ARIC Manuscript Proposal #4323

PC Reviewed: 9/12/23	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Gaussian graphical modeling to discover novel classes of circulating eicosanoid and related bioactive lipid markers of type 2 diabetes risk

b. Abbreviated Title (Length 26 characters): GGM for T2D markers

2. Writing Group [please provide a middle name if available; EX: Adam Lee Williams]:

Writing group members:

Arjana Begzati, abegzati@ucsd.edu
Karla P. Godinez-Macias, kpgodine@ucsd.edu
Tao Long, novutao@gmail.com
Jeramie D. Watrous, jeramie.watrous@gmail.com
Rafael Moranchel, moranchel.rafael@gmail.com
Edward D. Kantz, e4kantz@gmail.com
Jaakko Tuomilehto, jaakko.tuomilehto@helsinki.fi
Aki S. Havulinna, aki.havulinna@thl.fi
Teemu J Niiranen, Teemu.Niiranen@thl.fi
Pekka Jousilahti, pekka.jousilahti@thl.fi
Veikko Salomaa, veikko.salomaa@thl.fi
Elizabeth A. Winzeler, ewinzeler@ucsd.edu
Trey Ideker, tideker@health.ucsd.edu
Mohit Jain, mjain@health.ucsd.edu
Amit R. Majithia, amajithia@health.ucsd.edu
Bing Yu, Bing.Yu@uth.tmc.edu
Faye Norby, Faye.Norby@cshs.org
Casey Rebholz, crebhol1@jhu.edu
Elizabeth Selvin, eselvin@jhu.edu
Susan Cheng, susan.cheng@cshs.org

others welcome

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

First author: Arjana Begzati
Address: Department of Medicine, University of California San Diego
9500 Gilman Dr, MC0613
La Jolla, CA 92037
Phone: 510-499-7535
E-mail: abegzati@ucsd.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: Susan Cheng
Address: Smidt Heart Institute
127 S San Vicente Blvd #A3600
Los Angeles, CA 90048
Phone: 310-423-2726 Fax: 310-423-6795
E-mail: Susan.Cheng@cshs.org

3. Timeline: Statistical analysis: 6 months
Manuscript preparation: 1 month

4. Rationale:

Type 2 diabetes (T2D) is a major health concern not only because of the direct effects it has on the patient's quality of life, but also because it increases their risk of other health-related issues, such as heart and kidney diseases, and even diseases like Alzheimer's[1]. Although a person's blood glucose level is the most important predictor for T2D incidence, early diagnosis of T2D or people at increased risk for T2D remains challenging, and is of great importance given that multiple preventive strategies exist and are most effective when implemented early on[2]. Since T2D is preceded by and results in major disruptions to metabolism, measuring and studying the levels of metabolites other than glucose can provide important additional indicators of T2D risk. With the advances in untargeted metabolomics platforms, we can now measure thousands of chemically diverse metabolites from plasma at once. Our preliminary data shows that in our discovery cohort (FINRISK 2002) over one hundred circulating eicosanoid and related bioactive lipids are markers of incident T2D, including both previously described small molecule lipids and novel ones. Since many metabolites detected in human plasma remain yet to be characterized and assigned to chemical pathways[3,4], we will use a data driven approach on data from FINRISK 2002 to group eicosanoid and related bioactive lipids metabolites from similar chemical classes and pathways together, allowing us to not only study known metabolites but also novel ones. A clustering approach is motivated given prior data from our group and others indicating that eicosanoid and related bioactive lipid associations tend to cluster at least partly due to sharing common enzymatic pathways. Specifically, the goal of this proposal is to *discover novel classes of* eicosanoid and related *bioactive lipids that are incident T2D markers using Gaussian graphical modeling (GGM) to group both known and uncharacterized analytes by chemical class*[5].We hypothesize that (1) the incident T2D associated eicosanoid and related bioactive lipid markers measured by our LC-MS platform will be of various PUFA derived molecular and related chemical subclasses, (2) chemically related incident T2D eicosanoid markers will cluster together (in expected canonical as well as non-canonical patterns) in the GGM network, and (3) that eicosanoid and related bioactive lipid biomarkers of interest will replicate in ARIC (validation cohort).

5. Main Hypothesis/Study Questions:

Aim 1: Identify sets of chemically related, circulating incident T2D bioactive lipids and related metabolite biomakers in our discovery cohort (FINRISK 2002).

Hypothesis: Incident T2D metabolite biomarkers include a diversity of eicosanoids and related bioactive lipids.

Sub-aim 1a: Use Gaussian graphical modeling network analysis in our discovery cohort (FINRISK 2002) to group analytes of similar subclasses together and highlight sets of chemically related eicosnaoid and related bioactive lipid biomarkers of T2D risk.

Aim 2: Validate incident T2D associated eicosanoid and related bioactive lipid markers from clusters that are enriched for biomarkers in ARIC visit 2 (validation cohort).

Hypothesis: The majority of eicosanoid and related bioactive lipid biomarkers will validate within the ARIC cohort.

Sub-aim 2a: Determine if selected eicosanoid and related bioactive lipid metabolite biomarkers predict T2D risk equally well in both White and Black race groups in ARIC.

