ARIC Manuscript Proposal #4183

PC Reviewed: 1/10/23 **Status: _____** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title: Genome wide association study of plasma eicosanoids in a community-based population

b. Abbreviated Title (Length 26 characters): Eicosanoid GWAS

2. Writing Group:

Writing group members: Aditya Surapaneni, Pascal Schlosser, Eugene Rhee, Susan Cheng, Mohit Jain, Mona Alotaibi, Josef Coresh, Bing Yu, Bing Yu's student, Morgan Grams, *others welcome* (order TBD).

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

First author: Eugene Rhee
Address: 55 Fruit Street, Massachusetts General Hospital
Boston, MA 02114
E-mail: eprhee@partners.org

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: Morgan Grams
Address: 2024 E. Monument Street, Suite 2-638
Baltimore, Maryland 2187
Phone: 443-287-1827 Fax: 410-955-0485
E-mail: mgrams2@jhmi.edu

3. Timeline: Analyses will begin once the manuscript proposal has been approved. We anticipate that the manuscript will be written and submitted to the ARIC Publications Committee within one year of the manuscript proposal being approved.

4. Rationale: Eicosanoids are metabolites derived from the oxidation of arachidonic acid or other polyunsaturated fatty acids. They function in diverse physiological systems and pathological processes including inflammation, immune responses, endothelial function, and hemostasis.[1] There are multiple subfamilies of eicosanoids, including prostaglandins, thromboxanes, leukotrienes, and hydroxyeicosatetraenoic acids (HETEs), among others. The

major enzyme families involved in the synthesis of eicosanoids are known (Figure, page 6), and include cyclooxygenases (COXs), lipoxygenases (LOXs), and cytochrome P450 enzymes. However, to what extent genetic variation at these loci modulate blood levels of eicosanoids, and whether there are other genetic regulators, for example relevant to eicosanoid elimination, is unknown.

Recent advances in LC-MS based profiling now permit the high throughput measurement of hundreds of eicosanoids and eicosanoid-related metabolites in samples obtained from large human cohorts,[2] providing an opportunity to conduct a genome wide association study (GWAS) of circulating eicosanoids. Several GWAS of other circulating blood metabolites ("mGWAS") have been conducted to date. These studies have shown that many loci have relatively large effect sizes on metabolite levels, as compared to GWAS for common diseases. As a result, significant findings have emerged from samples as small as 284 subjects.[3] Many of these loci encode enzymes or transporters directly involved with the given metabolite's disposition, highlighting potential therapeutic targets. Other associations have been at loci previously associated with a complex disease, raising the possibility that the metabolite is a causal participant in the disease process. Building on this framework, we seek to conduct a GWAS of plasma eicosanoids, with the goal to understand the genetic determinants of circulating eicosanoid levels.

5. Main Hypothesis/Study Questions:

Our overarching hypothesis is that many eicosanoids will have identifiable genetic determinants that provide insight into their synthesis, transport, and/or metabolism.

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: We will conduct analyses of the ARIC cohort, treating Visit 2 (1991-1993) as the baseline visit.

Study Population: The study population will consist of European-American (EA) and African-American (AA) ARIC participants with eicosanoid data from Visits 2.

Exposure: Genetic markers in the genome (SNPs)

Outcomes: Plasma eicosanoid levels, measured by LC-MS

Statistical Analysis: We will use descriptive statistics, including means, medians, and proportions to summarize baseline characteristics at Visit 2. We anticipate that the distributions of eicosanoids will be skewed: we plan to transform (e.g., log base-2 or inverse-normal) to achieve a more normal distribution. Genotype data will be imputed to a common set of SNPs using TopMed (Freeze 5 on GRCh38) as a reference panel using the Michigan Imputation

Server.[4] SNPs will be removed if they are not bi-allelic, have poor imputation quality (<0.1), or have minor allele frequency <1%.

We will perform genome wide association studies on eicosanoids among individuals with genetic and eicosanoid data, examining EA and AA cohorts separately. The association between eicosanoids and the genetic variants will be estimated by linear regression, adjusted for age, sex, and the first five genetic principal components (PCs) and eicosanoid PCs, using Fast Association Tests (FAST) software.[5] We chose five PCs on the basis of previous studies, but in sensitivity analyses, we will also evaluate results adjusted for ten genetic and eicosanoid PCs. Statistical significance will be set at a threshold of 5×10^{-8}/# of eicosanoids according to Bonferroni adjustment in the EA cohort; we will explore meta-analysis with results in the replication AA cohort given our observation of preserved effect sizes in previous metabolite GWAS. For each eicosanoid, we will identify the index SNP as the variant with the lowest p value within a 1 Mb genomic radius. Index SNPs will be annotated through linkage with the SNiPA web-tool based on the 1000 Genomes phase 3 v5 and Ensembl v87 datasets.[6,7]

For index SNPs, we will conduct statistical fine-mapping using SuSiE.[8,9] Colocalization with gene expression signals in tissue and whole blood will be performed using the coloc.fast function from the R package 'gtx' with default parameters using the GTEx V8 data.[10] In addition, we will perform colocalization analyses of the replicated SNPs with existing GWAS summary statistics of 778 phenotypes and diseases performed in 450,000 UK Biobank participants using the coloc.fast function.[11]

Limitations: We acknowledge that our proposed study has a few limitations. First, the accuracy of eicosanoid identification is not always known. Second, we will not have replication in an external cohort. Third, our study will not consider rare variants.

7.a. Will the data be used for non-CVD analysis in this manuscript? ____ Yes ___X_ 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? __X__ Yes ____ 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"? __X__ Yes ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status.
ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/ARIC/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)?

