INTEGRATING MULTIPLE OMICS APPROACHES AND HUMAN BIOLOGY FOR UNDERSTANDING KIDNEY FUNCTION

By Zhi Yu

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

July 2020

Abstract

Chronic kidney disease (CKD) is a global public health problem that affects greater than 10% of adults and is associates with multiple complications. Prevention of CKD incidence and early progression informed by an accurate understanding of the risk factors and causes is important since later risks are high and hard to treat. Previous studies have examined several potential risk factors of CKD; however, few studies have evaluated longitudinal change in kidney function as an outcome. Furthermore, research addressing the underlying causal relationships and risk prediction has been limited. Advances in genomics and proteomics provide new opportunities for understanding CKD risk factors. We therefore investigated kidney function at three layers: association, causation, and prediction, integrating multiple omics approaches into our epidemiological studies. Chapter 1 provides an introduction. In Chapter 2 and 3, we rigorous examined risk factors for longitudinal kidney function decline using classical epidemiological approaches. We modeled 30-year kidney function trajectories and demonstrated the associations of hypertension and obesity with random effects models for kidney function decline over 30-year of follow-up. In Chapter 4, we used genome-wide association study (GWAS) results and Mendelian randomization methods to examine the causal directions between risk factors and kidney function. Not only did we demonstrate strong causal effects of lower kidney function on higher blood pressure, we devised a method using multiple markers to triangulate on the subset of genes that are likely to reflect kidney function susceptibility. In Chapter 5, we further examined the power of genetic susceptibility for kidney function to predict future risk and investigated the association between the genetic risk with an intermediate phenotype, plasma proteins. We demonstrated the link between genetic basis of kidney function measured as polygenic risk score (PRS) with incidence of CKD, end-stage kidney disease, kidney failure, and

acute kidney injury, supporting the use of genetic information as risk factors for kidney diseases. We also found that protein associations were stronger with kidney function than its genetic risk. Overall, using multiple types of data and methods, this doctoral thesis advances our understanding of multiple non-genetic and genetic risk factors for CKD and its progression.

Advisors: Josef Coresh, MD, PhD; Nilanjan Chatterjee, PhD; Adrienne Tin, PhD

Readers: Morgan E. Grams, MD, PhD; Elizabeth Selvin, PhD, MPH

Alternates: Scott L. Zeger, PhD; Jay H. Bream, PhD; Casey M. Rebholz, PhD, MS, MNSP, MPH

Table of Contents

Abstract ii

List of Tables vi

List of Figures vii

Chapter 1 Introduction 1

Chapter 2 Association between Hypertension and Kidney Function Decline: the Atherosclerosis Risk in Communities Study 6

Abstract 7

Introduction 9

Methods 10

Results 14

Discussion 18

Chapter 3 Association between Midlife Obesity and Kidney Function Trajectories: the Atherosclerosis Risk in Communities Study 37

Abstract 39

Introduction 41

Methods 42

Results 46

Disscussion 49

Chapter 4 Kidney Function and Blood Pressure: A Bidirectional Mendelian Randomization Study 77

Abstract 78

Introduction 79

Results 80

Discussion 85

Methods 88

Chapter 5 Polygenic Risk Scores of Kidney Function, Circulating Proteome, and Incident Kidney Diseases: the Atherosclerosis Risk in Communities Study 104

Abstract 106

Introduction 108

Methods 109

Results 117

Discussion 120

Chapter 6 Conclusions 166

References 171

Curriculum Vitae 196

List of Tables

Table 2-1 Baseline characteristics of the study population according to baseline hypertension status. 22

Table 2-2 Differences in estimated glomerular filtration rate decline per year according to hypertension categories in the total population and by race. 30

Table 2-3 Average predicted probability of developing chronic kidney disease over 30 years follow-up according to hypertension status among participants based on the baseline covariates. 32

Table 3-1 Characteristics of the study population according to baseline body mass index tertiles by sex and race. 60

Table 3-2 Association of estimated glomerular filtration rate decline slope with three measures of baseline obesity by sex and race. 62

Table 3-3 Hazard ratios for end-stage kidney disease according to baseline obesity status by sex and race. 64

Table 4-1 Basic characteristics of the studies that contributed summary statistics of kidney function and blood pressure. 94

Table 5-1 Characteristics of the study population. 126

Table 5-2 Risk for incident kidney diseases according to polygenic risk scores of kidney function 127

List of Figures

Figure 2-1 Distribution of estimated glomerular filtration rate and unadjusted and adjusted eGFR change over 30 years follow-up according to baseline hypertension status. 27

Figure 2-2 Distribution of predicted average annual change in estimated glomerular filtration rate within the ARIC population using baseline covariates according to baseline hypertension status. 29

Figure 3-1 Distribution of unadjusted predicted average annual change in estimated glomerular filtration rate within the ARIC population according to baseline body mass index tertile by sex and race. 55

Figure 3-2 Distribution of unadjusted predicted average annual change in estimated glomerular filtration rate within the ARIC population according to waist to hip ratio tertile by sex and race. 57

Figure 3-3 Distribution of unadjusted predicted average annual change in estimated glomerular filtration rate within the ARIC population according to baseline predicted percent fat tertile by sex and race. 59

Figure 4-1. The hypothesized bi-directional relations between kidney function and blood pressure depicted in a directed acyclic graph. 96

Figure 4-2 Selection of genetic instruments for glomerular filtration rate estimated from serum creatinine and systolic blood pressure. 97

Figure 4-3 Causal effect estimates of log(glomerular filtration rate estimated from serum creatinine) and log(blood urea nitrogen) on systolic blood pressure and diastolic blood pressure using the weighted mode method. 99

Figure 4-4 Causal effect estimates of systolic blood pressure and diastolic blood pressure on log(e glomerular filtration rate estimated from serum creatinine), log(blood urea nitrogen) and chronic kidney disease using the weighted model method. 100
Figure 5-1 Study design and workflow. 125
Figure 5-2 Association of deciles of LDPred polygenic risk score of kidney function with incident kidney diseases. 129
Figure 5-3 Scatter plots of correlations between protein and LDPred polygenic risk score for kidney function and correlation between protein and estimated glomerular filtration rate. 131

Chapter 1 Introduction

Chronic kidney disease (CKD), defined as reduced kidney function or abnormalities of kidney structure, is a global public health epidemic affecting over 10% of U.S. adults, with similarly high rates across the world.[1-3] The clinical impact of CKD is substantial: risks of end stage kidney disease (ESKD), cardiovascular disease (CVD), and premature mortality are all raised.[4-6] CKD has limited therapeutic options and is associated with high medical costs, which heightens the importance of prevention.

Accurate understanding of the risk factors and causes of kidney function decline as well as risk prediction are critical for guiding effective and efficient prevention strategies of CKD. Taking blood pressure as an example, identifying it as a genetic risk factor, i.e., cause, of kidney function decline may encourage earlier treatment.[7-10] We recognize that clinical trial data are already available and there are multiple risks in addition to CKD progression, yet hypertension awareness and control have much room for improvement.[11,12] Amassing very strong evidence and understanding the full range of risks, including CKD, coupled with identifying individuals at high risk of the disease can assist with enhanced prevention and treatment efforts.

Over the past decade, genome-wide association studies (GWAS) have revealed numerous genetic loci for kidney function, generating interests in the use of genetic information to examine biology, causal relationships and improve prediction of kidney diseases.[13-19] Mendelian randomization is a method to assess the causality of an observed association. It overcomes the

limitations of confounding and reverse causation in classical epidemiological studies by using genetic variants as instruments.[20] Polygenic risk scores (PRS) capture the polygenic architecture of complex diseases, including kidney diseases, by aggregating genome-wide genetic variation into a single score that reflects individual's disease risk. Both of these GWAS-based methods pose a new opportunity for better studying kidney function.

New methodologies for large scale proteomic measurement using aptamer technologies also provide an opportunity to assess the impact of genetic susceptibility to low kidney function on the plasma proteome.[21,22] Proteins are the basic building block of cells. There are reported to be around 10,000 to 12,000 proteins detected in human plasma or serum,[23,24] which are involved in various biological processes, including signaling, vascular and endothelial function, metabolism, and immune response.[25-27] Protein is a combination of nature, i.e. genetics, and nurture, i.e., environment, therefore it is very interesting to examine the balance of genetically predisposition vs. secondary influences on proteomic signals,[28] which is particularly important for kidney diseases since reduced kidney function results in elevations of many protein.

We thus comprehensively investigated kidney function by integrating multiple omics approaches into epidemiological studies. We used data from the Atherosclerosis Risk in Communities (ARIC) study, the UK Biobank (UKB), and summary statistics of GWAS meta-analysis of European-ancestry (EA) participants in the Chronic Kidney Genetic (CKDGen) Consortium, the UKB, and the International Consortium for Blood Pressure (ICBP). The ARIC study is a prospective population-based cohort (N=15,792; Mean age: 54; % female: 57%) with 30-year follow up. It collected blood sample and a rich set of traditional risk factors at baseline and

follow-up visits.[29] Collaborating with SomaLogic, the ARIC study will be able to assay ~5,000 proteins at two time points that are around 20 years apart.[21] The UKB is a prospective population-based cohort (EA: N=458,577; Mean age: 56.8; % female: 54.2%) with deep genetic and phenotypic data collected on participants across the United Kingdom.[30] The CKDGen Consortium is a trans-ethnic consortium (EA: N=567,460; Mean age: 50.1; % female: 51.8%) combined >75 epidemiologic studies, mostly population-based, with genome-wide genetic data and kidney function measurements.[13] The ICBP is an international consortium to investigate blood pressure genetics (EA: N=299,024; Mean age: 54.9; % female: 55.1%).[31]

Using the above-mentioned data, we conducted four studies to shed light on CKD progression, causes and genetics. We first rigorously examined the associations between potential risk factors, hypertension and obesity, and prospective kidney function decline and related kidney diseases. We used longitudinal data from the ARIC study. Kidney function was measured as glomerular filtration rate (GFR) estimated by the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) creatinine equation (eGFRcr) from visit 1 (1987-1989) to visit 6 (2016-2017)**.** We modeled long-term kidney function trajectories over 30 years of follow up and use mixed models to evaluate the associations of potential risk factors with future decline in kidney function and development of CKD and ESKD. The two studies of this step are detailed in Chapter 2 and Chapter 3.

The editorial to our Chapter 2 paper summarized our "several important results" and concluded that our paper was "a significant addition to the existing literature". But it also asked an important question: "is the association causal"?[32,33] To evaluate the causal relation between

kidney function and blood pressure, we used GWAS results from the CKDGen consortium, the UKB, and the ICBP to conduct a two-sample bi-directional Mendelian randomization (MR) analysis.[13,31] To obtain robust conclusions from our analyses, we paid particular attention to two critical aspects in this MR study. One being the use of serum creatinine for GFR estimation, which might link eGFRcr to genetic variants more related to creatinine metabolism than glomerular filtration function, making it difficult to interpret any causal findings between eGFRcr and blood pressure. To address this issue, we used additional data from large-scale meta-analysis of GWAS of blood urea nitrogen (BUN) as a complementary kidney function biomarker to select genetic instruments that are more likely to be specific to kidney function. The second being the assumption of the lack of horizontal pleiotropy of the genetic instruments, that is the genetic instruments must be associated with the outcome through the exposure only. This assumption is usually difficult to assess and verify.[34] To address this issue, we analyzed the data using multiple statistical MR methods and prioritized the method that are known to be most robust to the presence of horizontal pleiotropy.[35] The details of this study can be found in Chapter 4.

After successfully using GWAS results for examining causal relations, we further examined the power of genetic susceptibility for kidney function to predict future risk. To better understand GWAS results, we also investigated the association between genetically predicted kidney function decline with an intermediate phenotype, i.e., plasma proteins. Using large studies and new algorithms, we derived a range of PRS for kidney function measured as eGFRcr, including a genome-wide score as a weighted average of 1.2 million common SNPs (LDPred PRS).[36] We investigated the strengths of associations of PRS for kidney function with four incident kidney

diseases, CKD, ESKD, kidney failure, and acute kidney injury (AKI), among EA participants in the ARIC study. We then examined ~5,000 plasma proteins measured at two time points, visit 3 (1993-1995) to visit 5 (2011-2013)**,** in relation to both PRS for kidney function and the concurrent kidney function itself. This work corresponds to the content in Chapter 5.

Overall, supported by multiple methods and data sources, this doctoral thesis provided a unique and comprehensive view of kidney function decline from association to cause to risk prediction.

Chapter 2 Association between Hypertension and Kidney Function Decline: the Atherosclerosis Risk in Communities Study

Zhi Yu, BM, MS,[1,2,*] Casey M. Rebholz, PhD, MS, MNSP, MPH,[1,2,*], Eugenia Wong, MHS,[1,2] Yuan Chen, MHS,[3] Kunihiro Matsushita, MD, PhD,[1,2] Josef Coresh, MD, PhD,[1,2] Morgan E. Grams, MD, PhD[1,2,4]

[1] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

[2] Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore, Maryland

[3] Department of Biostatistics, Columbia University Mailman School of Public Health, New York, New York

[4] Division of Nephrology, Department of Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland

* ZY and CMR contributed equally to this work

ABSTRACT

Rationale & Objective: Hypertension is a risk factor for chronic kidney disease progression. The relationship between hypertension and change in kidney function over time among individuals with preserved kidney function, and whether it differs by race, has not been elucidated.

Study Design: Observational study.

Setting & Participants: 14,854 participants from the Atherosclerosis Risk in Communities (ARIC) study.

Predictors: Hypertension status was assessed at baseline (1987-89) and was categorized following the classification in the 2017 American College of Cardiology/American Heart Association Clinical Practice Guideline as: normal blood pressure; elevated blood pressure; stage 1 hypertension; stage 2 hypertension without medication; or stage 2 hypertension with medication.

Outcomes: Estimated glomerular filtration rate (eGFR) calculated using serum creatinine measured at baseline and 4 follow-up study visits over 30 years.

Results: At baseline, 13.2%, 7.3%, and 19.4% Caucasians and 15.8%, 14.9%, and 39.9% African Americans were categorized to stage 1 hypertension, stage 2 hypertension no medication, and stage 2 hypertension with medication. Individuals with hypertension had significantly greater decline in eGFR over the 30-year follow-up than those without hypertension. Adjusted annual difference in eGFR decline compared to normal blood pressure was fairly similar across race groups (Caucasians: elevated blood pressure: -0.11 ml/min/1.73 m^2; stage 1 hypertension: -0.15 ml/min/1.73 m^2; stage 2 hypertension without medication: -0.36 ml/min/1.73 m^2; stage 2 hypertension with medication: -0.17 ml/min/1.73 m^2; African

Americans: elevated blood pressure: -0.21 ml/min/1.73 m^2; stage 1 hypertension: -0.16 ml/min/1.73 m^2; stage 2 hypertension without medication: -0.50 ml/min/1.73 m^2; stage 2 hypertension with medication: -0.16 ml/min/1.73 m^2). The average 30-year predicted probability of developing CKD stage G3a+ with normal blood pressure, elevated blood pressure, stage 1 hypertension, stage 2 hypertension without medication, or stage 2 hypertension with medication was 54.4%, 61.6%, 64.7%, 78.1%, and 70.9%, respectively, among Caucasians and 55.4%, 62.8%, 60.9%, 76.1%, and 66.6%, respectively, among African Americans.

Limitations: Five eGFR measurements for estimating long-term trajectories; potential differential loss to follow-up for participants with hypertension at baseline.

Conclusions: Hypertension status at baseline was associated with faster kidney function decline over 30-year follow-up in a general population cohort.

INTRODUCTION

Hypertension ranks as the top risk factor for chronic disease worldwide.[37] People with hypertension have an increased risk for myocardial infarction, stroke, heart failure, and kidney failure.[38] According to the 2017 American College of Cardiology/American Heart Association Clinical Practice Guideline, the prevalence of hypertension among U.S. adults was 45.6%.[39,40] Hypertension is a risk factor for kidney disease progression in individuals with CKD,[8,41] but few studies have addressed the relationship between hypertension and longitudinal change in kidney function in the general population.[42] Furthermore, the extent to which hypertension precedes kidney function decline or is simply a consequence of lower kidney function continues to be an area of controversy.[43]

African Americans have a substantially higher risk of hypertension than Caucasians and, among those with hypertension, poorer hypertension control.[44-47] There are also profound racial disparities in kidney disease, with African Americans approximately being 1.5-times more likely to develop CKD and three times more likely to develop end-stage renal disease compared to Caucasians.[48-51] Racial disparities may in part be explained by a greater burden of risk factors among African Americans, including higher prevalence of hypertension, diabetes mellitus and the *APOL1* genetic risk variant.[52-55] However, it is also possible that the risk relationship between hypertension may be stronger in African Americans than Caucasians, either due to heightened susceptibility to disease or poorer risk factor control.

As such, the purpose of this study was to evaluate the association of hypertension with trajectories of estimated glomerular filtration rate (eGFR) and to assess whether the risk of

kidney outcomes associated with hypertension varied by race in a community-based cohort of 14,854 Caucasian and African American adults over 30 years of follow-up.

METHODS

Study Design & Study Population

The Atherosclerosis Risk in Communities (ARIC) study is a prospective cohort designed to investigate the etiology of atherosclerosis and its clinical consequences as well as to examine variability in disease risk according to characteristics of the study population.[29] The ARIC study enrolled a total of 15,792 middle-aged (45-64 years old at baseline), predominantly Caucasian and African American men and women from four communities in the U.S.: Forsyth County, North Carolina; Jackson, Mississippi; suburbs of Minneapolis, Minnesota; and Washington County, Maryland. The initial examination took place in 1987-1989 (baseline, study visit 1). Follow-up examinations occurred at approximately three-year intervals: 1990-1992 (study visit 2), 1993-1995 (study visit 3), 1996-1998 (study visit 4), more recently, in 2011-2013 (study visit 5), and in 2016-2017 (study visit 6). During each study visit, an extensive questionnaire was administered, a clinical examination was conducted, and blood and urine specimens were collected.

