

ARIC Manuscript Proposal # 1263

PC Reviewed: 6/ 5 /07

Status: A

Priority: 2

SC Reviewed:

Status:

Priority:

1.a. Full Title: Stage II of Genome-Wide Association Study for the Identification of Genetic Variants Associated with Renal Traits

b. Abbreviated Title (Length 26 characters): GWAS replication kidney

2. Writing Group:

Writing group members: Anna Kottgen, Linda Kao, Caroline Fox, Shih-Jen Hwang, Brad Astor, Eric Boerwinkle, Josef Coresh

Additional authors will be added to represent the Framingham collaboration.

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

First author: **Anna Kottgen**

Address: 2024 E. Monument St.

Welch Center for Prevention, Epidemiology, and Clinical Research

Baltimore, MD 21287

Phone: (410) 245-6897

Fax: (410) 955-4076

E-mail: akottgen@jhsp.edu

Corresponding/senior author (if different from first author correspondence will be sent to both the first author & the corresponding author): **Josef Coresh**

Address: 2024 E. Monument St.

Welch Center for Prevention, Epidemiology, and Clinical Research

Baltimore, MD 21287

Phone: (410) 955-0495

Fax: (410) 955-4076

E-mail: coresh@jhu.edu

3. Timeline: Study design and genotyping to start immediately, analyses completed in the next 3 months, additional genotyping and manuscript preparation over the next 6 month.

4. Rationale:

Chronic kidney disease (CKD) has been recognized as a public health problem which affects an estimated 19 million adults in the US (1). Progression of CKD may lead to end-stage renal disease (ESRD). The yearly mortality rates for individuals treated with dialysis for ESRD exceed 20% (2). Therefore, early identification of individuals at increased risk for CKD and effective intervention is essential. Previous studies have shown that sub-groups of individuals susceptible to kidney disease exist (3). Additionally, multiple studies have confirmed that kidney disease is heritable (4). Apart from major cardiovascular risk factors such as hypertension, genetic causes contribute directly to the complex disease CKD.

Association studies in large study populations provide greater power than family-based linkage studies for the identification of common genetic variants underlying complex diseases (5). Recently, genome-wide association studies (GWAS) to discover associations of common single nucleotide polymorphisms (SNPs) and a phenotype of interest have become feasible. Screens of 100,000-500,000 common SNPs across the genome have led to the discovery of variants significantly associated with age-related macular degeneration (6), type 2 diabetes (7-10), obesity (11, 12), inflammatory bowel disease (13), and heart disease (14-16). To date, there have not been any reports on genome-wide association studies with kidney disease as the phenotype.

In 2006, the NHLBI genotyped 116,204 markers on 1,320 related and unrelated Framingham Heart Study (FHS) participants not selected for a specific phenotype using the Affymetrix GeneChip Human Mapping 100K set. Kidney disease as the phenotype of interest was analyzed using 3 different traits: 1) continuous eGFR in ml/min/1.73m² (MDRD Study equation(17)), 2) CKD defined as eGFR<59 ml/min/1.73m² (women) or <64 ml/min/1.73m² (men), 3) continuous serum cystatin C levels (mg/l). Analyses were performed using both FBAT (18) and GEE models to account for relatedness among study persons. Lists containing the SNPs with the top 100 lowest p-values for association with each of the 3 traits as well as those showing p-values <0.01 with at least two of the traits were passed on to ARIC investigators for the selection of SNPs to be replicated in the ARIC cohort. A signed memorandum of understanding with the Framingham investigators exists.

We propose to replicate the association of these SNPs with traits related to kidney disease in the ARIC Study.

5. Main Hypothesis/Study Questions:

Main hypothesis: SNPs that were significantly associated with kidney traits in the ARIC Study will also be significantly associated with kidney traits in Caucasian ARIC participants.

Study questions:

1. Will the association between these SNPs and kidney traits observed in the FHS replicate in Caucasian ARIC participants?

