ARIC Manuscript Proposal #3473

PC Reviewed: 9/10/19  Status: _____  Priority: 2
SC Reviewed: _________  Status: _____  Priority: _____

1.a. Full Title:
Serum Proteome-wide Association Study for gout and hyperuricemia

b. Abbreviated Title (Length 26 characters):
Serum proteomics for gout

2. Writing Group:
Writing group members: Jingning Zhang, Joe Coresh, Adrienne Tin, Anna Kottgen, Bing Yu, Eric Boerwinkle, Nilanjan Chatterjee, and others welcome.

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

First author: Jingning Zhang
Address: Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe Street, E3527
Baltimore, MD 21205

Phone: 443-739-3979
E-mail: jzhan218@jhu.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).
Name: Nilanjan Chatterjee
Address: Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe Street, E3527
Baltimore, MD 21205

Phone: 410-955-3067
E-mail: nchatte2@jhu.edu

3. Timeline:
We will begin the analysis using the proteomic data for gout from visit 3 and focus on identifying proteins associated with the hyperuricemia and gout. The manuscript is expected to be completed in 2 years.

4. Rationale:

Understanding causal relationship between proteins and complex traits are critical for understanding biology and developing potential drug targets for intervention. Recent availability of high-throughput protein measurements together with genome-wide genotyping data provides the opportunity to test for causal relationship between proteins and specific traits in an agnostic and objective fashion. The availability of SomaLogic data on large number of individuals from the ARIC study, combined with GWAS data, provides unique opportunity to build models for predicting proteins based on cis-SNPs (in the 500kb on either side of the gene boundary) and then take these models to large GWAS for testing association between genetically predicted proteins and specific traits. Such approach, which relies on an underlying Mendelian Randomization analysis framework, has been widely popular for testing causal relationship between gene-expression and complex trait under the framework of transcriptome-wide association studies (TWAS).

Gout is the most common form of inflammatory arthritis and affects ~8 million U.S. adults. Gout attacks result from an inflammatory response to the deposition of monosodium urate crystals in particular tissues, including the synovial lining, causing rapidly intensifying painful and swollen joint(s). Thus, high serum urate levels (> 6 mg/dL) have long been recognized as a major risk factor for gout. While genome-wide association studies have discovered a large number of urate-associated loci, few causal genes have been identified and the regulation of serum urate levels is not fully understood. The association between genetically-predicted protein levels and serum urate may shed light on serum urate regulation and potential causal gene(s) in urate-associated loci. In addition, the association between genetically-predicted protein levels and gout may inform gout susceptibility beyond serum urate.

5. Main Hypothesis/Study Questions:

Identifying proteins that may be causally related to gout and hyperuricemia using proteome-wide association analysis.

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).

We will use SomaLogic proteomic and GWAS data to build models for predicting individual proteins based on SNPs in cis-region (in the 500kb on either side of the gene boundary) of pertinent genes. We use the cis-region because our proposed method utilizes the idea of two-sample Mendelian Randomization approach to identify causal relations between a specific protein and the traits, here are gout and hyperuricemia. Mendelian Randomization has the assumption that the SNPs we use should be independent of the traits conditional on this protein
level. Therefore, only cis SNPs can be used because trans SNPs may influence the traits from biological pathways other than this protein, which violates the independence assumption. In our study, the Mendelian Randomization is performed through predicting the protein level using multi-SNP linear models and then regressing traits on the predicted protein level to get the estimated effect size from this protein to the traits.

We will use elastic net method, an approach has been shown to be powerful for predicting gene-expression using SNPs data, for building multi-SNP models for predicting individual proteins. We adjust the protein level with covariates for sex, age, genetic ancestry, and multiple gene expression PCs. The race factor can thus be adjusted in our models. Further to improve power of protein prediction, we will incorporate information from prior models that have been built for predicting expression level for relevant genes using data from the large GTEx project.

We will then use models to perform proteome-wide association studies (PWAS) for hyperuricemia and gout using largest available GWAS dataset for these traits. Essentially the models for predicting proteins will be applied to the GWAS datasets to evaluate the association between genetically predicted proteins and the traits. The association tests could be performed even when only summary-level association statistics are available from external GWAS. We will identify proteins which is found to be significant after conservative Bonferroni correction. For identified proteins using this PWAS approach, we will evaluate association between actual level of measured protein and the outcomes in the smaller ARIC data.

Inclusion criteria:
ARIC participants with:
  1) SomaLogic proteomic data at visit 3 or visit 5
  2) Genotype imputed using Affy 6.0. Only SNPs within the cis region of the encoding proteins will be used.

Other variables of interest: serum urate at visits 1, 2, 4, 5, self-reported gout status collected at visits 4 and 5.

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

   b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = “CVD Research” for non-DNA analysis, and for DNA analysis RES_DNA = “CVD Research” would be used? _X__ Yes    ____ No
   (This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? __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”? 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)?
ARIC Manuscript Proposal #3388, Proteomic Profiling of Gout Risk in ARIC.

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

11.b. If yes, is the proposal
A. primarily the result of an ancillary study (list number* __2018.13__)
B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* __________ ____________)

*ancillary studies are listed by number at http://www.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 No.

References