In the present study, participants were excluded if they had missing data on hypertension status at baseline, missing measurement of serum creatinine at baseline, had eGFR <60 ml/min/1.73 m^2 at baseline, prevalent end-stage renal disease, self-reported race other than Caucasian or African American, or missing covariates. After these exclusions, the analytic sample size was 14,854

(94% of the original cohort). Study participants provided written documentation of informed consent and study protocols were approved by the institutional review board at each study site.

Assessment of Hypertension Status

Systolic blood pressure (SBP) and diastolic blood pressure (DBP) were measured twice at visit 4 and three times at other visits using a random-zero sphygmomanometer while seated after resting for five minutes in a separate, quiet room. Participants were requested to avoid vigorous physical activity, cigarette smoking, and consumption of food, caffeinated beverages, and alcohol for twelve hours prior to the study visit. The appropriate cuff size was selected after measuring arm circumference. The first and second blood pressure values at visit 4 and the second and third blood pressure values at other visits were averaged and used in the analysis. Baseline hypertension status was categorized according to the criteria in the 2017 American College of Cardiology/American Heart Association Clinical Practice Guideline as normal blood pressure (SBP <120 mm Hg and DBP <80 mm Hg), elevated blood pressure (120 mm Hg ≤SBP <130 mm Hg and DBP <80 mm Hg), stage 1 hypertension (130 mm Hg ≤SBP <140 mm Hg or 80 mm Hg ≤DBP <90 mm Hg), stage 2 hypertension without medication (SBP ≥140 mm Hg or DBP ≥90 mm Hg), and stage 2 hypertension with medication (use of anti-hypertensive medication in the last two weeks).[40]

Assessment of Kidney Function

Kidney function was assessed by measuring creatinine in serum or plasma specimens collected during each study visit, except for study visit 3. In our study, we used five eGFR measurements (visits 1, 2, 4, 5, and 6) for the estimation of trajectories. The modified kinetic Jaffe method was

used for the measurement of creatinine with standardization to the National Institute of Standards and Technology (NIST) standard and calibration across study visits using repeated measurements from a sample of 200 ARIC study participants.[56-58] The Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation was used to calculate estimated glomerular filtration rate (eGFR) based on creatinine.[59] For those participants who developed incident end-stage renal disease (ascertained via linkage to the United States Renal Data System), an eGFR value of 15 mL/min/1.73 m^2 was imputed on the date of initiation of renal replacement therapy (transplant, dialysis).

Assessment of Covariates

Demographic characteristics (date of birth for the calculation of age, sex, race, education, family income), lifestyle factor (smoking), and medical history (diabetes, coronary heart disease) were ascertained via questionnaire administered by trained interviewers at the baseline study visit. Study participants brought medications to the study visit and the names of all medications were transcribed, including antihypertensive medications. Body mass index was calculated using weight in kilograms divided by the square of height in meters measured during the study visit. Blood samples that were collected from study participants during the baseline study visit were assayed for the measurement of concentrations of high density lipoprotein cholesterol using an enzymatic method after precipitation with dextran sulfate-magnesium and glucose using the modified hexokinase/glucose-6-phosphate dehydrogenase method.[60] Diabetes was defined as fasting glucose ≥126 mg/dL, non-fasting glucose ≥200 mg/dL, self-report of diagnosed diabetes, or use of diabetes medication in the past two weeks.

Statistical Analysis

Baseline characteristics of the study population were compared by hypertension status and racial group using descriptive statistics and differences were tested using ANOVA for continuous variables and χ^2 tests for categorical variables.

Mixed models were used to evaluate the association between hypertension status at baseline (normal blood pressure / elevated blood pressure / stage 1 hypertension / stage 2 hypertension no medication / stage 2 hypertension with medication) and eGFR trajectories using random intercepts and random slopes to account for individual variation in eGFR at baseline and its change. Because the random slopes had higher variance among African Americans than Caucasians, our models were conducted overall and after stratifying by race. Covariates included in the adjusted models were age (continuous), sex, body mass index (BMI, continuous), race-center (Minneapolis, Minnesota, and Washington County, Maryland, where all participants were white; Jackson, Mississippi, where all participants were African American; and Forsyth County, North Carolina, which recruited both whites and African Americans, and was represented by two variables) or center only for race-stratified analysis, smoking (current / former / never), family income (annual income ≥$25,000 / <$25,000 / not reported), education (high school graduated / not graduated), high-density lipoprotein cholesterol (continuous), diabetes (yes / no), and history of coronary heart disease (yes / no) at baseline. We tested for interaction by race by adjusting for race and center separately and including a three-way product term of hypertension category, race, and time. We examined and plotted the patterns of eGFR change over time (i.e., trajectories from best linear unbiased prediction estimates) and estimated the differences in annual eGFR decline according to hypertension status.[61] Kernel density plots were used to illustrate the

distribution of unadjusted and adjusted annual predicted change in eGFR. The average probability (absolute risk) of developing different stages of chronic kidney disease (G3a+: eGFR <60 mL/min/1.73 m^2; G3b+: eGFR <45 mL/min/1.73 m^2; G4+: eGFR <30 mL/min/1.73 m^2) over 30 years of follow-up was estimated for the baseline population based on the baseline covariates. These probabilities were expressed using best linear unbiased predictions from race-stratified models according to hypertension status.[62]

In a sensitivity analysis, we examined the associations between blood pressure category at baseline (SBP <130 mm Hg and DBP <80 mm Hg; 130 mm Hg ≤SBP <140 mm Hg or 80 mm Hg ≤DBP <90 mm Hg; SBP ≥140 mm Hg or DBP ≥90 mm Hg) and eGFR trajectory after adjusting for hypertension medication status at baseline (yes / no). We also examined the associations of interest after imputing the eGFR at the time of initiation of renal replacement therapy using information supplied on the 2728 form to test the robustness of our main results. All analyses were conducted using Stata statistical software version 13 (StataCorp, College Station, Texas) and R version 3.3.3 (R Development Core Team).

RESULTS

Baseline Characteristics

The baseline characteristics of the 14,854 study participants (11,003 Caucasian and 3,851 African American) according to hypertension status category and racial group are shown in **Table 2-1**. At baseline, 13.2%, 7.3%, and 19.4% Caucasians and 15.8%, 14.9%, and 39.9% African Americans were categorized to stage 1 hypertension, stage 2 hypertension no medication, and stage 2 hypertension with medication. In both Caucasians and African

Americans, participants with hypertension, and particularly those with stage 2 hypertension, had higher body mass index ($p<0.001$). Individuals with stage 2 hypertension were more likely to have diabetes, annual family income less than $25,000, and less likely to be high school graduates ($p<0.001$ for all comparisons). Individuals in the stage 2 hypertension with medication category were more likely to have a history of coronary heart disease ($p<0.001$). Although there was a statistically significant difference in baseline kidney function by hypertension status ($p<0.001$), the absolute difference in eGFR values was relatively small.

Average eGFR Trajectories

There was a steady decline in eGFR over time among both Caucasians and African Americans in each of the five hypertension status categories (**Figure 2-1**). Compared with individuals without hypertension, the slopes of the trajectory among participants with hypertension were steeper, representing faster eGFR decline. After adjusting for risk factors, individuals with hypertension compared to those without hypertension had significantly greater decline in eGFR over the 30 years of follow-up. The decline in eGFR among individuals in the stage 2 hypertension with medication category was similar to that among those in stage 1 hypertension category (differences in eGFR decline per year: elevated blood pressure: -0.12 ml/min/1.73 m^2; stage 1 hypertension: -0.14 ml/min/1.73 m^2; stage 2 hypertension without medication: -0.39 ml/min/1.73 m^2; stage 2 hypertension with medication: -0.16 ml/min/1.73 m^2). Similar results were found when examining Caucasians and African Americans separately (Caucasians: elevated blood pressure: -0.11 ml/min/1.73 m^2; stage 1 hypertension: -0.15 ml/min/1.73 m^2; stage 2 hypertension without medication: -0.36 ml/min/1.73 m^2; stage 2 hypertension with medication: -0.17 ml/min/1.73 m^2; African Americans: elevated blood pressure: -0.21 ml/min/1.73 m^2; stage 1

hypertension: -0.16 ml/min/1.73 m^2; stage 2 hypertension without medication: -0.50 ml/min/1.73 m^2; stage 2 hypertension with medication: -0.16 ml/min/1.73 m^2) (**Table 2-2**). There was no interaction in the association between hypertension and eGFR decline with race except for the stage 2 hypertension with medication category (p for interaction: 0.01).

Variation in Annual Change in eGFR

Figure 2-2 shows the distribution of eGFR decline by blood pressure category. The unadjusted median (IQR) annual eGFR change of Caucasian and African American participants, respectively, were -1.32 (-1.51, -1.11) mL/min/1.73 m^2/year and -1.79 (-2.07, -1.45) mL/min/1.73 m^2/year among those with normal blood pressure, -1.48 (-1.67, -1.31) mL/min/1.73 m^2/year and -2.10 (-2.34, -1.77) mL/min/1.73 m^2/year among those with elevated blood pressure, -1.47 (-1.66, -1.26) mL/min/1.73 m^2/year and -2.00 (-2.28, -1.62) mL/min/1.73 m^2/year among those with stage 1 hypertension, -1.71 (-1.93, -1.51) mL/min/1.73 m^2/year and -2.39 (-2.64, -1.94) mL/min/1.73 m^2/year among those with stage 2 hypertension without medication, and -1.61 (-1.81, -1.40) mL/min/1.73 m^2/year and -2.25 (-2.55, -1.79) mL/min/1.73 m^2/year among those with stage 2 hypertension with medication. Overlap was greater when eGFR trajectories were adjusted for baseline covariates. Compared with Caucasians, African Americans had similar differences by hypertension status but larger mean and variance of the annual eGFR rate of decline.

Predicted Probability of Chronic Kidney Disease

The predicted probability of chronic kidney disease of 30 years was generally higher among people with hypertension (**Table 2-3**). African Americans had a similar predicted risk of

developing stage G3a+ but a greater predicted risk of developing stage G3b+ and G4+ CKD compared with Caucasians. The average 30-year predicted probability of developing CKD stage G3a+ (eGFR <30 mL/min/1.73 m^2) with normal blood pressure, elevated blood pressure, stage 1 hypertension, stage 2 hypertension no medication, or stage 2 hypertension with medication was 54.4%, 61.6%, 64.7%, 78.1%, and 70.9%, respectively, among Caucasians and 55.4%, 62.8%, 60.9%, 76.1%, and 66.6%, respectively, among African Americans; that of stage G4+ was 7.0%, 9.0%, 10.1%, 15.8%, and 12.4% among Caucasians and 22.3%, 26.0%, 25.4%, 32.5%, and 27.5% among African Americans.

Sensitivity Analysis

Replacing hypertension stages with two separate variables, blood pressure categories and antihypertensive medication, showed similar results. In the total population, individuals with higher blood pressure had significantly greater decline in eGFR over the 30 years of follow-up (120 ≤SBP <130 mm Hg and DBP <80 mm Hg: -0.15 ml/min/1.73 m^2 per year; 130 ≤SBP <140 mm Hg or 80 ≤DBP <90 mm Hg: -0.14 ml/min/1.73 m^2 per year; SBP ≥140 mm Hg or DBP ≥90 mm Hg: -0.42 ml/min/1.73 m^2 per year), compared to those with normal blood pressure (SBP <120 mm Hg and DBP <80 mm Hg), after adjusting for risk factors and hypertension medication status (**Supplementary Table 2-1**). Results were similar when examining Caucasians and African Americans separately. After accounting for blood pressure categories, hypertension medication was not associated with eGFR decline in the total population or in analyses stratified by race (p=0.99 for overall, 0.38 among Caucasians, and 0.21 among African Americans). When eGFR at the initiation of renal replacement therapy was imputed from the 2728 form rather than

using a value of 15 ml/min/1.73 m^2, results were nearly identical to those in the main analysis (**Supplementary Table 2-2**).

DISCUSSION

In this community-based population of 14,854 middle-aged adults, participants with hypertension at baseline experienced faster decline in kidney function than those without hypertension over 30 years of follow-up. The risk of developing CKD was greater with hypertension, especially stage 2 hypertension in both Caucasians and African Americans. Although there was no difference by race in the association between hypertension and eGFR decline, African Americans did have higher mean and dispersion in the rate of decline and risk of developing CKD stage G3b or worse, which translated into a greater absolute risk difference between those with and without hypertension.

The current study adds to existing literature by demonstrating that hypertension is a risk factor for eGFR decline over 30 years of follow-up in a population with relatively preserved kidney function. We defined hypertension at baseline in order to definitively establish the temporal relationship between onset of hypertension and eGFR decline. Other studies suggest that this is a conservative approach.[63] Previous work in the Chronic Renal Insufficiency Cohort (CRIC) study evaluated baseline blood pressure as a predictor of kidney function decline,[63] but this study included only participants with CKD.[64] The Multi-Ethnic Study of Atherosclerosis (MESA) demonstrated that higher SBP and variable pulse pressure were significantly associated with cystatin C-based eGFR (eGFRcys) decline in participants who attended a 5- and 10-year visit.[10]

High blood pressure was also an independent predictor of age-adjusted annual eGFR decline over 10 years among Italian individuals with preserved kidney function and type 2 diabetes.[65]

Our study demonstrates the variability in slope by hypertension and race among middle-aged men and women with preserved kidney function, finding that eGFR decline is more variable among African Americans than Caucasians. The higher variability in eGFR decline among African Americans resulted in disproportionately greater probability of development of advanced CKD, despite relatively small differences in annual eGFR decline. Greater variability in quality of medical care or control of risk factors among African Americans may play a role in the greater heterogeneity of disease progression.[66,67]

Our finding that African Americans had similar probability of early stage CKD compared with Caucasians but greater probability of late stage CKD is consistent with previous studies which found significantly faster renal progression in African Americans compared with Caucasians.[68] Previous prediction models developed among populations of healthy individuals also showed higher risk of developing end stage kidney disease (ESKD) for African Americans.[69] The difference may be attributable to both biological differences and treatment barriers.[55,70,71] In particular, genetic variants of Apolipoprotein L1 (*APOL1*) have been associated with worse kidney outcomes, and carriers of these variants are overwhelmingly of African descent.[55,72]

The mechanism underlying the association between hypertension and eGFR decline is not fully understood, but may be due in part to higher intraglomerular pressure and progressive arteriosclerosis.[73,74] However, low GFR may also increase blood pressure due to impairments in

salt and water excretion. For this reason, we evaluated blood pressure only at visit 1, in order to maintain strict temporality. Others have shown that time-updated blood pressure is more strongly associated with kidney disease risk, which may be due in part to this bidirectional association.[8]

There are several strengths to this study. The ARIC study is a large, prospective cohort with 30 years of follow-up and representation from 4 US communities with both Caucasians and African Americans. The long duration of the study allows for the characterization of kidney function decline in a population that was generally healthy at the outset. Multiple established risk factors were collected in a research protocolized manner.

There are also several study limitations to acknowledge. There were only up to five eGFR measurements for the estimation of long-term trajectories. That said, there are few longitudinal, population-based cohorts that have had more frequent measurements of eGFR over 30 years. Despite the likelihood that participants are healthier than the general population, some degree of misclassification may have occurred due to possible non-compliance with anti-hypertensive medication. Hypertension treatment practice has changed over time; however, we chose to use baseline rather than time-varying hypertension status as a more conservative approach with a clear temporal relationship with subsequent eGFR decline. Furthermore, we were concerned about the potential for a bidirectional association between eGFR and hypertension and the possibility of introducing time-varying confounding. Participants who develop ESKD are less likely to survive to attend subsequent study visits; we included an estimate of their trajectory by imputing eGFR at the time of ESKD onset as identified through linkage to the USRDS registry. There is the potential for differential loss to follow-up by baseline hypertension status. Our

analysis did not attempt to capture acute kidney injury (AKI) occurring over follow-up, precluding inferences about the impact of AKI on eGFR decline. For blood pressure measurement, we characterized hypertension based on 2 measurements at the same occasion rather than ≥2 occasions as stipulated by the 2017 American College of Cardiology/American Heart Association Clinical Practice Guideline. Also, blood pressure was measured by random-zero sphygmomanometer. Lastly, as ARIC only included Caucasians and African Americans, our results are not applicable to other ethnic groups.

In summary, hypertension status is an important risk factor for future decline in kidney function and the development of kidney disease in community-dwelling Caucasian and African American adults. Our study highlights the potential importance of preventing and treating hypertension as a strategy to preserve kidney function over time. Population-level efforts to lower blood pressure may help to reduce the onset of kidney disease.

Table 2-1 Baseline characteristics of the study population according to baseline hypertension status (N=14,854).

