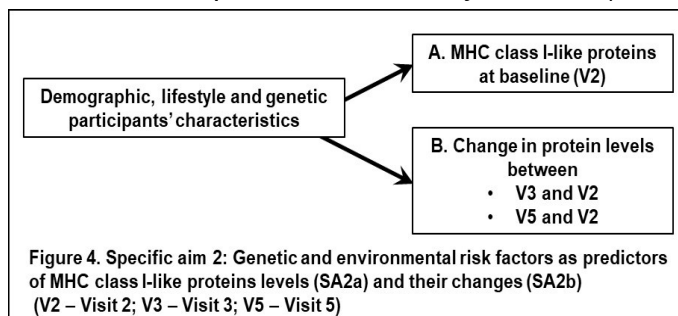
Linear regression model will be used to determine characteristics associated with each MHC-class I protein. The protein levels will be log-transformed.

The following characteristics will be considered: age, sex, race, BMI, smoking, physical activity, alcohol and aspirin use, diabetes, and the genetic variants including SNPs and microsatellite repeats. The SNPs in MHC class I-like genes (such as rs1051792, rs1051794, rs1131896, rs1063635, rs2516448, rs3763288, rs2395029 and rs3128982 in the *MICA* gene<sup>22</sup> and rs2516400, rs2855804, rs2855812, rs3132464, rs3132468, and rs3828903 in the *MICB* gene)<sup>38,39</sup> have been already genotyped in the ARIC study. Microsatellite repeats in these genes (such as A4, A5, A6, A7, A8, A9, A10 and A5.1 in the *MICA* gene)<sup>21,40</sup> are being currently identified in the ARIC study. Characteristics that showed a univariate association with the biomarker will be included into the multivariable model to identify those associated with that biomarker after adjusting for other covariates. Genetic variants will be also included into the model if they are not in linkage disequilibrium. In addition, we will compute a percent of variance explained by genetics for each protein via comparing the regression model with and without genetic variants. To account for multiple testing for six biomarkers, we will use Bonferroni correction with a significance level of  $0.05/6=0.008$ .

**Power analysis for SA2a.** In this this aim, a Bonferroni correction ( $\alpha=0.05/6=0.008$ ) will be applied to all the analyses. In the analysis of predictors at baseline among 12,589 cancer-free participants, we will have 80% power to detect a 2.81 pg/mL change in mean of biomarkers levels

**Specific aim 2b.** Identify environmental (demographic and lifestyle) and genetic characteristics in cancer-free ARIC participants that contribute to the short- and long-term changes in plasma levels of MHC class I-like proteins over time (Fig. 4b).

Mixed effect regression models with random coefficients (random intercept, random linear time) will be used to model the change in biomarker levels over time, and determine characteristics associated with short- and long-term changes. Similar to SA1, short-term change will be examined between Visits 3 and 2 and a long-term change between Visits 5 and 2. The potential risk factors discussed in SA2a will be considered. In addition to including baseline lifestyle characteristics, we will account for the change in the characteristics (such as BMI, smoking, diabetes, use of alcohol and aspirin) using inverse-probability weighting. Separate inverse-probability weights will be generated to model a short-term changes in time-dependent characteristics between Visits 2 and 3, and long-term change between Visits 2 and 5.





# PHS HUMAN SUBJECTS AND CLINICAL TRIALS INFORMATION

OMB Number: 0925-0001 and 0925-0002

Expiration Date: 03/31/2020

Are Human Subjects Involved

Yes  No

Is the Project Exempt from Federal regulations?

Yes  No

Exemption Number

1  2  3  4  5  6  7  8

Does the proposed research involve human specimens and/or data

Yes  No

If Yes, provide an explanation of why the application does not involve human subjects research

Prizment\_JustifNoHumanS.pdf

Other Requested information

## **JUSTIFICATION THAT NO HUMAN SUBJECTS ARE INVOLVED**

This proposal does not constitute research involving human subjects as defined by the NIH Office for Human Research Protections (OHRP) Guidance on Research Involving Coded Private Information or Biological Specimens.

### **Who is providing the data/biological specimens and their role in the proposed research**

The proposed study will involve accessing and analyzing de-identified existing data from a prospective cohort: Atherosclerosis Risk in Communities Study (ARIC, Contract PI: Aaron Folsom)). There will be no contact of participants, and no additional data collection.

### **Description of the identifiers that will be associated with the human specimens and data**

All participants in the ARIC study were assigned a unique identifier at the time of enrollment and only the participants' identifier will be used for this study. These assigned identifiers are linked to the data (demographic and lifestyle characteristics, protein and genetic data) that will be used in this proposal.

### **List of who has access to subjects' identities**

ARIC study will only share de-identified data with the investigators and staff involved in this research.

### **Manner in which the privacy of research participants and confidentiality of data will be protected.**

Participant data privacy and confidentiality will be protected through a multi-tiered approach covering data transmission, data handling, and data distribution processes to ensure anonymity both during and after the study. Individual identifiers are not part of the study database and reside only with the Coordinating Center of the ARIC study.

## **RESOURCE SHARING PLAN**

To ensure that maximal scientific benefit is derived from the proposed project, descriptive and association data will be rapidly and efficiently shared, consistent with NIH established policy. Descriptive and association data will be made available to the wider scientific community as set out by the NIH, Access to the study dataset and will be governed by policies developed by the leadership and publication committees of Atherosclerosis Risk in Communities Study (ARIC). These policies have been developed to maximize interactions will all investigators.