Manuscript # 4005: Eicosanoids and kidney outcomes in a community-based population

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*__2015.09___)

_ _ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* __2011.03_(Selvin for funding on visit 6 labs, Matsushita for funding of visit 3 labs))

*ancillary studies are listed by number at http://www.cscc.unc.edu/aric/forms/

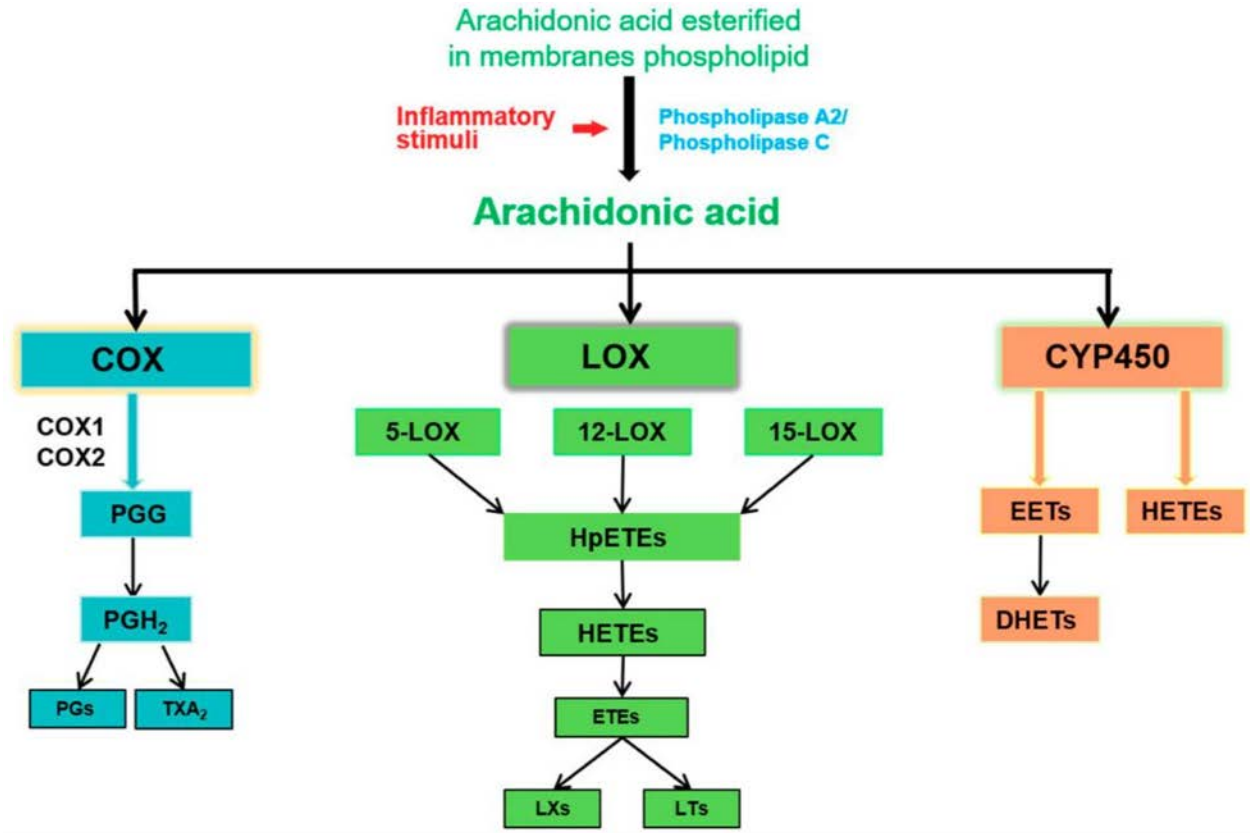
12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

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.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.

13. Per Data Use Agreement Addendum, approved manuscripts using CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes __**X**__ No.

References

1. Rand AA, Barnych B, Morisseau C, et al. Cyclooxygenase-derived proangiogenic metabolites of epoxyeicosatrienoic acids. *Proc Natl Acad Sci U S A.* 2017;114(17):4370-4375.
2. Lagerborg KA, Watrous JD, Cheng S, Jain M. High-Throughput Measure of Bioactive Lipids Using Non-targeted Mass Spectrometry. *Methods Mol Biol.* 2019;1862:17-35.
3. Gieger C, Geistlinger L, Altmaier E, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* 2008;4(11):e1000282.
4. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016;48(10):1284-1287.
5. Chanda P, Huang H, Arking DE, Bader JS. Fast association tests for genes with FAST. *PLoS One.* 2013;8(7):e68585.
6. Zerbino DR, Wilder SP, Johnson N, Juettemann T, Flicek PR. The ensembl regulatory build. *Genome Biol.* 2015;16:56.
7. Arnold M, Raffler J, Pfeufer A, Suhre K, Kastenmuller G. SNiPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics.* 2015;31(8):1334-1336.
8. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet.* 2018;19(8):491-504.
9. Wang G, Sarkar, A., Carbonetto, P., Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 2020;82:1273-1300
10. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015;348(6235):648-660.
11. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet.* 2018;50(11):1593-1599.


Arachidonic acid esterified
in membranes phospholipid
Inflammatory
stimuli
Phospholipase A2/
Phospholipase C
Arachidonic acid
COX
COX1
COX2
PGG
PGH2
PGs
TXA2
LOX
5-LOX
12-LOX
15-LOX
HpETEs
HETEs
ETEs
LXs
LTs
CYP450
EETs
HETEs
DHETs