Characteristic[a]	Caucasian N=11,003					African American N=3,851				
	Normal BP[b] N=5,341	Elevated BP[b] N=1,281	Stage 1 HTN[b] N=1,448	Stage 2 HTN without MED[b] N=801	Stage 2 HTN with MED[b] N=2,132	Normal BP[b] N=859	Elevated BP[b] N=273	Stage 1 HTN[b] N=610	Stage 2 HTN without MED[b] N=573	Stage 2 HTN with MED[b] N=1,536
Age, years[c]	53.5	56.1	55.0	56.7	56.8	52.2	54.5	52.9	54.6	55.0
	(5.5)	(5.6)	(5.7)	(5.6)	(5.4)	(5.6)	(6.1)	(5.6)	(5.7)	(5.7)
Female[c]	3052	639	630	381	1107	537	179	328	277	1039
	(57.1)	(49.9)	(43.5)	(47.6)	(51.9)	(62.5)	(65.6)	(53.8)	(48.3)	(67.6)
Smoking status[c]										
Current	1512	322	290	152	458	268	89	179	200	416
	(28.3)	(25.1)	(20.0)	(19.0)	(21.5)	(31.3)	(32.7)	(29.3)	(34.9)	(27.1)
Former	1723	456	560	308	826	209	65	161	122	377
	(32.3)	(35.6)	(38.7)	(38.5)	(38.8)	(24.4)	(23.9)	(26.4)	(21.3)	(24.6)

Never	2103	503	598	341	847	379	118	270	251	742
	(39.4)	(39.3)	(41.3)	(42.6)	(39.7)	(44.3)	(43.4)	(44.3)	(43.8)	(48.3)
High school	4573	1040	1214	637	1660	587	162	377	318	822
graduate[c]	(85.7)	(81.3)	(83.9)	(79.5)	(77.9)	(68.4)	(59.3)	(62.1)	(55.7)	(53.6)
Annual family	1150	335	345	248	696	456	169	372	389	1054
income <$25,000[c]	(21.5)	(26.2)	(23.8)	(31.0)	(32.6)	(53.1)	(61.9)	(61.0)	(67.9)	(68.6)
Center[c]										
Forsyth county	1772	420	386	243	569	135	46	51	49	170
	(33.2)	(32.8)	(26.7)	(30.3)	(26.7)	(15.7)	(16.8)	(8.4)	(8.6)	(11.1)
Jackson	0	0	0	0	0	705	223	553	518	1349
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(82.1)	(81.7)	(90.7)	(90.4)	(87.8)
Minneapolis	1879	433	606	303	656	9	4	1	3	5
	(35.2)	(33.8)	(41.9)	(37.8)	(30.8)	(1.0)	(1.5)	(0.2)	(0.5)	(0.3)
Washington	1690	428	456	255	907	10	0	5	3	12
county	(31.6)	(33.4)	(31.5)	(31.8)	(42.5)	(1.2)	(0.0)	(0.8)	(0.5)	(0.8)
Systolic BP, mm	106.0	123.9	128.5	148.2	128.0	107.9	124.0	126.4	152.2	132.7
Hg[c]	(8.5)	(2.8)	(7.6)	(11.8)	(17.5)	(7.3)	(2.9)	(8.1)	(18.4)	(21.3)

Diastolic BP, mm	65.9	71.2	79.2	84.2	75.7	69.2	72.4	81.2	91.0	81.8
Hg[c]	(7.2)	(6.2)	(6.7)	(10.3)	(10.1)	(6.1)	(5.9)	(5.9)	(12.8)	(12.1)
BMI, kg/m^2 [c]	25.7	27.3	27.7	28.3	29.1	27.8	29.0	29.0	29.2	31.0
	(4.1)	(4.8)	(5.0)	(5.4)	(5.4)	(5.1)	(6.0)	(6.0)	(6.3)	(6.4)
HDL-C, mg/dL[c]	52.4	50.8	50.0	50.1	46.2	56.0	57.4	56.8	56.8	52.8
	(17.1)	(17.0)	(16.5)	(16.8)	(15.5)	(17.4)	(17.2)	(19.2)	(18.9)	(16.2)
Diabetes[c]	198	88	92	61	326	88	34	65	80	374
	(3.7)	(6.9)	(6.4)	(7.6)	(15.3)	(10.2)	(12.5)	(10.7)	(14.0)	(24.3)
CHD[c]	190	60	51	26	234	24	4	16	12	94
	(3.6)	(4.7)	(3.5)	(3.2)	(11.0)	(2.8)	(1.5)	(2.6)	(2.1)	(6.1)
eGFR,	101.5	99.5	99.5	99.0	96.5	114.8	115.7	114.4	113.2	109.6
mL/min/1.73 m^2[c]	(11.2)	(10.9)	(11.3)	(11.3)	(12.7)	(15.5)	(14.0)	(15.9)	(15.7)	(18.6)

[a] Mean (standard deviation) for continuous variables and % (n) for categorical variables.

[b] Normal blood pressure defined as systolic blood pressure <120 mm Hg and diastolic blood pressure <80 mm Hg. Elevated blood pressure defined as systolic blood pressure 120 ≤systolic blood pressure <130 mm Hg and diastolic blood pressure <80 mm Hg. Stage 1 hypertension defined as 130 ≤systolic blood pressure <140 mm Hg or 80 ≤diastolic blood pressure <90 mm Hg. Stage 2 hypertension defined as systolic blood pressure ≥140 mm Hg or diastolic blood pressure ≥90 mm Hg and this group was stratified by use of anti-hypertensive medication in the last two weeks.

[c] P-value for comparing the hypertension groups within each race group < 0.05; P-value calculated by ANOVA for continuous variables and χ^2 test for categorical variables

BMI, body mass index; BP, blood pressure; CHD, coronary heart disease; eGFR, estimated glomerular filtration rate; HDL-C, high density lipoprotein cholesterol; HTN, hypertension; MED, medication

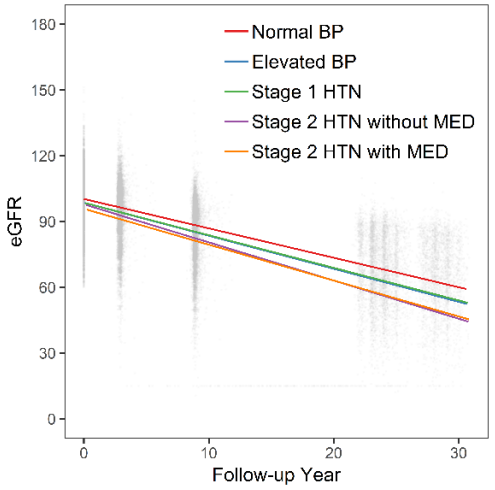
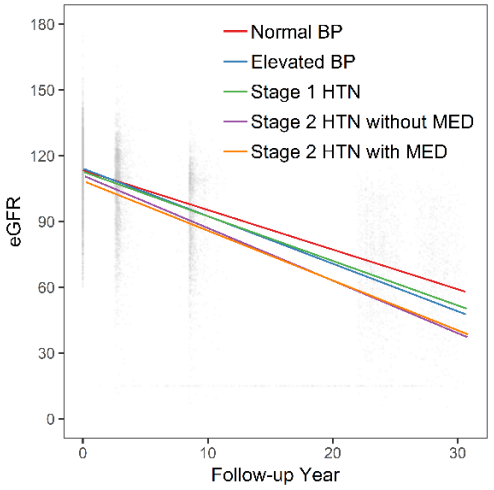
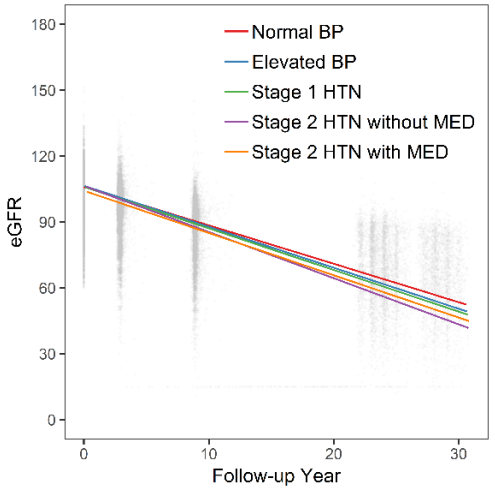
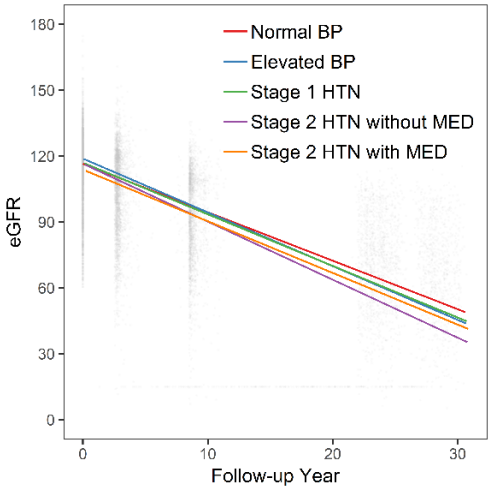
	Caucasian	African American
Unadjusted	 **A**	 **B**
Adjusted	 **C**	 **D**

Figure 2-1 Distribution of eGFR (unit: ml/min per 1.73 m^2) and unadjusted and adjusted eGFR change over 30 years follow-up according to baseline hypertension status. (A) unadjusted eGFR change among Caucasian (B) unadjusted eGFR change among African American (C) adjusted eGFR change among Caucasian (D) adjusted eGFR change among African American.

For adjusted eGFR changes, model adjusted for age (centered at 50 years old), sex (reference group: male), center (reference group: Forsyth County, North Carolina), baseline smoking status (reference group: current smoker), baseline education level (reference group: non-high school graduate), baseline annual family income (reference group: <$25,000), baseline body mass index (centered at 25 kg/m^2), baseline high density lipoprotein cholesterol level (centered at 40 mg/dL), baseline history of diabetes (reference group: no diabetes), baseline history of coronary heart disease (reference group: no coronary heart disease), as well as their interaction with follow-up time. For adjusted predicted average annual changes among African Americans, African Americans in the Minnesota and Washington County cohorts were excluded in adjusted model because of small numbers. The number of participants at each visit are: visit 1: Caucasian: N=11,003, African American: N=3,851; visit 2: Caucasian: N=10,297 (93.6% of the original cohort), African American: N=3,224 (83.7% of the original cohort); visit 4: Caucasian: N=8,616 (78.3%), African American: N=2,373 (61.6%); visit 5: Caucasian: N=4,758 (43.2%), African American: N=1,375 (35.7%); visit 6: Caucasian: N=2,995 (27.2%), African American: N=1,013 (26.3%). The total number of eGFR measurements was 49,502, and the median (IQR) of eGFR measurements was 3 (3,4).
BP, blood pressure; HTN, hypertension; MED, medication

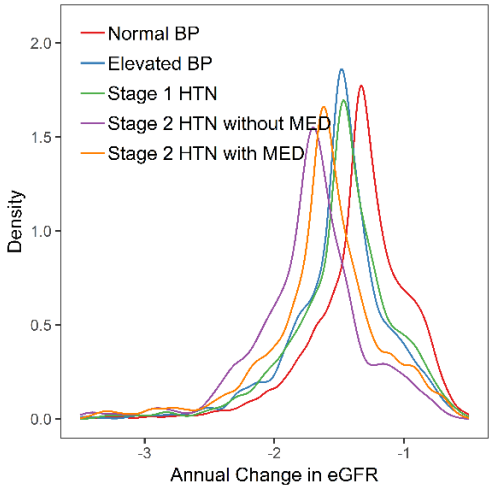
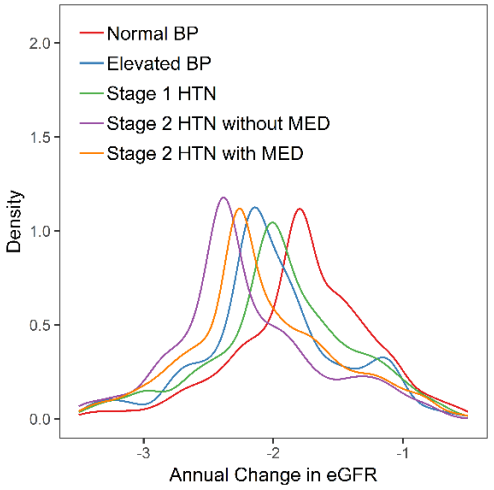
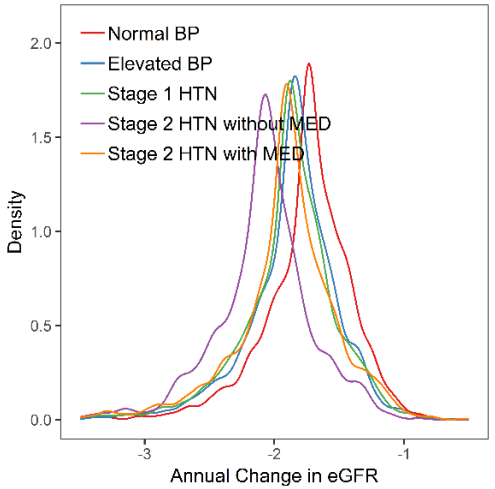
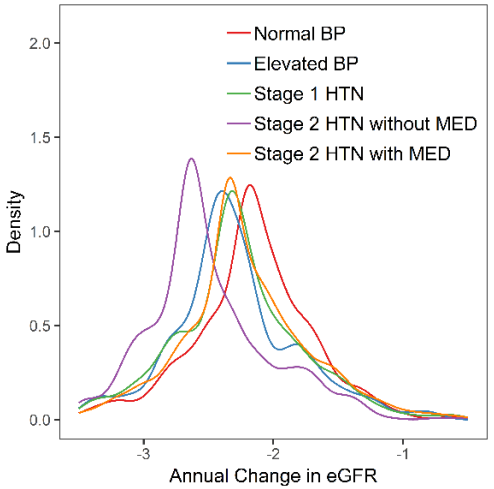
	Caucasian	African American
Unadjusted	 **A**	 **B**
Adjusted	 **C**	 **D**

Figure 2-2 Distribution of predicted average annual change in eGFR (unit: ml/min per 1.73 m^2) within the ARIC population using baseline covariates according to baseline hypertension status. (A) unadjusted predicted average annual change among Caucasian (B) unadjusted predicted average annual change among African American (C) adjusted predicted average annual change among Caucasian (D) adjusted predicted average annual change among African American.

For adjusted predicted average annual changes, slopes were estimated from a mixed model adjusted for age (centered at 50 years old), sex (reference group: male), center (reference group: Forsyth County, North Carolina), baseline smoking status (reference group: current smoker), baseline education level (reference group: non-high school graduate), baseline annual family income (reference group: <$25,000), baseline body mass index (centered at 25 kg/m^2), baseline high density lipoprotein cholesterol level (centered at 40 mg/dL), baseline history of diabetes (reference group: no diabetes), baseline history of coronary heart disease (reference group: no coronary heart disease), as well as their interaction with follow-up time. For adjusted predicted average annual changes among African Americans, African Americans in the Minnesota and Washington County cohorts were excluded in adjusted model because of small numbers.

BP, blood pressure; HTN, hypertension; MED, medication

Table 2-2 Differences (95% CIs) in eGFR decline (unit: ml/min per 1.73 m^2) per year according to hypertension categories in the total population and by race.

Population	Differences (95% CIs) in eGFR decline (unit: ml/min per 1.73 m^2) per year				
	Normal BP[a]	**Elevated BP[a]**	**Stage 1 HTN[a]**	**Stage 2 HTN**	**Stage 2 HTN**
	N=6,200	**N=1,554**	**N=2,058**	**without MED[a]**	**with MED[a]**
				N=1,374	**N=3,668**
Unadjusted	Ref	-0.22 (-0.16, -0.29)	-0.24 (-0.19, -0.30)	-0.58 (-0.50, -0.65)	-0.49 (-0.44, -0.53)
Adjusted[b,c]	Ref	-0.12 (-0.06, -0.18)	-0.14 (-0.08, -0.19)	-0.39 (-0.31, -0.46)	-0.16 (-0.11, -0.21)
Caucasian	**N=5,341**	**N=1,281**	**N=1,448**	**N=801**	**N=2,132**
Unadjusted	Ref	-0.17 (-0.11, -0.22)	-0.14 (-0.09, -0.20)	-0.40 (-0.32, -0.47)	-0.30 (-0.24, -0.35)
Adjusted[d]	Ref	-0.11 (-0.05, -0.17)	-0.15 (-0.09, -0.20)	-0.36 (-0.28, -0.44)	-0.17 (-0.11, -0.22)
African American	**N=859**	**N=273**	**N=610**	**N=573**	**N=1,536**
Unadjusted	Ref	-0.36 (-0.13, -0.59)	-0.23 (-0.05, -0.40)	-0.59 (-0.40, -0.78)	-0.46 (-0.32, -0.61)
Adjusted[d,e]	Ref	-0.21 (0.01, -0.43)	-0.16 (0.01, -0.33)	-0.50 (-0.31, -0.69)	-0.16 (-0.01, -0.31)

[a] Normal blood pressure defined as systolic blood pressure <120 mm Hg and diastolic blood pressure <80 mm Hg. Elevated blood pressure defined as systolic blood pressure 120 ≤systolic blood pressure <130 mm Hg and diastolic blood pressure <80 mm Hg. Stage

1 hypertension defined as 130 ≤systolic blood pressure <140 mm Hg or 80 ≤diastolic blood pressure <90 mm Hg. Stage 2 hypertension defined as systolic blood pressure ≥140 mm Hg or diastolic blood pressure ≥90 mm Hg and this group was stratified by use of anti-hypertensive medication in the last two weeks.

[b] Model adjusted for age, sex, race-center, baseline body mass index, baseline smoking status, baseline education level, baseline annual family income, baseline high density lipoprotein cholesterol level, baseline history of diabetes, baseline history of coronary heart disease, as well as their interaction with follow-up time.

[c] P-value for interaction race × elevated blood pressure × time = 0.16, race × stage 1 hypertension × time = 0.95, race × stage 2 hypertension no medication × time = 0.08, race × stage 2 hypertension with medication × time = 0.01.

[d] Model adjusted for age, sex, center, baseline body mass index, baseline smoking status, baseline education level, baseline annual family income, baseline high density lipoprotein cholesterol level, baseline history of diabetes, baseline history of coronary heart disease, as well as their interaction with follow-up time.

[e] African Americans in the Minnesota and Washington County cohorts was excluded in adjusted model because of small numbers.

BP, blood pressure; HTN, hypertension; MED, medication

Table 2-3 Average predicted probability of developing chronic kidney disease over 30 years follow-up according to hypertension status among participants based on the baseline covariates.