2. Will this association also be present in African American ARIC participants? Furthermore, can differences in frequencies of such genetic variants between Caucasians and African Americans account for part of the differences in disease risk?
3. Will the association be observed cross-sectionally as well as prospectively (incident CKD), and with different kidney traits?
4. How will such associations be influenced by adjustment for or stratification on important CKD risk factors such as age, hypertension, or type 2 diabetes?
5. Will the observed associations be for variants lying in genes, and if so, in those with a known pathophysiological role in kidney disease? Can new hypotheses for kidney disease mechanisms be generated?

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:

SNPs will be selected for replication based on statistical, biological, and cost considerations. Initially, we selected the 8 SNPs (rs10509132, rs6831700, rs1613631, rs2228210, rs2419912, rs2827732, rs4553158, rs6495446) significantly associated with each of the 3 kidney traits studied in FHS at $p < 0.01$ for replication in the entire ARIC cohort. Genotyping will be carried out using the Sequenom iPLEX platform at the ARIC DNA lab in Houston (E. Boerwinkle). SNPs failing on this platform will be typed using TaqMan assay. In addition, SNPs in kidney disease candidate genes that were on any of the lists received from FHS will be typed (rs2839235, rs10520688, rs1455177, rs3779748, rs1743955, rs2061063, rs4835136, rs4148686). Further SNPs may be added to this list in coordination with the ARIC DNA lab and following Dr. Boerwinkle's direction as chair of the laboratory committee.

Inclusions/Exclusion: Individuals who did not consent to DNA research as well as those not of Caucasian or African American race will be excluded. Quality control will be performed and genotypes not meeting the standards (see below) will be excluded.

Outcome: Kidney disease traits will be evaluated cross-sectionally as well as prospectively. For cross-sectional analyses, the association of genotypes with eGFR, $\ln(\text{eGFR})$, albumin-to-creatinine ratio (ACR), and $\text{eGFR} < 60 \text{ ml/min/1.73m}^2$ will be studied at ARIC study visit 1 (eGFR) and 4 (eGFR, ACR). For prospective analyses, incident CKD will be defined as a rise in serum creatinine $> 0.4 \text{ mg/dl}$ above baseline or kidney disease hospitalization or death, or as incident $\text{eGFR} < 60 \text{ ml/min/1.73m}^2$ or kidney disease hospitalization or death.

Other variables of interest: age, gender, systolic and diastolic blood pressure, antihypertensive medication, diabetes mellitus, smoking, triglycerides, HDL cholesterol, education.

Data analysis:

Quality control: SNPs exhibiting a call rate is <90%, a p-value for the test of Hardy-Weinberg equilibrium of $<10^{-4}$, and a minor allele frequency of <5% will be dismissed from analyses.

In all analyses, an additive genetic model will be assumed. All analyses will be carried out stratified by race.

The **primary analysis** to be conducted will be **cross-sectional** and consistent with the analysis conducted in the FHS to avoid having to adjust for multiple testing. Specifically, multivariable regression models adjusted for age, sex, systolic blood pressure, hypertension treatment, HDL-cholesterol, smoking, diabetes, and body mass index will be run. Consistent with the FHS analyses, we will run this model in i) a linear regression analysis with continuous eGFR as the outcome, and ii) a logistic regression model with eGFR <60 ml/min/1.73m² as the outcome. Additional **secondary analyses** will be conducted to investigate the association of genotypes with the albumin-to-creatinine ratio (ACR) in a multivariable linear regression model. For all continuous kidney traits, means will be calculated and compared for the three genotypes at each SNP using ANOVA.

In additional secondary analyses, the association of the genetic variants and incident CKD will be studied. For these **prospective analyses**, follow-up time will be counted from ARIC visit 1 until the visit date at which the creatinine rise / eGFR decline occurred, or the date of CKD hospitalization discharge or death, or the earlier of the two dates for participants meeting both definitions. Non-cases will be censored at the earlier of the date of last contact (or date of non-CKD death) or December 31, 2002. Incidence rates of CKD will be calculated using person-time methods. Kaplan-Meier estimates of mortality will be computed, and log-rank tests will be used to compare survival curves among the genotypes. Each SNP will be tested for association with incident CKD in multivariable Cox proportional hazard analyses, adjusted for the same covariates as the primary analysis. We will estimate the effect size (RH of incident CKD) and 95% confidence interval for each SNP.

