ARIC Manuscript Proposal #2252

PC Reviewed: 11/12/13  Status: A  Priority: 2
SC Reviewed: _________  Status: _____  Priority: ____

1.a. Full Title: Identification of Exonic Variants for Blood Electrolyte Balance

b. Abbreviated Title (Length 26 characters): Exome Electrolyte

2. Writing Group:
Writing group members: Adrienne Tin, Yong Li, Elizabeth Selvin, Mandy Li, Nathan Bihlmeyer, Pam Lutsey, Eric Boerwinkle, Dan Arking, Joe Coresh, Linda Kao, Anna Kottgen; other ARIC authors are invited. Different members of the writing group may participate in the different analyses outlined below (such as exome chip vs. exome sequencing).

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

First author/analyst (publication will likely occur in the setting of large consortia):
Adrienne Tin, Ph.D.
Address: Johns Hopkins Bloomberg School of Public Health
615 N Wolfe St, Room W6521
Baltimore, MD 21205
Phone: 201-281-9577
E-mail: atin@jhsph.edu

Corresponding/senior author (must be an ARIC investigator for the proposal but can be different in the published paper; correspondence will be sent to both the first author & the corresponding author):

Senior author:
Anna Kottgen, MD MPH
Address: Dept. of Internal Medicine IV
Freiburg University Medical Center
Berliner Allee 29
79110 Freiburg
Germany

Phone: +49 761 270-78050  Fax: +49 761 270-78040
E-mail: anna.koettgen@uniklinik-freiburg.de, akottgen@jhsph.edu

3. Timeline:
Data analysis can begin immediately after approval of the manuscript proposal.
4. **Rationale:**

The blood electrolytes included in this proposal are: phosphate, magnesium, calcium, potassium, and sodium. The balances of these electrolytes are highly regulated. Based on a study of female twins in British, the heritability estimates are 33% for serum calcium, 58% for serum phosphate, and 27% for serum magnesium.\(^1\) The heritability of serum sodium is found to be higher in female (0.44 to 0.67) than in males (0.18 to 0.24).\(^2\) Genome-wide association studies have identified common variants associated with serum calcium,\(^3,4\) phosphate,\(^5\) magnesium and potassium.\(^6\) These associated variants explain a small proportion of the estimated heritability and are mostly in non-exomic region. The biological mechanisms underlying the association of the common variants remain to be investigated.

A potential role of rare coding and splice variants in electrolyte homeostasis is supported by the literature. For examples, rare exonic variants in genes involved in blood potassium and sodium imbalance have been implicated in hypertension,\(^7\) and a loss of function mutation in TRPV4 is associated with hyponatremia.\(^8\) The availability of the exome chip and exome sequence data are opportunities for identifying exonic or splice variants associated with electrolyte balance.

5. **Main Hypothesis/Study Questions:**

The main objectives of the proposed analyses:

1. Identify exonic or splice variants that are associated with measures of serum sodium, magnesium, calcium, potassium, and phosphate.

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 for genotype data collection:**

1. Exome chip data are from a cohort design including all individuals in the ARIC study with self-reported white or black ethnicity (N=10895 for whites and 3853 for blacks)

2. Exome sequence data are available from a case-cohort design (CHARGE-S, N~3173).

We will exclude individuals without genetic consent and those who do not have genotype or phenotype data. Analyses for each electrolyte may have specific exclusion criteria. For example, it may be appropriate to exclude individuals with low estimate glomerular filtration rate (eGFR) for serum Mg or phosphate because low eGFR leads to imbalance in the levels of these electrolytes.

**Outcome variables:** serum sodium, magnesium, calcium, potassium, and phosphate at ARIC visit 1.

**Exposure variables:**
1. Genotype data from exome chip
2. Genotype data from exome sequencing

Covariates: Age, sex, field center, principle components if necessary. The analyses for each electrolyte may include specific covariates. For example eGFR and diuretics use may be relevant as covariates for serum Mg.

**Analysis approach for both exome chip and sequence data.** We will perform single SNP and gene-based analyses stratified by race:

- **Single SNP analyses:** we will perform linear regression for each outcome on additive coding of SNP genotype. Additional exact p-value or permutation p-value will be provided for rare variants (MAF<5%). Number of replicates in the permutation should be >1/pval from regression.
- **Gene-based analyses**
  1. Burden test. This test is more powerful when the effect directions of all variants in a gene region are consistent.
  2. Sequence kernel-based association test (SKAT). This is more powerful than the burden test when a region contains protective, deleterious and null variants.

**Exome chip specific analysis.** Variants will be recoded such that the CHARGE-wide minor allele will be used as the coded allele. All analyses will be performed using the seqmeta (formerly known as skatMeta) package developed by the CHARGE Analysis Committee. The prepScores (formerly known as skatCohort) function will be used for generating score statistics from the data of a single cohort. The burdenMeta, skatMeta, and singleMeta methods will be used for generating association test results for the burden, SKAT, and single variant test for a single cohort and for meta-analysis of all cohorts. All variants, including monomorphic ones, will be included in the analysis. This way, SNP selection criteria for burden test and SKAT test, including MAF threshold and SNP function, can later be decided at the meta-analysis stage. For exploratory analyses within ARIC only, we will test different thresholds such as MAF of 0.01 and 0.05. For processing of the data and the format of the results, we will adhere to established CHARGE standards as detailed in the skatMeta_Cookbook.pdf developed by the CHARGE Analysis Committee.

For CHARGE-S data, we will apply the appropriate sampling weight in our analysis.

**General approach in following up a SNP of interest.** We will apply the following criteria to narrow down our list of SNPs for follow up: 1) non-synonymous coding SNPs; 2) SNPs with high statistical significance; 3) novel SNPs (not in any public databases); 4) SNPs associated with a splice site; 5) SNPs predicted to be functional; 6) SNPs that are in LD with the initial GWAS index SNP and have predicted functional significance; 7) SNPs in regulatory regions of the gene.

**Limitations/Challenges:**
7.a. Will the data be used for non-CVD analysis in this manuscript?  ____ 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?  ____ Yes  ____ No

(This file ICTDER03 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  ____ 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  ____ 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)?
None yet, the thyroid function measures have just become available and there are no manuscript proposals in this area yet.

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*) 2009.24, 2006.03, 2007.02
  ____ 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/

12. 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
References