Stage of CKD (eGFR, mL/min/1.73 m^2)	Caucasian					African American[a]				
	Normal BP	Elevated BP	Stage 1 HTN	Stage 2 HTN without MED	Stage 2 HTN with MED	Normal BP	Elevated BP	Stage 1 HTN	Stage 2 HTN without MED	Stage 2 HTN with MED
G3a+ (<60)	54.4	61.6	64.7	78.1	70.9	55.4	62.8	60.9	76.1	66.6
G3b+ (<45)	22.5	27.8	30.0	41.8	35.0	35.2	41.6	40.1	52.4	44.5
G4+ (<30)	7.0	9.0	10.1	15.8	12.4	22.3	26.0	25.4	32.5	27.5

[a] African Americans in the Minnesota and Washington County cohorts were excluded in adjusted model because of small numbers.

BP, blood pressure; HTN, hypertension; MED, medication

Supplementary Table 2-1. Difference in eGFR decline (unit: ml/min per 1.73 m^2) per year according to blood pressure categories in the total population and by race.

	Differences (95% CIs) in eGFR decline (unit: ml/min per 1.73 m^2) per year			
	SBP <120 mm Hg and DBP <80 mm Hg	**120 ≤SBP <130 mm Hg and DBP <80 mm Hg**	**130 ≤SBP <140 mm Hg or 80 ≤DBP <90 mm Hg**	**SBP ≥140 mm Hg or DBP ≥90 mm Hg**
Total Population	**N=7,150[a]**	**N=2,063[a]**	**N=3,120[a]**	**N=2,521[a]**
Unadjusted	Ref	-0.26 (-0.32, -0.20)	-0.27 (-0.32, -0.22)	-0.66 (-0.71, -0.60)
Adjusted[b,c]	Ref	-0.15 (-0.20, -0.09)	-0.14 (-0.19, -0.09)	-0.42 (-0.48, -0.36)
Caucasian	**N=5,961**	**N=1,625**	**N=2,055**	**N=1,362**
Unadjusted	Ref	-0.19 (-0.24, -0.13)	-0.17 (-0.22, -0.12)	-0.45 (-0.51, -0.38)
Adjusted[c,d]	Ref	-0.12 (-0.18, -0.07)	-0.15 (-0.20, -0.10)	-0.37 (-0.43, -0.30)
African American	**N=1,189**	**N=438**	**N=1,065**	**N=1,159**
Unadjusted	Ref	-0.39 (-0.58, -0.20)	-0.20 (-0.34, -0.06)	-0.64 (-0.79, -0.49)
Adjusted[c,d,e]	Ref	-0.29 (-0.47, -0.10)	-0.18 (-0.31, -0.04)	-0.55 (-0.70, -0.40)

Note: Differences (95% confidence intervals) in eGFR decline (mL/min/1.73 m^2) per year, compared to SBP <120 mm Hg and DBP <80 mm Hg as reference.

Abbreviations: BP, blood pressure; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; MED, medication; Ref, reference; SBP, systolic blood pressure.

[a] The number of participants on antihypertensive MED of each blood pressure category is as follows: SBP <120 mm Hg and DBP <80 mm Hg, n=905; 120 ≤SBP <130 mm Hg and DBP <80 mm Hg, n=509; 130 ≤SBP <140 mm Hg or 80 ≤DBP <90 mm Hg, n=1,062; SBP ≥140 mm Hg or DBP ≥90 mm Hg, n=1,147.

[b] Model adjusted for age, sex, race-center, baseline body mass index, baseline smoking status, baseline education level, baseline total family income, antihypertensive MED status, baseline high density lipoprotein cholesterol level, baseline history of diabetes, baseline history of coronary heart disease, as well as their interaction with follow-up time.

[c] P-values for interaction between antihypertensive MED status and follow-up time are 0.99, 0.38, and 0.21 for total population, Caucasian, and African American respectively.

[d] Model adjusted for age, sex, center, baseline body mass index, baseline smoking status, baseline education level, baseline total family income, antihypertensive MED status, baseline high density lipoprotein cholesterol level, baseline history of diabetes, baseline history of coronary heart disease, as well as their interaction with follow-up time.

[e] African Americans in the Minnesota and Washington County cohorts was excluded in adjusted model because of small numbers.

Supplementary Table 2-2. Difference in eGFR decline (unit: ml/min per 1.73 m^2) per year according to hypertension categories in the total population and by race with imputing the eGFR from the 2728 form.

Population	Differences (95% CIs) in eGFR decline (unit: ml/min per 1.73 m^2) per year				
	Normal BP[a] N=6,200	Elevated BP[a] N=1,554	Stage 1 HTN[a] N=2,058	Stage 2 HTN w/o MED[a] N=1,374	Stage 2 HTN w/ MED[a] N=3,668
Unadjusted	Ref	-0.23 (-0.30, -0.16)	-0.25 (-0.31, -0.19)	-0.60 (-0.68, -0.53)	-0.51 (-0.56, -0.46)
Adjusted[b,c]	Ref	-0.13 (-0.20, -0.07)	-0.14 (-0.19, -0.08)	-0.38 (-0.45, -0.31)	-0.17 (-0.22, -0.12)
Caucasian	**N=5,341**	**N=1,281**	**N=1,448**	**N=801**	**N=2,132**
Unadjusted	Ref	-0.17 (-0.23, -0.11)	-0.15 (-0.20, -0.09)	-0.41 (-0.49, -0.33)	-0.31 (-0.36, -0.26)
Adjusted[d]	Ref	-0.12 (-0.18, -0.06)	-0.15 (-0.20, -0.09)	-0.36 (-0.44, -0.28)	-0.19 (-0.24, -0.13)
African American	**N=859**	**N=273**	**N=610**	**N=573**	**N=1,536**
Unadjusted	Ref	-0.39 (-0.63, -0.14)	-0.23 (-0.42, -0.05)	-0.64 (-0.84, -0.44)	-0.50 (-0.66, -0.35)
Adjusted[d,e]	Ref	-0.27 (-0.49, -0.04)	-0.16 (-0.33, 0.01)	-0.50 (-0.69, -0.31)	-0.17 (-0.31, -0.02)

Note: Differences (95% confidence intervals) in eGFR decline (mL/min/1.73 m^2) per year, compared to normal BP as reference.

Abbreviations: BP, blood pressure; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; HTN, hypertension; MED, medication; Ref, reference; SBP, systolic blood pressure.

[a] Normal BP defined as SBP < 120 mm Hg and DBP < 80 mm Hg. Elevated BP defined as 120 ≤ SBP < 130 mm Hg and DBP < 80 mm Hg. Stage 1 HTN defined as 130 ≤ SBP < 140 mm Hg or 80 ≤ DBP < 90 mm Hg. Stage 2 HTN defined as SBP ≥ 140 mm Hg or DBP ≥ 90 mm Hg, and this group was stratified by the use of antihypertensive MED in the last 2 weeks.

[b] Model adjusted for age, sex, race-center, baseline body mass index, baseline smoking status, baseline education level, baseline annual family income, baseline high density lipoprotein cholesterol level, baseline history of diabetes, baseline history of coronary heart disease, as well as their interaction with follow-up time.

[c] P-value for interaction race × elevated BP × time = 0.16, race × stage 1 HTN × time = 0.94, race × stage 2 HTN w/o MED × time = 0.08, race × stage 2 HTN w/ MED × time = 0.01.

[d] Model adjusted for age, sex, center, baseline body mass index, baseline smoking status, baseline education level, baseline annual family income, baseline high density lipoprotein cholesterol level, baseline history of diabetes, baseline history of coronary heart disease, as well as their interaction with follow-up time.

[e] African Americans in the Minnesota and Washington County cohorts was excluded in adjusted model because of small numbers.

Chapter 3 Association between Midlife Obesity and Kidney Function Trajectories: the Atherosclerosis Risk in Communities Study

Zhi Yu, BM, MS[1,2], Morgan E. Grams, MD, PhD[1,2,3], Chiadi E. Ndumele, MD, PhD[1,3], Lynne Wagenknecht, PhD[4], Eric Boerwinkle, PhD[5,6], Kari North, PhD[7], Casey M. Rebholz, PhD[1,2], Edward L. Giovannucci, MD, ScD[8,9,10], Josef Coresh, MD, PhD[1,2,3]

[1] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

[2] Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore, Maryland

[3] Department of Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland

[4] Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina

[5] Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas

[6] Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas

[7] Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

[8] Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts

[9] Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts

[10] Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, Massachusetts

ABSTRACT

Rationale & Objective: Obesity has been related to risk of chronic kidney disease (CKD). However, the associations of different measures of midlife obesity with long-term kidney function trajectories and whether they differ by sex and race is unknown.

Study Design: Observational study.

Setting & Participants: 13,496 participants from the ARIC study.

Predictors: Midlife obesity status as measured by body mass index (BMI), waist to hip ratio, and predicted percent fat at baseline.

Outcomes: Estimated glomerular filtration rate (eGFR) calculated using serum creatinine measured at 5 study visits and end-stage kidney disease (ESKD).

Analytical Approach: Mixed models with random intercepts and random slopes for eGFR. Cox proportional hazards models for ESKD.

Results: Baseline mean age was 54 years, median eGFR was 103 ml/min/1.73 m^2, and median BMI was 27 kg/m^2. Over 30 years of follow-up, midlife obesity measures were associated with eGFR decline in white and black women but not consistently in men. Adjusted for age, center, smoking and coronary heart disease (CHD), the differences (95% CI) in eGFR decline slope (unit: ml/min per 1.73 m^2 per decade) per standard deviation higher BMI, waist to hip ratio, and predicted percent fat were 0.09 (-0.18, 0.36), -0.25 (-0.50, 0.01) and -0.14 (-0.41, 0.13) for white men, -0.91 (-1.15, -0.67), -0.82 (-1.06, -0.58) and -1.02 (-1.26, -0.78) for white women, -0.70 (-1.54, 0.14), -1.60 (-2.42, -0.78) and -1.24 (-2.08, -0.40) for black men, and -1.24 (-2.08, -0.40), -1.50 (-2.05, -0.95) and -1.43 (-2.00, -0.86) for black women. Obesity indicators were independently associated with risk of ESKD for all sex-race groups except white men.

Limitations: Loss to follow-up during three decades of follow-up with five eGFR measurements.

Conclusions: Obesity status is a risk factor for future decline in kidney function and development of ESKD in black and white women with less consistent associations among men.

INTRODUCTION

Kidney function trajectories have long been used in the estimation of time to end-stage kidney disease (ESKD).[75] Recently, kidney function decline over time has been related not only to ESKD but also to all-cause mortality and cardiovascular disease risk.[76-78] Understanding risk factors for different patterns of kidney function trajectories is important so that individuals at risk for rapid progression may be targeted early for interventions. Successful interventions may prevent the development of disease among individuals with normal kidney function and slow the progression among those with kidney disease.

Obesity may be a targetable risk factor in the prevention of kidney function decline. Higher body-mass index has been associated with increased risk of incident chronic kidney disease (CKD), including greater kidney function decline among healthy, young adults.[79-81] However, BMI may not be the best marker of obesity-related risk, and associations may differ across sex and race.[82-86] Much less is known about the relationship between other obesity indicators, such as waist to hip ratio and the recently developed predicted percent fat,[87] and long-term kidney function decline.

This study evaluated the associations of midlife obesity with subsequent trajectories of estimated glomerular filtration rate (eGFR) and risks of developing ESKD across sex-race groups in a community-based cohort of 13,496 middle-aged white and black men and women over 30 years of follow-up. We examined several different obesity measures given controversy over the optimal method of estimating obesity, with the goal of defining the etiologic associations between obesity and kidney function decline.

METHODS

Study Design & Study Population

The Atherosclerosis Risk in Communities (ARIC) study is a prospective cohort study designed to evaluate risk factors for the development of cardiovascular disease.[29] It enrolled a total of 15,792 middle-aged (45-64 years old at baseline), predominantly white and black men and women from four communities in the U.S.: Forsyth County, North Carolina; Jackson, Mississippi; suburbs of Minneapolis, Minnesota; and Washington County, Maryland. The first examination took place in 1987-1989 (baseline, study visit 1), with follow-up examinations initially at approximately three-year intervals: 1990-1992 (study visit 2), 1993-1995 (study visit 3), 1996-1998 (study visit 4), and more recently, in 2011-2013 (study visit 5) and in 2016-2017 (study visit 6). During each study visit, an extensive questionnaire was administered, a clinical examination was conducted, and blood and urine specimens were collected.

In the present study, we excluded study participants who at baseline had ESKD (n=150), eGFR below 60 mL/min/1.73 m^2 (n=341), diabetes (n=1,797), or BMI below 18.5 kg/m^2 (n=143). Thus, the analytic sample size was 13,496 (85.5% of the original cohort). We excluded participants with prevalent diabetes since clinical diabetes can lead to intentional and unintentional weight loss. Study participants provided written documentation of informed consent and study protocols were approved by the institutional review board at each study site.

Assessment of Obesity Status

BMI was calculated as measured weight (in kilograms) divided by measured height (in meters) squared. Waist to hip ratio, which has been shown by some studies to be the more appropriate metric for obesity-related risk stratification among older adults,[88] was calculated from the measurements of the circumference at umbilical level (waist) and maximum buttocks (hip) to the nearest centimeter. Predicted percent fat was derived using sex-specific anthropometric prediction equations including information on age, race, weight, height, and waist circumference. The equations were reported to explain a large amount of the variation in percent fat (R^2 of 0.73 for men and 0.65 for women).[87] We evaluated predicted percent fat as an indicator of obesity because it was more strongly correlated with obesity-related biomarkers compared to BMI in previous studies.[87] The anthropometric prediction equations are:

Men: Percent fat (%) = 0.02 + 0.00*age (year) − 0.07*height (cm) − 0.08*weight (kg) + 0.48*WC (cm) + 0.32*Mexican + 0.02*Hispanic − 0.65*Black + 1.12*Other ethnicity

Women: Percent fat (%) = 50.46 + 0.07*age (year) − 0.26*height (cm) + 0.27*weight (kg) + 0.10*WC (cm) + 0.89*Mexican + 0.49*Hispanic − 1.57*Black + 0.43*Other ethnicity

We modeled all obesity measurements both as continuous variables and in tertiles.

Assessment of Kidney Function

Kidney function was assessed by measuring creatinine in serum or plasma specimens collected during each study visit, except for study visit 3. A modified kinetic Jaffe method was adapted to

measure creatinine level, and it was standardized to the National Institute of Standards and Technology standard and calibrated across study visits using repeated measurements from a sample of 200 ARIC study participants.[56-58] The Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation was used to calculate eGFR based on creatinine.[59] For those participants who developed incident ESKD (ascertained via linkage to the United States Renal Data System), an eGFR value of 15 mL/min/1.73 m^2 was imputed on the date of initiation of renal replacement therapy (transplant, dialysis).

Assessment of Other Variables

Demographic characteristics (date of birth for the calculation of age, sex, and race) and medical history (coronary heart disease (CHD)) were ascertained via questionnaire administered by trained interviewers at the baseline study visit. Systolic blood pressure (SBP) was measured three times using a random-zero sphygmomanometer. The average of the second and third measurements were used in the analysis. Study participants brought medications to the study visit and the names of all medications were transcribed. Blood samples that were collected from study participants during the baseline study visit were assayed for the measurement of concentrations of high-density lipoprotein cholesterol using an enzymatic method after precipitation with dextran sulfate-magnesium. Bioelectrical impedance (BIA), measured using the BIA 101-F device (Akern/RJL, Florence, Italy), was utilized to measure percent body fat and fat mass at visit 5.

Statistical Analysis

Baseline characteristics of the study population were compared by baseline BMI tertile, sex, and racial group using descriptive statistics and differences were tested using ANOVA for continuous variables and χ^2 tests for categorical variables. Spearman's correlations between BIA-measured percent fat at visit 5 and BMI, waist to hip ratio, or predicted percent fat at visit 5 were examined within each sex and racial group. Scatter plots of waist to hip ratio and predicted percent fat against BMI tertile at baseline were shown by sex and race. We estimated the differences in annual eGFR decline slope according to baseline obesity status tertiles[61]. Kernel density plots were used to illustrate the distribution of unadjusted annual predicted change in eGFR. Mixed models were used to evaluate the association between obesity status at baseline and eGFR trajectories using random intercepts and random slopes to account for individual variation in eGFR at baseline and its change. Cox proportional hazards models were used to estimate the association between baseline obesity status and ESKD. All models were stratified by sex and race (white / black) since the association of eGFR decline showed an interaction with baseline obesity across race-sex groups ($p<0.001$ for all obesity measurements) as well larger variance in blacks. Three models were constructed. Model 1 was unadjusted. Model 2 was adjusted for age (continuous), center (Minneapolis, Minnesota / Washington County, Maryland / Jackson, Mississippi / Forsyth County, North Carolina), current smoker (yes / no), and history of CHD (yes / no) at baseline. For Model 3, we further adjusted for hypertension medication use (yes / no), SBP (continuous), high-density lipoprotein cholesterol (HDL) (continuous), and eGFR (continuous; ESKD model only) at baseline, to assess the associations of obesity and kidney function decline independent of other obesity associated comorbidities. As socioeconomic status may be associated with both weight change and kidney function, we also additionally adjust for

family income (annual income ≥$25,000 / <$25,000 / not reported), education (high school graduated / not graduated) and tested their interactions with obesity measurements and time.

In sensitivity analyses, we categorized the baseline obesity measurements (BMI, waist to hip ratio, and predicted percent fat), into tertiles by sex and race and examined their associations with eGFR trajectories using the same mixed models as main analysis. We examined the associations of interest only among participants with valid information on visit 6 using the same methods to test the robustness of our main results. As smoking can lead to weight loss and modify the associations with obesity-related health conditions, we conducted a sensitivity analysis excluding current smokers.[89] As obesity is a risk factor for increased mortality,[90] we conducted a Fine-Gray competing risks analysis.[91] All analyses were conducted using R version 3.3.3 (R Development Core Team).

RESULTS

Baseline Characteristics

The baseline characteristics of the 13,496 study participants (10,222 white and 3,274 black) according to baseline BMI tertile, sex, and racial group are shown in **Table 3-1**. For all four sex and racial groups, participants with higher baseline BMI, particularly those in the highest tertile, were more likely to have higher waist to hip ratio, predicted percent fat, SBP, and a history of coronary heart disease. They were also more likely to take anti-hypertensive medication, have lower levels of HDL, and less likely to be current smokers ($p<0.001$ for all comparisons).

