ARIC Manuscript Proposal #2347

PC Reviewed: 4/8/14  Status: A  Priority: 2
SC Reviewed: _________  Status: _____  Priority: ____

1.a. Full Title: Gene-specific function prediction for non-synonymous mutations in monogenic diabetes genes

b. Abbreviated Title (Length 26 characters): MODY SNP prediction

2. Writing Group:
   Writing group members: Quan Li, Xiaoming Liu, Richard Gibbs, Eric Boerwinkle, Constantin Polychronakos, Hui-Qi Qu

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

First author: Quan Li
   Address: Endocrine Genetics Lab, The McGill University Health Center (Montreal Children's Hospital), Montréal, Québec, Canada
   Phone: 5149341934x23204   Fax: E-mail: quan.li@mail.mcgill.ca

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: Xiaoming Liu
   Address: 1200 Herman Pressler ST, E529, Houston, TX 77030
   Phone: 7135009820   Fax: E-mail: xiaoming.liu@uth.tmc.edu

3. Timeline: Completion of the manuscript is anticipated in April 2014.

4. Rationale:
   Functional prediction methods for non-synonymous SNPs are commonly used in sequencing-based Mendelian disease studies. However, as different methods are
optimized based on their own training data set, their prediction accuracy for different genes can be very different. We propose a comparison of those methods specifically for non-synonymous SNPs of MODY (Mature Onset Diabetes of the Young) genes, in order to help classification of non-synonymous SNPs of unknown clinical significance in the future. We propose to compare the predictions of those methods using known non-synonymous SNPs that causing MODY and those observed in cohort populations and unlikely causing MODY. We further propose to adjust thresholds for those prediction methods for gene-specific optimal prediction.

Non-synonymous SNPs identified by the Exome sequencing of the ARIC cohort will be used as controls to compare with non-synonymous SNPs causing MODY. This set of SNPs will be combined with other non-synonymous SNPs in MODY genes that were reported by the 1000 genomes project and the ESP project to form a neutral SNP set.

5. Main Hypothesis/Study Questions:
The main study question is to compare functional prediction methods and identify the methods that have the best prediction accuracy for MODY genes.

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).
MODY causing non-synonymous SNPs were collected from HGMD database. Non-MODY causing non-synonymous SNPs were collected from large exome sequencing projects including the 1000 genomes project, the ESP project and the exome sequence data of ARIC samples from the CHARGE-S project. No individual-level phenotype data or genotype data were used in the study. Exome sequence data of ARIC samples from the freeze 3 CHARGE-S data after QC were used to collect non-synonymous SNPs in MODY genes and then combined with other such SNPs from the 1000 genomes project and the ESP project.

Prediction performance was compared using ROC curves and AUC. Optimal thresholds for prediction were determined based on the ROC curves.

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”?
   _X_ Yes    ____ No

10. 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)?
    No related manuscript proposal found.

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* __________)  
      ___ B. primarily based on ARIC data with ancillary data playing a minor
            role (usually control variables; list number(s)* (2009.12)

*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.

     Agree.

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
     shows you which journals automatically upload articles to Pubmed central.

     Agree