Aim 3: Perform structural elucidation of eicosanoid and related bioactive lipid biomarkers from 2-3 clusters that are enriched for incident T2D biomarkers.

Hypothesis: Identifying a couple of eicosanoid and related bioactive lipids from a cluster by follow up MS2 collection is sufficient to identify the whole class of relevant metabolites. Structural elucidation, or metabolite identication, of relevant metabolites will be done by collecting MS2 data (i.e. fragmentation information) or matching m/z and retention time to library of commercial standards.

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).

Information pertaining to validation with ARIC cohort:

Study design

Aim 2: Prospective observational study using selected plasma eicosanoid and related bioactive lipids data from ARIC visit 2 and incident T2D through 2019.

Study population

- Inclusion criteria: ARIC participants with eicosanoid and related bioactive lipid data assayed at visit 2.
- Exclusion criteria: ARIC participants with prevalent T2D at visit 2 (defined using established ARIC definitions). Those missing age, sex and BMI.

Variables

- Exposure: At visit 2, levels of eicosanoid and related bioactive lipids of interest will be extracted from the data of all the eicosanoid and related bioactive lipids already profiled using the LC-MS bioactive lipids platform. We will consider visit 2 as the baseline.
    - We will apply an appropriate transformation and batch correction to the metabolite measurements.

- Outcomes: Incident T2D until 2019. We will use ARIC established definitions of incident T2D and compare results from using the different definitions available.
- Covariates: The primary covariates for adjustment will be age, sex, and BMI, with then further adjustment made for additional traits often associated with T2D or potentially related separately to eicosanoid and related bioactive lipids variation including race/center, triglycerides, HDL, fasting glucose, A1C, eGFR, and fasting time.

Statistical analysis

- Baseline characteristics will be described for the total cohort, and then by sex and race, using mean ± SD for continuous variables and proportions for categorical variables.

Aim 1 (involving the FINRISK 2002 cohort):

- Cox proportional hazards models will be used to test for association between each measured metabolite and incident T2D, adjusting initially for age, sex, and BMI and then secondarily for additional conventional risk factors for T2D, in the FINRISK 2002 cohort.
- Gaussian graphical modeling will be used to estimate partial correlation coefficients between metabolite pairs and create a network in which nodes represent metabolites and edges represent statistically significant partial correlations. To break down the network into modular units that only contain tightly connected metabolites, the Louvain method of community detection will be applied.

Aim 2 (involving ARIC):

- Primary analysis: Cox proportional hazards models will be used to validate the association between rank selected eicosanoid and related bioactive lipids and incident T2D, adjusting initially for age, sex, and BMI and then secondarily for additional conventional risk factors for T2D. We will also consider discrete-time methods to account for the relatively non-continuous event timing of T2D event capture.
- Secondary analysis: We will report the association by race group given prior literature indicating there may be racial/ethnic/ancestral variation in bioactive lipids and their associations (*Glob Heart 2017;12:141-150, ESC Heart Fail 2020;7:1700-10*).

The proportional hazards assumption will be examined using the Schoenfeld residuals. A conservative Bonferroni corrected P-value will be used to determine significant associations. We will also secondarily consider FDR and other less conservative methods.

7.a. Will the data be used for non-ARIC analysis or by a for-profit organization in this manuscript? ____ Yes _X_ No

b. If Yes, is the author aware that the current derived consent file ICTDER05 must be used to exclude persons with a value RES_OTH and/or RES_DNA = "ARIC only" and/or "Not for Profit" ? ____ Yes ____ No
(The file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? ____ Yes _X_ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the current derived consent file ICTDER05 must be used to exclude those with value RES_DNA = "No use/storage DNA"? ____ Yes ____ No

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

_____ 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)?

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? ____ Yes _X_ No

11.b. If yes, is the proposal

___ A. primarily the result of an ancillary study (list number* _____)

___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____ _____)

*ancillary studies are listed by number https://sites.cscc.unc.edu/aric/approved-ancillary-studies

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

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PubMed Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

References:

1. Reed J, Bain S, Kanamarlapudi V. A Review of Current Trends with Type 2 Diabetes Epidemiology, Aetiology, Pathogenesis, Treatments and Future Perspectives. *Diabetes Metab. Syndr. Obes.* 2021;14:3567–3602. doi: 10.2147/DMSO.S319895.

2. Tuomilehto J, Uusitupa M, Gregg EW, Lindström J. Type 2 Diabetes Prevention Programs—From Proof-of-Concept Trials to National Intervention and Beyond. *J. Clin. Med. Res.* 2023:12:1876. doi: 10.3390/jcm12051876.
3. Da Silva RR, Dorrestein PC, Quinn RA. Illuminating the Dark Matter in Metabolomics. *PNAS*. 2015;112:12549–12550. doi: 10.1073/pnas.1516878112.
4. Nagana Gowda GA, Gowda YN, Raftery D. Expanding the Limits of Human Blood Metabolite Quantitation Using NMR Spectroscopy. *Anal. Chem.* 2014;87:706–715. doi: 10.1021/ac503651e.
5. Krumsiek J, Suhre K, Illig T, Adamski J, Theis FJ. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst. Biol.* 2011;5:1–16. doi: 10.1186/1752-0509-5-21.