Correlations of Obesity Measures

Percent fat measured by BIA at visit 5 was strongly correlated with all concomitant obesity measures overall and in all four sex-race groups (all $p<0.001$, **Supplementary Figure 3-1**). However, the strength of the correlation and differences between sex-race groups varied substantially. BMI and predicted percent fat had the highest correlations with BIA-measured percent fat (range 0.62 to 0.85 across sex-race categories). In contrast, waist to hip ratio was has a weaker overall correlation of -0.15 (0.18 to 0.46 across sex-race categories). Predicted percent fat was the only measurement that captured the sex difference in obesity measurements resulting in convergence of regression lines for men and women. Density plots of distributions across baseline BMI tertiles showed that while waist to hip ratio and predicted percent fat increased with BMI, there was substantial overlap across tertiles indicating the measures would not classify the obesity status of individuals identically (**Supplementary Figure 3-2**)

Variation in Annual Change in eGFR by Race and Sex

Among women, annual decline in eGFR was more rapid among participants in the 2nd and 3rd tertiles of all three baseline obesity measurements; however, there was substantial overlap across categories. Black men in the 2nd and 3rd tertiles of waist to hip ratio and predicted percent fat also had more substantial decline in eGFR, but we did not observe such trend with BMI tertiles in this group (**Figure 3-1, 3-2, 3-3**). The median (25th, 75th percentile) annual eGFR declines in the low-tertile, mid-tertile, and high-tertile of baseline waist to hip ratio were 1.31 (1.10, 1.53), 1.37 (1.16, 1.61), and 1.32 (1.11, 1.52) for white men, 1.32 (1.14, 1.49), 1.43 (1.27, 1.59), and 1.51 (1.35, 1.67) for white women, 1.61 (1.41, 1.77), 1.69 (1.45, 1.85), and 1.90 (1.60, 2.07) for black

men, 1.76 (1.51, 2.02), 1.94 (1.67, 2.18), and 2.10 (1.86, 2.28) for black women, and that for baseline predicted percent fat were 1.29 (1.08, 1.51), 1.38 (1.16, 1.62), and 1.34 (1.14, 1.55) for white men, 1.32 (1.13, 1.49), 1.44 (1.26, 1.59), and 1.52 (1.36, 1.68) for white women, 1.62 (1.40, 1.82), 1.70 (1.43, 1.89), and 1.92 (1.66, 2.04) for black men, 1.81 (1.53, 2.07), 1.88 (1.62, 2.12), and 2.13 (1.88, 2.33) for black women.

Difference in eGFR decline by Markers of Obesity

In women, all obesity indicators were associated with eGFR decline over 30 years follow-up. In black men, only waist to hip ratio and predicted percent fat were associated with eGFR decline; in white men, no measures of obesity were associated with eGFR decline. Adjusted for age, center, smoking and CHD, the difference (95% confidence interval) in eGFR decline slope (unit: ml/min per 1.73 m^2 per decade) per SD higher baseline BMI, waist to hip ratio and predicted percent fat were 0.09 (-0.18, 0.36), -0.25 (-0.50, 0.01) and -0.14 (-0.41, 0.13) for white men, -0.91 (-1.15, -0.67), -0.82 (-1.06, -0.58) and -1.02 (-1.26, -0.78) for white women, -0.70 (-1.54, 0.14), -1.60 (-2.42, -0.78) and -1.24 (-2.08, -0.40) for black men, and -1.24 (-2.08, -0.40), -1.50 (-2.05, -0.95) and -1.43 (-2.00, -0.86) for black women (**Table 3-2**). None of the interactions between family income or education and obesity measurements with kidney function decline were significant after adjusting for multiple comparison. Results were similar to the main analysis in sensitivity analyses that examined the association between tertiles of obesity indicators and eGFR decline to check the impact of potential non-linearity (**Supplementary Table 3-1**), that only included participants who attended visit 6 (**Supplementary Table 3-2**), that excluded current smokers (**Supplementary Table 3-3**) , and that accounting for the competing risk of death before ESKD (**Supplementary Table 3-4**).

Risk of Developing ESKD

All obesity indicators were associated with increased risk of ESKD for all sex-race groups except among white men. Adjusted for age, center, smoking and CHD, the hazard ratios (95% confidence interval) for ESKD per SD of BMI, waist to hip ratio and predicted percent fat were 1.26 (0.98, 1.62), 1.12 (0.86, 1.48) and 1.19 (0.91, 1.54) for white men, 1.51 (1.14, 2.01), 1.79 (1.26, 2.53) and 1.72 (1.31, 2.26) for white women, 1.75 (1.29, 2.36), 1.99 (1.38, 2.87) and 1.86 (1.36, 2.55) for black men, and 1.68 (1.33, 2.13), 1.78 (1.32, 2.40) and 1.68 (1.32, 2.14) for black women. (**Table 3-3**). The 30-year difference in risk of ESKD across tertiles ranged from 0.8% to 5.8% (**Supplementary Figure 3-3**).

DISSCUSSION

In this community-based population of 13,496 middle-aged adults, we observed obesity status, measured by BMI, waist to hip ratio, and predicted percent fat, was generally associated with more rapid future decline in kidney function and higher risk of developing ESKD over 30-year of follow-up. Associations were observed in white and black women as well as black men; however, there was no evidence supporting associations between markers of obesity and eGFR decline or ESKD in white men. The more novel measure of obesity - predicted percent fat, a sex-specific equation that incorporates age, race, weight, height, and waist circumference - was highly correlated with BIA-measured percent fat. Our study suggests prevention of obesity in midlife may slow future rates of kidney function decline, at least in white and black women and black men. Documenting the full range of benefits of obesity prevention is important as obesity

prevention may require significant effort and cost which is justified by multiple health benefits. The stronger associations with ESKD suggest the overall benefit of obesity prevention will be greater for higher risk groups.

The current study adds to the existing literature by demonstrating that midlife obesity is a risk factor of kidney function decline and development of ESKD in later life in women and black men and by quantifying the mean rate and range of decrease in kidney function over 30 years of follow-up. Existing research on kidney function trajectories has been more focused on individuals with kidney diseases rather than those with preserved kidney function.[92,93] Our study addressed this gap by evaluating baseline obesity categories as a predictor of kidney function decline among individuals with preserved kidney function. We examined kidney function from midlife to older age where the prevalence of kidney disease is highest and ESKD most often occurs.

Our study demonstrated that the association of midlife obesity with decline in kidney function differed by race and sex. Measures of obesity were not associated with kidney function decline in white men, and BMI but not waist to hip ratio or predicted percent fat predicted kidney outcomes in black men. Differences by sex in susceptibility to kidney outcomes associated with obesity among middle-aged adults has been described in previous literature. For example, previous studies have suggested that higher BMI was a significant risk factor for the development of CKD in women, but not men, in a Japanese community cohort.[85] However, the underlying mechanisms for this difference are unknown. The lack of association in white men may be due to the

combination of lower obesity and lower kidney disease progression among whites reducing power. Alternatively, there may be greater variation in muscle mass as a non-GFR influence of creatinine among men compared to women.[94-98] Given the greater magnitude and range of measurements of obesity in women than in men, it is also possible that the power to observe associations between obesity and future kidney function decline was larger and indeed obesity explained more variance in rates of decline in women than in men.

In general, the ARIC cohort follows the expected trends from previous studies with weight gain dominating in mid-life and weight loss increasing at older age.[99-102] The mix of intentional and unintentional weight loss is unknown. Therefore, we focused on baseline weight to provide clear temporality and minimize reverse causation. Longitudinal tracking of obesity rank simplifies the interpretation of midlife obesity as a risk factor for kidney disease progression over the subsequent decades.[103]

Our study used three indicators to model obesity: BMI, waist to hip ratio, and predicted percent fat. BMI has been most widely used in clinical and public health settings; however, whether it is the most suitable measure for all scenarios has been debated.[82-84] Some studies have proposed that waist to hip ratio is more appropriate than BMI for gauging risk in middle to older-age adults since generalized obesity in older ages has been thought to provide protection against injury, nutritional reserve against illness, and better weight bearing bone formation.[104-107] Also, BMI has been criticized for not being able to discriminate individuals with different body composition of fat mass and lean body mass, which has been suggested as the reason for the

"obesity paradox", a phenomenon that overweight and obese individuals have better health outcomes compared with the normal-weight counterparts, in some settings.[108,109] In our study, we observed high sex/race specific correlations between BMI and BIA-measured percent fat. Waist to hip ratio was distinct from BMI and was relatively weakly correlated with BIA-measured percent fat, although their associations with kidney outcomes were similar. The novel obesity marker, predicted percent fat, appeared to be not only highly correlated with BIA-measured percent fat, but also a risk factor with considerable magnitude of both kidney function decline and risk of ESKD in women and black men. However, the associations between different indicators of obesity and kidney outcomes were in general similar. This is probably because our study population was generally healthy while BMI loses its value mostly at advanced disease stage when loss of lean mass is important.[109-111] Our results suggested that BMI is a good measure in a general population cohort for kidney outcomes and the improvement with more sophisticated measures in this setting is likely marginal.

There are several strengths to this study. The ARIC study is a large, prospective cohort with 30 years of follow-up. The long duration of the study allows for the characterization of kidney function decline in a population that was generally healthy at the outset. Given the inclusion of both whites and blacks as well as both men and women from four distinct U.S. communities, we were able to examine the association between obesity and kidney outcomes by sex and race. BMI was measured and not self-reported. Also, multiple established risk factors were collected in a standardized manner according to a research protocol. The main limitation of this study is that there were only up to five eGFR measurements for the estimation of long-term trajectories. That said, there are few longitudinal, population-based cohorts that have had more frequent

measurements of eGFR over 30 years. Participants who develop ESKD were less likely to survive to attend subsequent study visits; we included an estimate of their trajectory by imputing an eGFR value of 15 mL/min/1.73 m^2 at the time of ESKD onset. The potential differential loss to follow-up may also occur for people in higher tertiles of obesity at baseline.

In conclusion, we observed in community-dwelling adults that midlife obesity status was a risk factor for future decline in kidney function and development of ESKD in all sex-race subgroups except for white men. The lack of associations in white men suggests that the role of obesity and its optimal quantification for kidney disease risk requires further study.

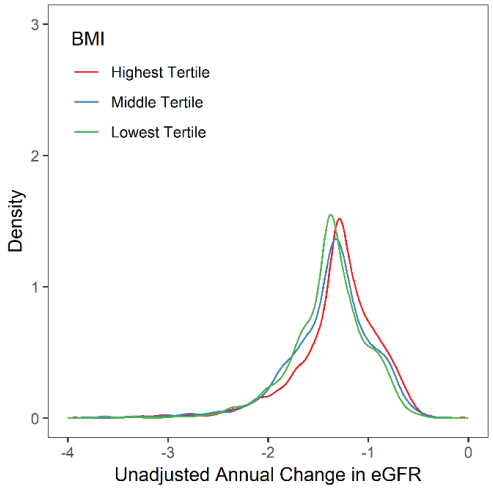
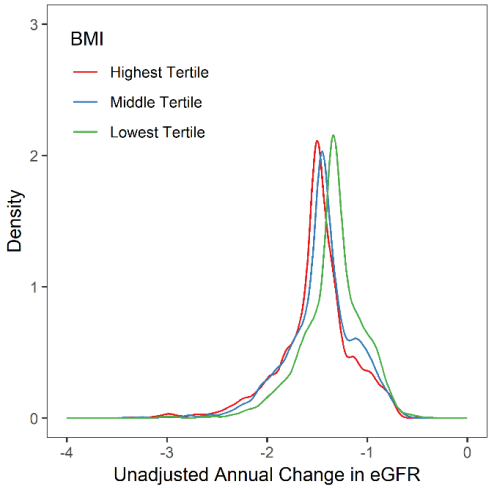
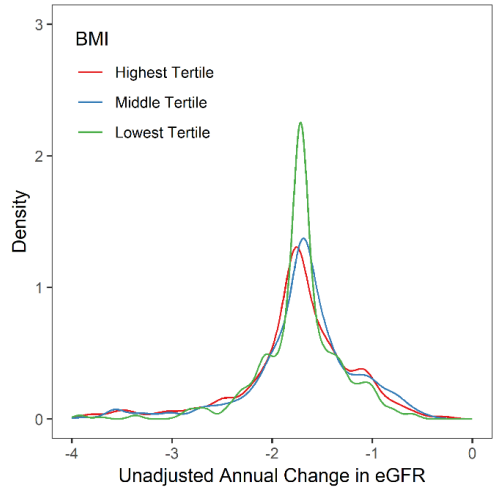
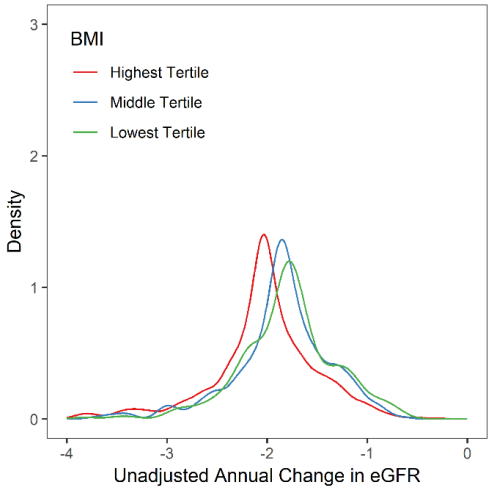
	Male	Female
White	 **A**	 **B**
Black	 **C**	 **D**

Figure 3-1 Distribution of unadjusted predicted average annual change in estimated glomerular filtration rate (eGFR; unit: ml/min per 1.73 m^2) within the ARIC population according to baseline body mass index (BMI) tertile by sex and race. (A) Unadjusted eGFR change among white men (B) Unadjusted eGFR change among white women (C) Unadjusted eGFR change among black men (D) Unadjusted eGFR change among black women.

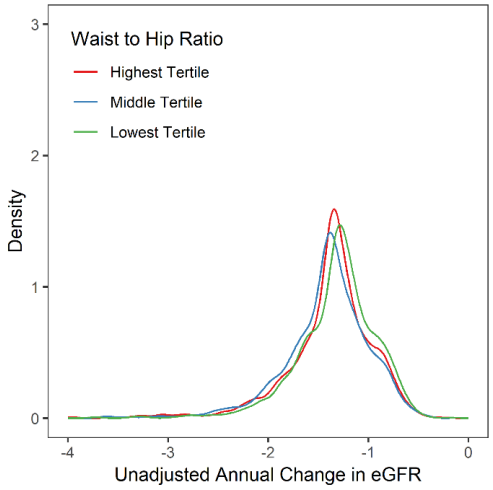
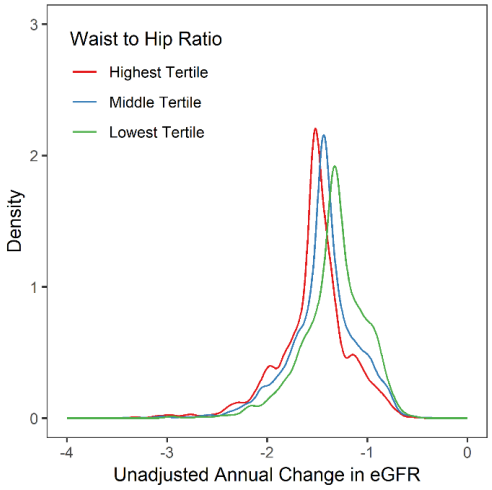
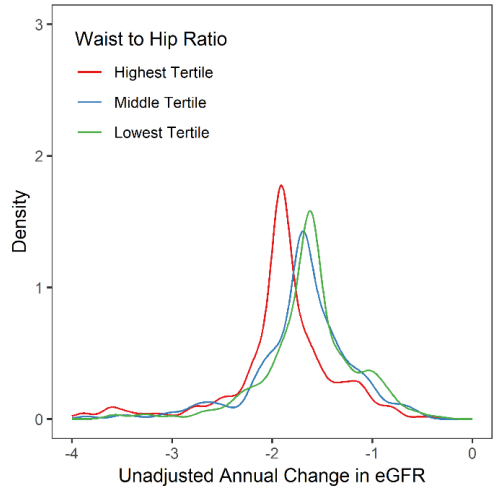
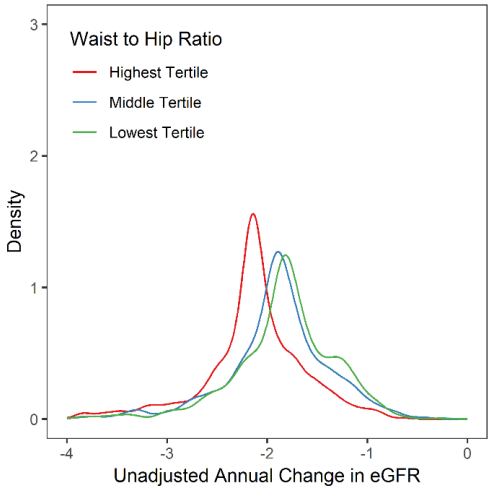
	Male	Female
White	 **A**	 **B**
Black	 **C**	 **D**

Figure 3-2 Distribution of unadjusted predicted average annual change in estimated glomerular filtration rate (eGFR; unit: ml/min per 1.73 m^2) within the ARIC population according to waist to hip ratio tertile by sex and race. (A) Unadjusted eGFR change among white men (B) Unadjusted eGFR change among white women (C) Unadjusted eGFR change among black men (D) Unadjusted eGFR change among black women.

