1. (a) **Full Title:** Genome wide association study of copy number with kidney disease-related traits

   (b) **Abbreviated Title:** Copy number GWAS

2. **Writing group.** Members: Robert B. Scharpf, Ingo Ruczinski, Josef Coresh, Anna Kottgen, Eric Boerwinkle, Linda Kao, Mandy Li, Rafael Irizarry

   I, the first author, confirm that all co-authors have given their approval for this manuscript proposal. RBS

3. **First author:** Robert B. Scharpf
   
   615 N. Wolfe St, Rm E3537
   
   Baltimore, MD 21205
   
   Phone: (410) 614-7837
   
   E-mail: rscharpf@jhsph.edu
   
   Fax: (410) 955-0958

   **ARIC author** to be contacted if there are questions about the manuscript and the first author does not respond or can not be located (this must be an ARIC investigator).

   Name: Josef Coresh
   
   Phone: (410) 955-0495
   
   E-mail: coresh@jhu.edu
   
   Fax: (410) 955-0476

4. **Timeline:** Data analysis to start immediately (August, 2008). First draft of manuscript expected September, 2008.

5. **Rationale:**

   Variation of chromosomal DNA between individuals occurs at the level of entire chromosomes, segmental changes spanning multiple loci, and small genomic regions, including single nucleotide polymorphisms (SNPs). Susceptibility to common diseases such as cancer and diabetes is influenced by genetic variation (e.g., [Feuk et al.](#) 2006; [Estivill and Armengol](#) 2007).

   Two forms of genetic variation are particularly common and can be assessed on a genomic scale using recent the Affymetrix 6.0 genotyping platform: single nucleotide polymorphisms (SNPs) and copy number variants (CNVs). Most existing GWAs ignore the information in genomic copy number as well as the roughly 1 million nonpolymorphic probes on the 6.0 array. The goal of this manuscript proposal is to assess the contribution of genetic variation from CNVs to traits relevant to kidney disease, including uric acid levels and electrolytes. Statistical methods developed as a result of this research will be publicly available as part of the Bioconductor project ([Gentleman et al.](#) 2004).

6. **Main Hypothesis/Study Questions:**
We propose to study the association of CNV and traits relevant to kidney disease among the white ARIC participants. Traits of interest include uric acid levels, kidney disease, and electrolyte measurements. We hypothesize that metabolic traits such as uric acid levels may be influenced by CNV.

7. 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).

The goal of this manuscript proposal is to identify CNVs among ARIC participants, quantify the uncertainty of inferred CNV regions, and assess the association of CNV with phenotypes relevant to kidney disease. While traits such as uric acid levels are known to have high heritability (0.63 after adjustment [Yang et al., 2005]) and several polymorphisms with strong genetic effects, the relative contribution of CNV to such traits is unknown.

Our analysis will proceed in several stages that we summarize briefly. First, we will preprocess the Affymetrix 6.0 CEL files using a robust-to-outlier procedure, SNP-RMA [Carvalho et al., 2007], that summarizes the raw fluorescence intensities to the level of the allele. We will obtain locus-specific estimates of copy number by modeling the preprocessed intensities as a function of optical background, non-specific binding, and specific-binding [Wu and Irizarry, 2007]. We will improve the locus-specific estimates of copy number by using a hidden Markov model (HMM) to infer CNVs spanning multiple loci [Scharpf et al., 2008]. Association of CNVs and phenotypes will be assessed using t-tests, ANOVAs, and rank-based procedures for continuous outcomes and chi-square tests for binary outcomes. We will develop alternative procedures for assessing association and quantifying copy number as needed. For instance, we will explore methods that borrow-strength across samples, such as a principal components-based approach, to identify regions of CNV shared by a subset of the participants.

Study design: All ARIC participants with available GWAS data. CEL files are available on all ARIC participants in Phase 1 and 2. This subgroup is being used in GWAs analysis by other papers. Phase 3 data will be integrated if available. Phase 1 and 2 includes approximately 8,800 white participants who will be the primary focus of this paper. Data on African-Americans was funded by CARE and its use will follow CARE procedures. The primary analysis will focus on the whites and the analysis of African-Americans will be guided by CARE procedures.

Primary outcomes: We will coordinate analysis with several phenotype working groups. Writing group members Drs. Coresh, Kao and Kottgen lead the ARIC GWAS Kidney disease working group and will ensure this analysis is coordinated with the other workgroup analyses.

8. (a) Will the data be used for non-CVD analysis in this manuscript? __ Yes \textbf{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? __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.)

9. (a) Will the DNA data be used in this manuscript? \textbf{X} Yes __ No
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

If yes, is the author aware that some DNA data is not allowed to be used by ‘for profit’ groups. Is this data being used by a ‘for profit’ organization? _X_ No If yes, is the author aware that the participants with RES DNA = ‘not for profit’ restriction must be excluded? _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

11. 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)?

#1343 Stage II of Genome-wide Association Study for Genetic Variants Associated with Uric Acid Levels and Gout

12. (a) Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? _Yes _ X No

(b) If yes, is the proposal underline A. primarily the result of an ancillary study (list number 1); _ _ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s))*

13. 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.

References


1ancillary studies are listed by number at [http://www.cscc.unc.edu/aric/forms/]