The distribution of baseline characteristics in the study population by genotype for significantly associated SNPs as well as by outcome will be computed using t-tests, chi-square tests and ANOVA as applicable.

Determination of statistical significance: For each of the SNPs in a plausible CKD candidate gene, results will be considered significant at a p-value of <0.05 (19). For the other 8 SNPs, a Bonferroni correction will be applied, resulting in a significance level of $\alpha = 0.00625$ (two-sided).

Power: The power in ARIC is high for moderate associations. A simplified analysis for incident CKD of cases vs. non-cases (ignoring follow-up time) shows that assuming $\alpha = 0.05$, an OR of 1.3 (or 1.6 if larger genetic effects are hypothesized) and allele frequencies of 0.2, 0.1, 0.05 and 0.01 results in calculated power of 98.5% (100%), 88.3% (100%), 64.4% (98.9%), and 19.0% (53.7%).

Limitations: We will only be able to investigate the most significantly associated SNPs as well as those in the most interesting candidate genes. Undetected population substructures might possibly impact our findings. However, the impact of population stratification on false-positive findings has been reported as low (20). In addition, we will be getting data from ancestry informative makers from CIDR shortly. This information can then be incorporated into our analyses.

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 ICTDER02 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 ICTDER02 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 ICTDER02 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.csc.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)?

MP 1203: Association between candidate genetic variants and incident chronic kidney disease: The Atherosclerosis Risk in Communities (ARIC) Study

MP 1237: Association between genetic variants conferring risk for type 2 diabetes mellitus and incident chronic kidney disease

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* genotyping will be funded through ancillary study #2006.16, albuminuria data from ARIC ancillary study #2002.02)**

_____ **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.csc.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.

References

1. Coresh J, Astor BC, Greene T, Eknoyan G, Levey AS. Prevalence of chronic kidney disease and decreased kidney function in the adult US population: Third National Health and Nutrition Examination Survey. *Am J Kidney Dis.* 2003;41(1):1-12.
2. National Kidney Foundation. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Am J Kidney Dis.* 2002;39(2 Suppl 1):S1-266.
3. Klag MJ, Whelton PK, Randall BL, Neaton JD, Brancati FL, Ford CE, et al. Blood pressure and end-stage renal disease in men. *N Engl J Med.* 1996;334(1):13-8.
4. Freedman BI, Satko SG. Genes and renal disease. *Curr Opin Nephrol Hypertens.* 2000;9(3):273-7.
5. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science.* 1996;273(5281):1516-7.
6. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. *Science.* 2005;308(5720):385-9.
7. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature.* 2007;445(7130):881-5.
8. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, et al. Replication of Genome-Wide Association Signals in U.K. Samples Reveals Risk Loci for Type 2 Diabetes. *Science.* 2007;.
9. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants. *Science.* 2007;.
10. Saxena R, Voight BF, Lyssenko V, Burt NP, de Bakker PI, Chen H, et al. Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels. *Science.* 2007;.
11. Herbert A, Gerry NP, McQueen MB, Heid IM, Pfeuffer A, Illig T, et al. A common genetic variant is associated with adult and childhood obesity. *Science.* 2006;312(5771):279-83.
12. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science.* 2007;316(5826):889-94.

13. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science*. 2006;314(5804):1461-3.
14. Arking DE, Pfeufer A, Post W, Kao WH, Newton-Cheh C, Ikeda M, et al. A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. *Nat Genet*. 2006;38(6):644-51.
15. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, et al. A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction. *Science*. 2007;.
16. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, et al. A Common Allele on Chromosome 9 Associated with Coronary Heart Disease. *Science*. 2007;.
17. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med*. 1999;130(6):461-70.
18. Horvath S, Xu X, Laird NM. The family based association test method: strategies for studying general genotype--phenotype associations. *Eur J Hum Genet*. 2001;9(4):301-6.
19. Curtis D, Vine AE, Knight J. A pragmatic suggestion for dealing with results for candidate genes obtained from genome wide association studies. *BMC Genet*. 2007;8:20.
20. Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet*. 2006;7(10):781-91.