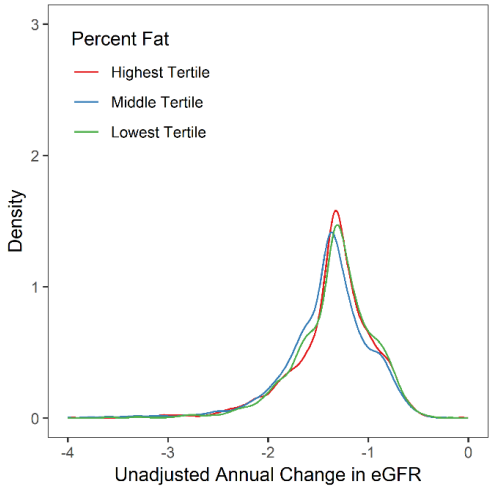
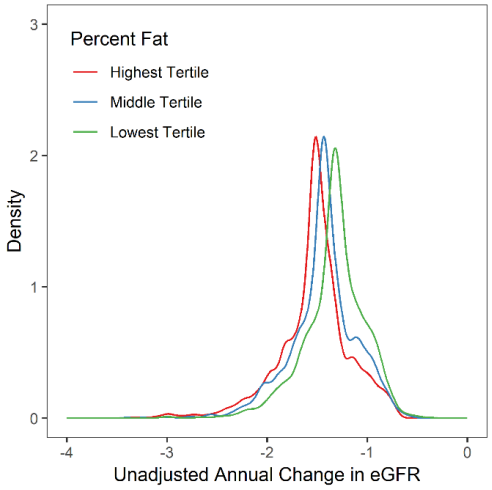
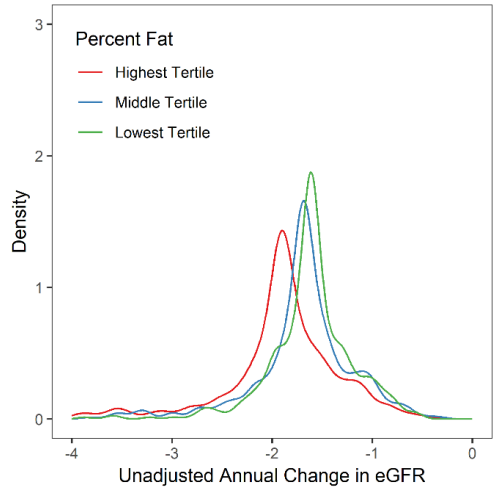
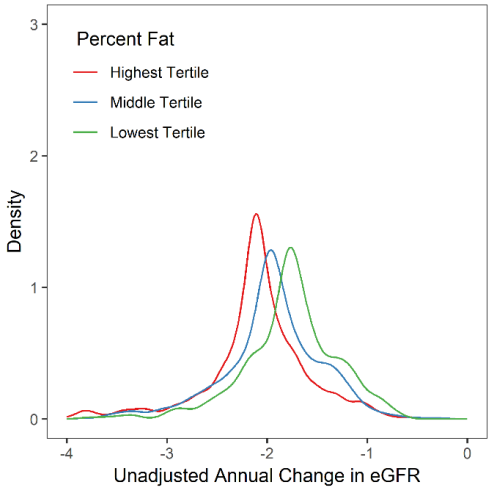
	Male	Female
White	 **A**	 **B**
Black	 **C**	 **D**

Figure 3-3 Distribution of unadjusted predicted average annual change in estimated glomerular filtration rate (eGFR; unit: ml/min per 1.73 m^2) within the ARIC population according to baseline predicted percent fat tertile by sex and race. (A) Unadjusted eGFR change among white men (B) Unadjusted eGFR change among white women (C) Unadjusted eGFR change among black men (D) Unadjusted eGFR change among black women.

Table 3-1 Characteristics of the study population according to baseline body mass index (BMI) tertiles by sex and race (N=13,496 in 1987-1989).

Characteristic[a]	White (N=1,0222)						Black (N=3,274)					
	Men (N=4,802)			Women (N=5,420)			Men (N=1,270)			Women (N=2,004)		
	Low-tertile (N=1,601)	Mid-tertile (N=1,600)	High-tertile (N=1,601)	Low-tertile (N=1,809)	Mid-tertile (N=1,803)	High-tertile (N=1,808)	Low-tertile (N=423)	Mid-tertile (N=424)	High-tertile (N=423)	-Low-tertile (N=668)	Mid-tertile (N=668)	High-tertile (N=668)
BMI, kg/m^2	23.5	26.8	31.4	21.6	25.3	32.2	22.7	26.9	32.2	24.2	29.3	37.4
	(1.5)	(0.8)	(3.1)	(1.2)	(1.1)	(4.2)	(1.7)	(1.1)	(3.5)	(2.0)	(1.4)	(5.2)
WHR[b]	0.93	0.97	1.00	0.84	0.88	0.93	0.90	0.93	0.97	0.85	0.90	0.94
	(0.05)	(0.04)	(0.04)	(0.07)	(0.07)	(0.07)	(0.05)	(0.04)	(0.05)	(0.07)	(0.07)	(0.07)
Predicted percent	25.1	28.2	32.2	35.4	39.1	45.5	22.3	25.9	30.8	36.1	40.9	48.5
fat, %[b]	(2.3)	(2.0)	(3.2)	(1.5)	(1.4)	(3.9)	(2.6)	(2.3)	(4.1)	(2.1)	(1.6)	(4.8)
eGFR, mL/min/1.73	99.5	97.8	97.0	103.0	101.2	100.4	112.3	107.9	105.4	115.4	114.8	115.4
m^{2}[b]	(11.1)	(11.0)	(11.6)	(10.3)	(11.4)	(12.0)	(15.5)	(15.4)	(16.7)	(16.3)	(15.9)	(16.2)
Age, years[b]	55.3	55.1	54.9	53.9	54.5	54.5	54.6	53.8	53.6	53.3	53.4	53.2
	(5.7)	(5.7)	(5.6)	(5.7)	(5.6)	(5.7)	(6.0)	(5.9)	(5.7)	(5.8)	(5.6)	(5.7)
Current smoker[b]	510	358	325	558	438	339	217	157	113	226	158	126
	(31.9)	(22.4)	(20.3)	(30.9)	(24.3)	(18.8)	(51.3)	(37.1)	(26.7)	(33.9)	(23.7)	(18.9)
HTN MED use[b]	203	265	380	169	283	471	89	127	166	199	255	310
	(12.7)	(16.7)	(23.9)	(9.4)	(15.7)	(26.2)	(21.1)	(30.1)	(39.4)	(29.9)	(38.4)	(46.8)

SBP, mm/Hg[b]	116.6	119.2	122.0	111.6	115.7	121.3	127.7	129.4	130.7	123.3	125.2	129.9
	(15.6)	(15.6)	(15.2)	(16.3)	(17.1)	(16.8)	(22.6)	(21.6)	(20.5)	(20.8)	(18.4)	(20.3)
HDL-C, mg/L[b]	47.1	42.6	39.9	64.8	58.6	51.7	57.6	50.0	46.1	63.9	59.0	55.3
	(13.9)	(11.5)	(10.2)	(17.6)	(16.3)	(14.0)	(20.3)	(14.2)	(12.6)	(18.7)	(17.3)	(14.2)
Prevalent CHD[b]	123	120	132	17	27	35	15	17	22	14	12	12
	(7.9)	(7.6)	(8.4)	(1.0)	(1.5)	(2.0)	(3.6)	(4.1)	(5.2)	(2.2)	(1.8)	(1.8)
High school graduate[b]	1329	1316	1321	1625	1553	1416	225	253	251	487	411	375
	(83.2)	(82.4)	(82.6)	(89.9)	(86.1)	(78.4)	(53.4)	(60.0)	(59.6)	(73.0)	(61.6)	(56.2)
Annual family income	335	284	308	423	496	612	253	217	209	384	440	484
<$25,000[b]	(20.9)	(17.8)	(19.2)	(23.4)	(27.5)	(33.8)	(59.8)	(51.2)	(49.4)	(57.5)	(65.9)	(72.5)

[a] Mean (standard deviation) for continuous variables and % (n) for categorical variables.

[b] P-value for comparing the BMI groups <0.001; P-value calculated by ANOVA for continuous variables and χ^2 test for categorical variables

[c] The cut-off points between the low- and the mid-tertiles and between mid- and high-tertiles are 25.4 and 28.3 for white men, 23.4 and 27.5 for white women, 25.1 and 28.8 for black men, and 27.0 and 32.0 for black women.

CHD, coronary heart disease; eGFR, estimated glomerular filtration rate; HDL-C, high density lipoprotein cholesterol**;** HTN, hypertension; MED, medication; SBP, systolic blood pressure; WHR: waist to hip ratio

Table 3-2 Association of estimated glomerular filtration rate (eGFR) decline slope (unit: ml/min per 1.73 m^2 per decade) with three measures of baseline obesity by sex and race.

	White (N=1,0222)		Black (N=3,274)	
	Men (N=4,802)	**Women (N=5,420)**	**Men (N=1,270)**	**Women (N=2,004)**
	Body mass index[c.d], per standard deviation			
Mean (SD), kg/m^2	27.23 (3.86)	26.33 (5.09)	27.28 (4.53)	30.29 (6.36)
Model 1[a,b]	0.21 (-0.06, 0.48)[e]	-0.76 (-1.00, -0.52)***	-0.50 (-1.34, 0.34)	-1.22 (-1.79, -0.65)***
Model 2[a,b]	0.09 (-0.18, 0.36)	-0.91 (-1.15, -0.67)***	-0.70 (-1.54, 0.14)	-1.35 (-1.92, -0.78)***
Model 3[a,b]	0.52 (0.23, 0.80)***	-0.27 (-0.53, -0.02)*	0.03 (-0.85, 0.91)	-0.85 (-1.45, -0.25)**
	Waist to hip ratio[c.d], per standard deviation			
Mean (SD)	0.97 (0.05)	0.89 (0.08)	0.94 (0.05)	0.90 (0.08)
Model 1[a,b]	-0.28 (-0.55, -0.01)*	-0.87 (-1.09, -0.65)***	-1.58 (-2.4, -0.76)***	-1.54 (-2.09, -0.99)***
Model 2[a,b]	-0.25 (-0.50, 0.005)	-0.82 (-1.06, -0.58)***	-1.60 (-2.42, -0.78)***	-1.50 (-2.05, -0.95)***
Model 3[a,b]	0.15 (-0.13, 0.43)	-0.16 (-0.41, 0.09)	-0.40 (-1.32, 0.52)	-1.07 (-1.68, -0.46)***
	Predicted percent fat[c.d], per standard deviation			
Mean (SD), %	28.49 (3.88)	39.99 (4.92)	26.31 (4.65)	41.83 (5.99)

Model 1[a,b]	-0.12 (-0.39, 0.15)	-0.89 (-1.11, -0.67)***	-1.08 (-1.92, -0.24)**	-1.32 (-1.89, -0.75)***
Model 2[a,b]	-0.14 (-0.41, 0.13)	-1.02 (-1.26, -0.78)***	-1.24 (-2.08, -0.40)**	-1.43 (-2.00, -0.86)***
Model 3[a,b]	0.30 (0.02, 0.59)*	-0.32 (-0.58, -0.06)*	-0.26 (-1.17, 0.65)	-0.90 (-1.5, -0.29)**

[a] Model 1 is unadjusted random effects model showing rate of decline (the interaction of each variable with follow-up time); model 2 adjusted for age (continuous), center (categorical), current smoker (yes / no), and prevalent coronary heart disease (yes / no) at baseline; model 3 additionally adjusted for hypertension medication (yes / no), systolic blood pressure (continuous), total cholesterol (continuous), high-density lipoprotein cholesterol (continuous), triglyceride (continuous, log transformed), education level (high school graduated / not graduated), and annual family income (categorical) at baseline.

[b] Black participants in the Minnesota and Washington County centers were excluded in the model because of small numbers.

[c] Centered at median of each race-gender group.

[d] In an overall unadjusted model combining sex-race groups, the p-value for interaction between sex-race and the association of eGFR decline with baseline obesity was <0.001 for all obesity measurements.

[e] Estimate (95% confidence interval) for all such values.

* indicates p<0.05, ** indicates p<0.01, *** indicates p<0.001.

Table 3-3 Hazard ratios for end-stage kidney disease (ESKD) according to baseline obesity status by sex and race.

	White (n/N=95/1,0222)		Black (N=95/3,274)	
	Men (n/N=57/4,802)	**Women (n/N=43/5,420)**	**Men (n/N=39/1,270)**	**Women (n/N=56/2,004)**
	Body mass index[c,d]**,** per standard deviation			
Mean (SD), kg/m²	27.23 (3.86)	26.33 (5.09)	27.28 (4.53)	30.29 (6.36)
Model 1[a,b]	1.22 (0.95, 1.57)[d]	1.47 (1.13, 1.92)**	1.66 (1.23, 2.25)***	1.63 (1.29, 2.06)***
Model 2[a,b]	1.26 (0.98, 1.62)	1.51 (1.14, 2.01)**	1.75 (1.29, 2.36)***	1.68 (1.33, 2.13)***
Model 3[a,b]	1.04 (0.79, 1.39)	1.06 (0.76, 1.46)	1.51 (1.09, 2.11)*	1.65 (1.29, 2.13)***
	Waist to hip ratio[c]**,** per standard deviation			
Mean (SD)	0.97 (0.05)	0.89 (0.08)	0.94 (0.05)	0.90 (0.08)
Model 1[a,b]	1.19 (0.91, 1.56)	2.11 (1.51, 2.94)***	2.00 (1.40, 2.88)***	1.77 (1.31, 2.37)***
Model 2[a,b]	1.12 (0.86, 1.48)	1.79 (1.26, 2.53)**	1.99 (1.38, 2.87)***	1.78 (1.32, 2.40)***
Model 3[a,b]	0.92 (0.69, 1.23)	1.26 (0.86, 1.84)	1.73 (1.18, 2.53)**	1.74 (1.26, 2.40)***
	Predicted percent fat[c]**,** per standard deviation			
Mean (SD), %	28.49 (3.88)	39.99 (4.92)	26.31 (4.65)	41.83 (5.99)
Model 1[a,b]	1.19 (0.92, 1.55)	1.74 (1.35, 2.24)***	1.84 (1.34, 2.52)***	1.64 (1.29, 2.08)***

Model 2[a,b]	1.19 (0.91, 1.54)	1.72 (1.31, 2.26)***	1.86 (1.36, 2.55)***	1.68 (1.32, 2.14)***
Model 3[a,b]	0.94 (0.71, 1.25)	1.27 (0.92, 1.74)	1.63 (1.16, 2.30)**	1.61 (1.24, 2.09)***

[a] Model 1 is unadjusted random effects model showing rate of decline (the interaction of each variable with follow-up time); model 2 adjusted for age (continuous), center (categorical), current smoker (yes / no), and prevalent coronary heart disease (yes / no) at baseline; model 3 additionally adjusted for hypertension medication (yes / no), systolic blood pressure (continuous), total cholesterol (continuous), high-density lipoprotein cholesterol (continuous), triglyceride (continuous, log transformed), estimated glomerular filtration rate (continuous), education level (high school graduated / not graduated), and annual family income (categorical) at baseline.

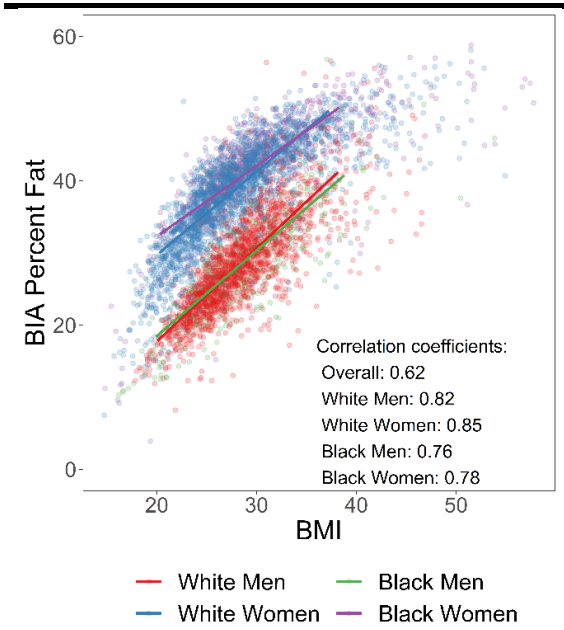
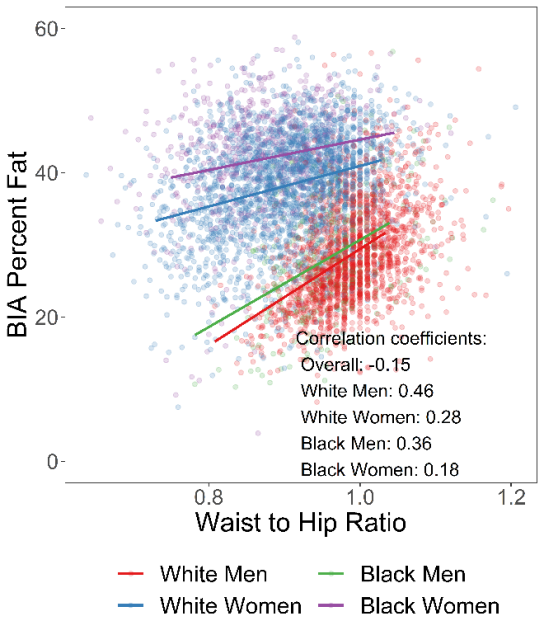
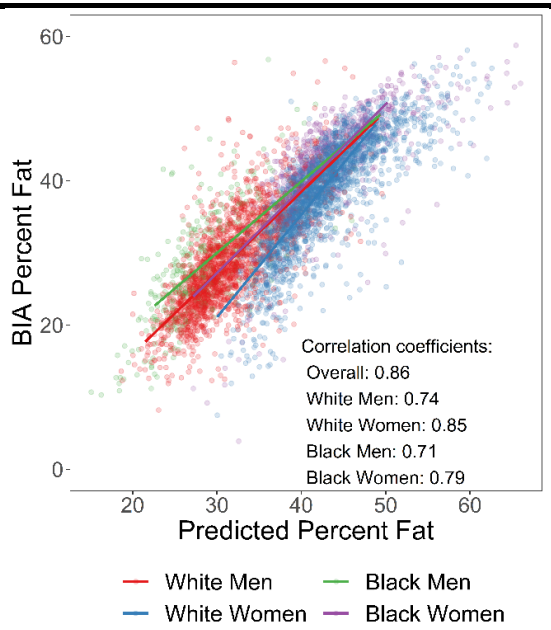
[b] Black participants in the Minnesota and Washington County centers were excluded in the model because of small numbers.

[c] Centered at median of each race-gender group.

[d] Hazard ratio (95% confidence interval) for all such values.

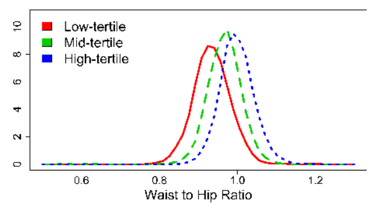
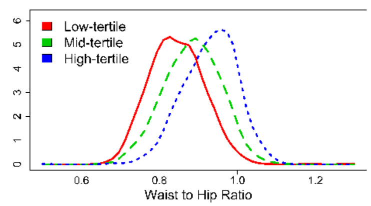
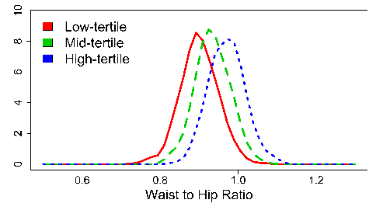
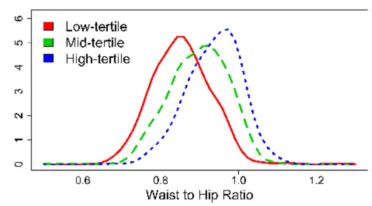
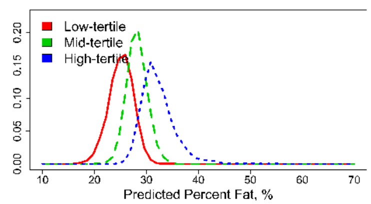
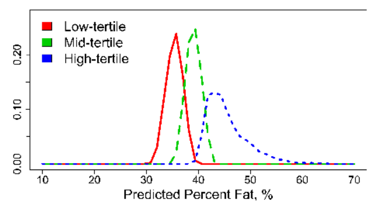
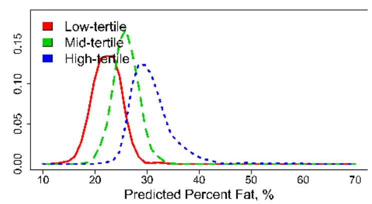
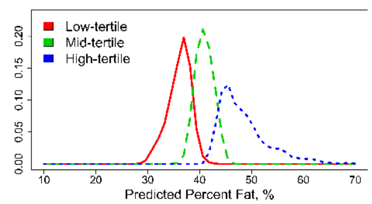
* indicates $p<0.05$, ** indicates $p<0.01$, *** indicates $p<0.001$.

Supplementary Figure 3-1 Scatter plot of bioelectrical impedance analysis (BIA) measured percent fat and body mass index (BMI), waist to hip ratio, or predicted percent fat at visit 5 with generalized additive regression line.

BIA percent fat and BMI	BIA percent fat and waist to hip ratio	BIA percent fat and predicted percent fat
		

All p-values are <0.001

Supplementary Figure 3-2 Distributions of waist to hip ratio and predicted percent fat according to baseline body mass index (BMI) tertile by sex and race.

	White (N=1,0222)		Black (N=3,274)	
	Men (N=4,802)	Women (N=5,420)	Men (N=1,270)	Women (N=2,004)
Waist to hip ratio				
Predicted percent fat				

Supplementary Table 3-1 Difference in estimated glomerular filtration rate (eGFR) decline slope (unit: ml/min per 1.73 m^2 per decade) according to baseline obesity status tertile by sex and race.

	White (N=1,0222)						Black (N=3,274)					
	Men (N=4,802)			Women (N=5,420)			Men (N=1,270)			Women (N=2,004)		
	Low-tertile (N=1,601)	Mid-tertile (N=1,600)	High-tertile (N=1,601)	Low-tertile (N=1,809)	Mid-tertile (N=1,803)	High-tertile (N=1,808)	Low-tertile (N=423)	Mid-tertile (N=424)	High-tertile (N=423)	Low-tertile (N=668)	Mid-tertile (N=668)	High-tertile (N=668)
						BMI						
Model 1[a,b]	Ref	0.20	0.85*	Ref	-0.92***	-1.61***	Ref	0.69	-0.50	Ref	-1.57*	-2.90***
Model 2[a,b]	Ref	0.01	0.56	Ref	-1.09***	-1.93***	Ref	0.45	-0.98	Ref	-1.70*	-3.19***
Model 3[a,b]	Ref	0.56	1.46***	Ref	-0.50	-0.52	Ref	0.82	0.45	Ref	-1.06	-1.92*
						Waist to hip ratio						
Model 1[a,b]	Ref	-0.96**	-0.62	Ref	-1.10***	-1.99***	Ref	-0.78	-2.98**	Ref	-0.68	-3.21***
Model 2[a,b]	Ref	-0.92**	-0.57	Ref	-1.02***	-1.89***	Ref	-0.83	-3.00**	Ref	-0.59	-3.13***
Model 3[a,b]	Ref	-0.36	0.28	Ref	-0.28	-0.40	Ref	0.78	-0.07	Ref	-0.22	-1.86*
						Predicted percent fat						
Model 1[a,b]	Ref	-0.68*	-0.17	Ref	-1.12***	-1.98***	Ref	-0.81	-3.02**	Ref	-1.83**	-3.35***
Model 2[a,b]	Ref	-0.69*	-0.19	Ref	-1.25***	-2.22***	Ref	-1.10	-3.36**	Ref	-1.91**	-3.61***
Model 3[a,b]	Ref	-0.28	0.72*	Ref	-0.50	-0.68*	Ref	0.21	-0.95	Ref	-1.36	-2.22**

[a] Model 1 is unadjusted random effects model showing rate of decline (the interaction of each variable with follow-up time); model 2 adjusted for age (continuous), center (categorical), current smoker (yes / no), and prevalent coronary heart disease (yes / no) at baseline; model 3 additionally adjusted for hypertension medication (yes / no), systolic blood pressure (continuous), total cholesterol (continuous), high-density lipoprotein cholesterol (continuous), triglyceride (continuous, log transformed), education level (high school graduated / not graduated), and annual family income (categorical) at baseline.

[b] Black participants in the Minnesota and Washington County centers were excluded in the model because of small numbers.

* indicates $p<0.05$, ** indicates $p<0.01$, *** indicates $p<0.001$.

Supplementary Table 3-2 Difference in estimated glomerular filtration rate (eGFR) decline slope (unit: ml/min per 1.73 m^2 per decade) according to baseline obesity status tertile by sex and race among participants who had visit 6 information.

	White (N=2,802)						Black (N=799)					
	Men (N=1,237)			Women (N=1,565)			Men (N=253)			Women (N=546)		
	Low-tertile (N=430)	Mid-tertile (N=412)	High-tertile (N=395)	Low-tertile (N=575)	Mid-tertile (N=564)	High-tertile (N=426)	Low-tertile (N=73)	Mid-tertile (N=102)	High-tertile (N=78)	Low-tertile (N=202)	Mid-tertile (N=186)	High-tertile (N=158)
Attended visit 6, %	26%	25%	24%	31%	31%	23%	17%	23%	16%	29%	26%	21%
						BMI						
Model 1[a,b]	Ref	0.17	0.62	Ref	-1.00**	-1.22***	Ref	0.46	0.15	Ref	-1.11	-2.99***
Model 2[a,b]	Ref	0.07	0.42	Ref	-1.20***	-1.57***	Ref	0.41	-0.14	Ref	-1.14	-3.02***
Model 3[a,b]	Ref	0.45	0.98**	Ref	-0.76*	-0.53	Ref	0.08	-0.59	Ref	-1.10	-2.67**
						Waist to hip ratio						
Model 1[a,b]	Ref	-0.98**	-0.04	Ref	-1.03***	-1.38***	Ref	-0.44	-0.62	Ref	-0.95	-2.16**
Model 2[a,b]	Ref	-1.01**	-0.08	Ref	-1.02***	-1.40***	Ref	-0.45	-0.55	Ref	-0.91	-2.11**
Model 3[a,b]	Ref	-0.68*	0.30	Ref	-0.41	-0.44	Ref	-0.37	-0.44	Ref	-0.92	-1.76*
						Predicted percent fat						
Model 1[a,b]	Ref	-0.36	0.08	Ref	-0.91**	-1.39***	Ref	-0.63	-1.03	Ref	-1.77*	-3.50***

Model 2[a,b]	Ref	-0.43	0.10	Ref	-1.05***	-1.65***	Ref	-0.85	-1.09	Ref	-1.81**	-3.55***
Model 3[a,b]	Ref	-0.16	0.66	Ref	-0.53	-0.56	Ref	-1.02	-1.37	Ref	-1.98**	-3.32***

[a] Model 1 is unadjusted random effects model showing rate of decline (the interaction of each variable with follow-up time); model 2 adjusted for age (continuous), center (categorical), current smoker (yes / no), and prevalent coronary heart disease (yes / no) at baseline; model 3 additionally adjusted for hypertension medication (yes / no), systolic blood pressure (continuous), total cholesterol (continuous), high-density lipoprotein cholesterol (continuous), triglyceride (continuous, log transformed), education level (high school graduated / not graduated), and annual family income (categorical) at baseline.

[b] Black participants in the Minnesota and Washington County centers were excluded in the model because of small numbers.

* indicates $p<0.05$, ** indicates $p<0.01$, *** indicates $p<0.001$.

Supplementary Table 3-3 Association of estimated glomerular filtration rate (eGFR) decline slope (unit: ml/min per 1.73 m^2 per decade) with three measures of baseline obesity by sex and race excluding current smokers.

	White (N=1,0222)		Black (N=3,274)	
	Men (N=3,608)	**Women (N=4,081)**	**Men (N=782)**	**Women (N=1,490)**
	BMI[c], per standard deviation			
Model 1[a,b]	0.01 (-0.28, 0.29)[d]	-0.76 (-1.00, -0.52)***	-0.35 (-1.30, 0.60)	-1.59 (-2.19, -0.99)***
Model 2[a,b]	-0.05 (-0.33, 0.23)	-0.84 (-1.08, -0.59)***	-0.34 (-1.28, 0.61)	-1.60 (-2.20, -1.00)***
Model 3[a,b]	0.31 (0.02, 0.61)*	-0.27 (-0.53, 0.001)	0.11 (-0.87, 1.09)	-1.22 (-1.86, -0.58)***
	Waist to hip ratio[c], per standard deviation			
Model 1[a,b]	-0.36 (-0.63, -0.09)**	-0.76 (-0.99, -0.53)***	-1.18 (-2.11, -0.25)*	-1.47 (-2.07, -0.88)***
Model 2[a,b]	-0.35 (-0.62, -0.07)*	-0.70 (-0.94, -0.46)***	-1.17 (-2.10, -0.24)*	-1.47 (-2.06, -0.87)***
Model 3[a,b]	-0.07 (-0.35, 0.22)	-0.12 (-0.39, 0.14)	-0.51 (-1.52, 0.51)	-1.24 (-1.89, -0.59)***
	Predicted percent fat[c], per standard deviation			
Model 1[a,b]	-0.28 (-0.56, 0.003)	-0.87 (-1.11, -0.64)***	-0.85 (-1.81, 0.11)	-1.66 (-2.26, -1.07)***
Model 2[a,b]	-0.27 (-0.55, 0.01)	-0.92 (-1.17, -0.68)***	-0.83 (-1.79, 0.13)	-1.67 (-2.27, -1.07)***
Model 3[a,b]	0.07 (-0.22, 0.37)	-0.31 (-0.58, -0.04)*	-0.30 (-1.32, 0.72)	-1.29 (-1.94, -0.65)***

[a] Model 1 is unadjusted random effects model showing rate of decline (the interaction of each variable with follow-up time); model 2 adjusted for age (continuous), center (categorical), and prevalent coronary heart disease (yes / no) at baseline; model 3 additionally adjusted for hypertension medication (yes / no), systolic blood pressure (continuous), total cholesterol (continuous), high-density lipoprotein cholesterol (continuous), triglyceride (continuous, log transformed), education level (high school graduated / not graduated), and annual family income (categorical) at baseline.

[b] Black participants in the Minnesota and Washington County centers were excluded in the model because of small numbers.

[c] Centered at median of each race-gender group.

[d] Estimate (95% confidence interval) for all such values.

* indicates $p<0.05$, ** indicates $p<0.01$, *** indicates $p<0.001$.

Supplementary Table 3-4 Subhazard ratios for end-stage kidney disease (ESKD) according to baseline obesity status by sex and race with using a Fine-Gray competing risk model.

	White (n/N=95/1,0222)		Black (N=95/3,274)	
	Men (n/N=57/4,802)	**Women (n/N=43/5,420)**	**Men (n/N=39/1,270)**	**Women (n/N=56/2,004)**
	Body mass index[c.d]**,** per standard deviation			
Mean (SD), kg/m^2	27.23 (3.86)	26.33 (5.09)	27.28 (4.53)	30.29 (6.36)
Model 1[a,b]	1.20 (0.95, 1.52)	1.45 (1.19, 1.78)***	1.71 (1.3, 2.24)***	1.58 (1.25, 1.99)***
Model 2[a,b]	1.21 (0.96, 1.54)	1.46 (1.17, 1.83)***	1.73 (1.31, 2.28)***	1.61 (1.28, 2.02)***
Model 3[a,b]	1.05 (0.82, 1.35)	1.05 (0.81, 1.36)	1.51 (1.07, 2.13)*	1.58 (1.24, 2.03)***
	Waist to hip ratio[c]**,** per standard deviation			
Mean (SD)	0.97 (0.05)	0.89 (0.08)	0.94 (0.05)	0.90 (0.08)
Model 1[a,b]	1.11 (0.85, 1.44)	1.99 (1.43, 2.78)***	1.68 (1.16, 2.43)**	1.62 (1.23, 2.12)***
Model 2[a,b]	1.05 (0.80, 1.37)	1.70 (1.19, 2.43)**	1.78 (1.19, 2.65)**	1.70 (1.29, 2.24)***
Model 3[a,b]	0.91 (0.69, 1.19)	1.27 (0.87, 1.86)	1.5 (0.97, 2.31)	1.67 (1.21, 2.3)**
	Predicted percent fat[c]**,** per standard deviation			
Mean (SD), %	28.49 (3.88)	39.99 (4.92)	26.31 (4.65)	41.83 (5.99)

Model 1[a,b]	1.13 (0.86, 1.48)	1.71 (1.37, 2.13)***	1.71 (1.27, 2.3)***	1.55 (1.25, 1.92)***
Model 2[a,b]	1.11 (0.85, 1.46)	1.67 (1.30, 2.15)***	1.74 (1.30, 2.34)***	1.59 (1.29, 1.96)***
Model 3[a,b]	0.96 (0.73, 1.25)	1.27 (0.92, 1.75)	1.44 (1.02, 2.03)*	1.53 (1.21, 1.93)***

[a] Model 1 is unadjusted random effects model showing rate of decline (the interaction of each variable with follow-up time); model 2 adjusted for age (continuous), center (categorical), current smoker (yes / no), and prevalent coronary heart disease (yes / no) at baseline; model 3 additionally adjusted for hypertension medication (yes / no), systolic blood pressure (continuous), total cholesterol (continuous), high-density lipoprotein cholesterol (continuous), triglyceride (continuous, log transformed), estimated glomerular filtration rate (continuous), education level (high school graduated / not graduated), and annual family income (categorical) at baseline.

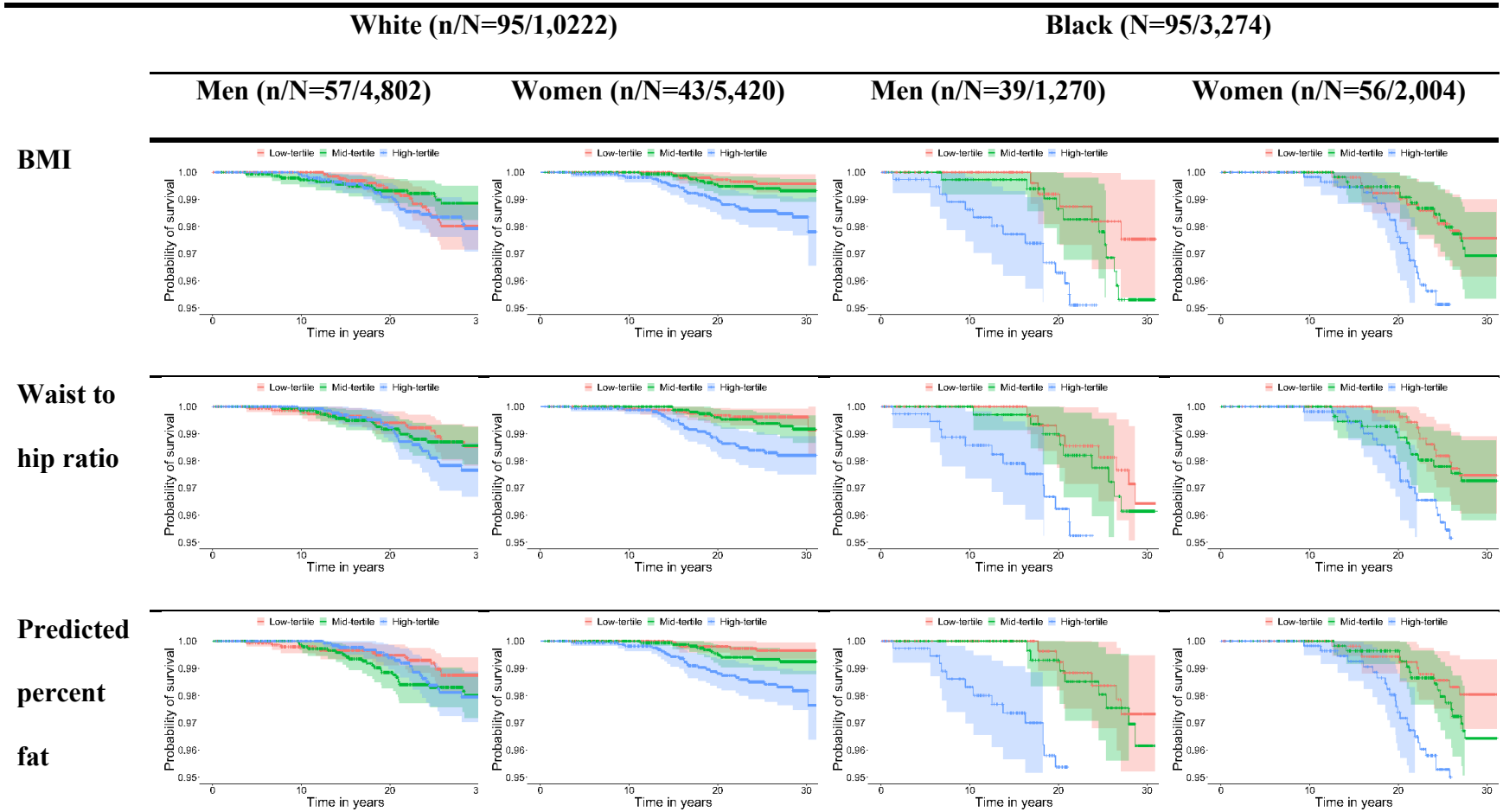
[b] Black participants in the Minnesota and Washington County centers were excluded in the model because of small numbers.

[c] Centered at median of each race-gender group.

[d] Hazard ratio (95% confidence interval) for all such values.

* indicates $p<0.05$, ** indicates $p<0.01$, *** indicates $p<0.001$.

Supplementary Figure 3-3 Kaplan-Meier survival free of end-stage kidney disease (ESKD) by tertile of baseline obesity measure within each sex-race group showing absolute risk of ESKD.



Chapter 4 Kidney Function and Blood Pressure: A Bidirectional Mendelian Randomization Study

Zhi Yu, BM, MS[1,2], Josef Coresh, MD, PhD[1,2], Guanghao Qi[3], Morgan Grams, MD, PhD[2,4], Eric Boerwinkle, PhD[5,6], Harold Snieder, PhD[7], Alexander Teumer, PhD[8,9], Cristian Pattaro, PhD[10], Anna Köttgen, MD, MPH[1,11], Nilanjan Chatterjee, PhD[3,4], Adrienne Tin, PhD[1,2,12]

[1] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

[2] Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore, Maryland

[3] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

[4] Department of Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland

[5] Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas

[6] Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas

[7] Department of Epidemiology, University Medical Centre Groningen, Groningen, Netherlands

[8] Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

[9] DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, Greifswald, Germany

[10] Eurac Research, Institute for Biomedicine (affiliated to the University of Lübeck), Bolzano, Italy

[11] Institute of Genetic Epidemiology, Department of Biometry, Epidemiology and Medical Bioinformatics, Faculty of Medicine and Medical Centre - University of Freiburg, Freiburg, Germany

[12] Division of Nephrology, Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi

ABSTRACT

Blood pressure and kidney function have a bidirectional relation. Hypertension has long been considered as a risk factor for kidney function decline. However, whether intensive blood pressure control could promote kidney health has been uncertain. The kidney is known to have a major role in affecting blood pressure through sodium extraction and regulating electrolyte balance. This bidirectional relation makes causal inference between these two traits difficult. To examine the causal relations between these two traits, we performed two-sample Mendelian randomization analyses using summary statistics of large-scale genome-wide association studies. We selected genetic instruments more likely to be specific for kidney function using complementary kidney function biomarkers (glomerular filtration rate estimated from serum creatinine [eGFRcr], *N*=567,460, and blood urea nitrogen, *N*=243,031, from the CKDGen Consortium). Systolic and diastolic blood pressure summary statistics were from the International Consortium for Blood Pressure and UK Biobank (*N*=757,601). Significant evidence supported the causal effects of higher kidney function on lower blood pressure. Based on the mode-based Mendelian randomization method, the effect estimates for 1 SD higher in eGFRcr was -0.17 SD unit (95 % CI: -0.09 to -0.24) in systolic blood pressure and -0.15 SD unit (95% CI: -0.07 to -0.22) in diastolic blood pressure. In contrast, the causal effects of blood pressure on kidney function were not statistically significant. Our results support causal effects of higher kidney function on lower blood pressure and suggest preventing kidney function decline can reduce the public health burden of hypertension.

INTRODUCTION

Hypertension and chronic kidney disease (CKD) are two interconnected global public health burdens. The estimated prevalence of hypertension is 31%, and CKD affects ~10% of adults.[1-3] Both CKD and hypertension are major risk factors for cardiovascular disease (CVD) and mortality.[112-114] Blood pressure (BP) and kidney function have a bidirectional relation. Hypertension has long been considered as a risk factor for kidney function decline based on observational studies.[7-10] However, it is uncertain whether intensive BP control could promote kidney health based on results from randomized controlled trials.[11,12] On the relation between kidney function decline and higher BP, two observational studies reported significant association between lower kidney function and incident hypertension. However, the kidney function biomarkers with significant association were cystatin C and beta-2 microglobulin, whereas serum creatinine, the most commonly used kidney function biomarker, was not significant.[115,116] These inconsistent results make inferring causal relation between BP and kidney function difficult. Evaluating the causal relations between kidney function and BP (Figure 1A) can inform disease prevention and treatment strategies.

Mendelian randomization (MR) is an approach employing genetic variants as instrumental variables of the exposure to estimate causal effects of the exposure on an outcome overcoming the confounding inherent in observational studies.[20] Two-sample MR analysis is an extension of the MR method that allows the use of summary statistics of genome-wide association studies (GWAS) for MR studies without directly analyzing individual-level data. Using two-sample MR analysis, Liu *et al.* found causal effects of systolic and diastolic BP on CKD.[117] Morris et al. reported significant causal effect of lower kidney function on higher diastolic BP (DBP) and not

on systolic BP (SBP).[118] Taking advantage of the recent large-scale meta-analysis of kidney function GWAS that included complementary glomerular filtration rate (GFR) biomarkers, we performed two-sample MR analyses to evaluate the potential bidirectional causal relation between kidney function and BP. The primary kidney function trait was estimated GFR based on serum creatinine (eGFRcr)[119] with blood urea nitrogen (BUN) as secondary. The primary BP trait was SBP with DBP as secondary.

To obtain robust conclusions from our analyses, we paid particular attention to two critical aspects in this MR study. One being the use of serum creatinine for GFR estimation, which might link eGFRcr to genetic variants related to creatinine metabolism and not GFR, making it difficult to interpret any causal findings between eGFRcr and BP. To address this issue, we used data from large-scale meta-analysis of GWAS of BUN, a complementary GFR biomarker, to select genetic instruments that are more likely to be specific to kidney function. The second being the assumption of the lack of horizontal pleiotropy of the genetic instruments, that is the genetic instruments must be associated with the outcome through the exposure only. This assumption is difficult to assess and verify.[34] To address this issue, we used multiple MR methods and prioritized the method that are known to be robust to the presence of horizontal pleiotropy and the influence of outlying genetic instruments.[35]

RESULTS

Summary of population characteristics

The meta-analysis of the GWAS of kidney function included largely adult population-based cohorts (among cohorts in the Chronic Kidney Genetic [CKDGen] Consortium, median age: 50.1 years; median of male %: 48.2; median eGFRcr: 91.4 mL/min/1.73m^2; N=567,460 for eGFRcr and N=243,031 for BUN, **Table 4-1**). The meta-analysis of the GWAS of blood pressure also included largely population-based cohorts (mean age of 54.9 years for the International Consortium for Blood Pressure [ICBP], 56.8 years for UK Biobank [UKB]; male %: 44.9% for ICBP, 45.8% for UKB; mean SBP: 134.3 mmHg for ICBP, 141.1 mmHg for UKB; N=299,024 for ICBP, N=458,577 for UKB, **Table 4-1**)

Selection of kidney function genetic instruments

The genetic instruments of kidney function were selected from the summary statistics of the meta-analysis of GWAS of European-ancestry participants of the CKDGen Consortium.[13] Of 256 reported eGFRcr index SNPs, after the removal of SNPs associated with potential confounders (Methods, **Supplementary Table 1**), 213 SNP remained (**Figure 4-2**, **Supplementary Table 2**). Using BUN as the complementary kidney function marker, we retained 40 index SNPs based on their association with BUN having direction consistency in kidney function and satisfying Bonferroni-corrected significance threshold (Methods, **Supplementary Table 3**). For example, the index SNP at *GATM*, which encodes an enzyme in creatine metabolism[120], was removed due to insignificant association with BUN (rs1145077, eGFRcr P=6.9×10^{-142}, BUN P=0.92). After pairwise-linkage disequilibrium (LD) clumping and matching of coding allele between exposure and outcome, 35 index SNPs remained. Finally, Steiger filtering removed the index SNPs at *FGF5* and *SPI1* (**Supplementary Table 4**) resulting in 33 genetic instruments for eGFRcr, which explained 1.33% of the variance of log(eGFRcr).

None of these 33 genetic instruments were associated with urine sodium-to-creatinine ratio or urine potassium-to-creatinine ratio at genome-wide significance level ($p<5x10^{-8}$) in a large-scale GWAS of UKB participants (N=327,613, **Supplementary Table 5)**.[121] Using similar selection procedures, the number of genetic instruments retained for BUN was 24, which explained 1.18% of the variance of log(BUN). For example, using eGFRcr as the complementary kidney function biomarker, the BUN index SNP at *SLC14A2*, a urea transporter[122,123] was removed due to insignificant association with eGFRcr (rs41301139, *P*=0.14) (**Supplementary Table 6)**. The numbers of SNPs retained after each selection step are reported in **Supplementary Table 7**.

Significant causal effect of kidney function on BP

We identified significant evidence for causal effects of higher kidney function on lower BP. Based on weighted mode, our primary method, the causal effect estimates for each standard deviation (SD) higher log(eGFRcr) were -0.17 SD in SBP (95% confidence interval [CI]: -0.24 to -0.09; $P=9.92\times10^{-5}$) and -0.15 SD in DBP (95% CI: -0.22 to -0.07; $P=5.02\times10^{-4}$, **Figure 4-3**). These causal effects were equivalent to 10% higher eGFRcr leading to 2.35 mmHg lower in SBP and 1.14 mmHg lower in DBP. If the relation between eGFRcr and blood pressure is linear,[124,125] a 50% lower in eGFRcr would lead to 17.5 mmHg higher SBP and 8.4 mmHg higher DBP. We also observed significant causal effects of BUN, the secondary kidney function trait, on SBP and DBP (weighted mode method, SBP $P=4.92\times10^{-4}$; DBP $P=3.88\times10^{-6}$). Using other MR methods: inverse-variance-weighted fixed-effects (IVW-FE) method, MR-Egger, weighted median, and MR analysis using mixture models (*MRMix*),[126-129] all causal effect estimates were in the same direction as those from weighted mode and statistically significant, providing support for causal effects of lower eGFRcr on higher SBP and DBP (**Supplementary Table 8**). MR-PRESSO

analysis with the explicit removal of outlying instruments showed similar results (**Supplementary Table 9**). The scatter plots with the regression line from all MR methods are presented in **Supplementary Figures 1A-4A.** The forest plots of single SNP effects from each of the kidney function traits to each of the BP traits are presented in **Supplementary Figure 1B-4B**. Multiple sensitivity analyses resulted in similar causal effect estimates for kidney function on BP (**Supplementary Tables 10 and 11**, Supplementary Results).

Selection of genetic instruments for SBP and DBP

The genetic instruments of SBP and DBP were selected from the summary statistics of the meta-analysis of GWAS from the UKB-ICBP.[31] Of the 551 reported index SNPs of SBP, after the removal of SNPs associated with potential confounders, LD clumping, matching of coding allele, 250 index SNPs remained (**Supplementary Tables 1, 2, 7, Figure 4-2).** When eGFRcr was used as the outcome, Steiger filtering removed 10 SBP index SNPs including those at *UMOD/PDILT* and *PRKAG2* (**Supplementary Table 4**), resulting in 240 genetic instruments for SBP explaining 2.46% of the variance of SBP. Of the 537 reported DBP index SNPs, after the removal of SNPs associated with potential confounders, LD clumping, and matching of coding allele, 250 remained (**Supplementary Tables 1, 2, 7**). When eGFRcr was used as the outcome, Steiger filtering removed 8 DBP index SNPs including those at *UMOD/PDILT* and *PRKAG2* (**Supplementary Table 4**), resulting in 234 genetic instruments for DBP explaining 2.85% of the variance of DBP. With the same SNP selection algorithms, 248 SBP and 237 DBP genetic instruments were selected when CKD was the outcome, and 243 SBP and 233 DBP genetic instruments were selected when BUN was the outcome (**Supplementary Tables 4** and **7**).

Causal effect estimates of BP on kidney function

Based on weighted mode, our primary method, we observed that the causal effect estimates of BP on kidney function were generally not significant accounting for multiple testing (P<0.025, Methods). The effect estimate for each SD higher SBP was -0.09 SD in log(eGFRcr) (95% CI: -0.18 to -0.002; P=4.71×10^{-2}, **Supplementary Figure 4**, **Supplementary Table 8**). This effect estimate is equivalent to 10 mmHg higher in SBP leading to 0.6% lower in eGFRcr. The causal effect estimates of SBP on eGFRcr from other MR methods were also modest (10 mmHg higher in SBP for <1% difference in eGFRcr). Similar non-significant and modest causal effect estimates were observed for DBP on the three kidney function outcomes (eGFRcr, CKD, and BUN) using weighted mode as well as MRMix, one of the methods most robust to horizontal pleiotropy (**Supplementary Table 8**).[130] In contrast, using IVW-FE, the causal effect estimates were significant across both SBP and DBP on kidney function outcomes. The results using MR-PRESSO, an extension of IVW-FE with outlier removal were consistent with the IVW-FE results (**Supplementary Table 9**). The scatter plots with the regression line from all MR analyses are presented in **Supplementary Figures 5A-10A.** The forest plots of single SNP effects from BP traits to kidney function traits are presented in **Supplementary Figures 5B-10B**. Sensitivity analysis using BP summary statistics from UKB only as exposure resulted in similar causal estimates for SBP and DBP on eGFRcr, BUN, and CKD (**Supplementary Table 10**).

DISCUSSION

Extensive MR analyses, based on the largest GWAS summary statistics available to date on kidney function and BP, showed evidence of a causal role of kidney function on BP levels. Specifically, we observed that 50% lower in eGFRcr results in 17.5 mmHg higher SBP and 8.4 mmHg higher DBP. In contrast, the causal role of BP on kidney function levels were not supported across MR methods. The significant causal effect of lower kidney function on higher BP suggests preventing kidney function decline can reduce the public health burden of hypertension.

Kidney function is difficult to measure and is generally estimated using biomarkers.[131] The systematic measurement errors in GFR estimation due to the biomarker post challenges for the study of kidney function, particularly for early kidney function decline using eGFRcr because the systematic measurement errors increase as GFR approaches or within the normal range[119]. In two population-based association studies that reported significant association between lower kidney function and incident hypertension, the kidney function biomarkers with the significant association were cystatin C and beta-2 microglobulin, whereas serum creatinine, the most commonly used GFR biomarker, was not significant.[115,116] In this MR study, we used a novel approach of combining GFR biomarkers in selecting the genetic instruments for kidney function to overcome the systematic measurement errors in eGFRcr and found significant causal effects for lower kidney function on higher BP at the population level. This finding is consistent with the important physiological role of the kidney in affecting BP through the regulation of sodium excretion and electrolyte balance[132] and with the genetics of hypertension-attributed kidney

disease in African Americans, in whom the *APOL1* high-risk genotype confers twice the risk of CKD progression and appears to directly affect kidney function rather than BP.[95,133,134]

Our study extends previous MR studies of kidney function on BP. Significant bidirectional causal effects between higher albuminuria, an indicator of kidney damage, and higher SBP and DBP have been reported using the large-scale UKB data.[135] In contrast, the causal effects of eGFRcr on SBP and DBP had inconsistent results. Morris *et al*. reported significant causal effect of eGFRcr on DBP but not on SBP.[118] Our study used a complementary GFR biomarker approach to identify genetic instruments that were more likely to be specific for GFR and found robust causal effects of eGFRcr on both SBP and DBP.

On the causal direction from BP on kidney function, significant causal effects of higher SBP and DBP on CKD have been reported using the IVW-FE method, which requires strong assumption on the sum of horizontal pleiotropy to be zero to provide consistent estimates.[136] In our study, the IVW-FE method also provided significant causal effect estimates of SBP and DBP on CKD with effect sizes similar to those previously reported.[117] If the assumption on horizontal pleiotropy of the IVW-FE method is valid, the IVW-FE method would be more powerful. Given the substantial heterogeneity of blood pressure genetic instruments on kidney function, the horizontal pleiotropy assumption of the IVW-FE method may not hold. Across all methods, the effect estimates of BP on kidney function were small (<1% difference in eGFRcr per 10 mmHg difference in SBP). Our inconclusive MR results of BP on kidney function might be due to measurement inaccuracies in kidney function contributed by GFR biomarkers leading to a

dilution of the causal effect estimates. Another potential source of these inconclusive results might be the ability of the kidney to adapt to small differences in blood pressure.[137,138] Some have hypothesized that only severe hypertension would lead to kidney function decline.[139] It should also be noted that our results mirror, to some extent, the results of BP control in randomized controlled trials, which have failed to show consistent protective effects for kidney function, particularly within the trial period.[140-142]

Our use of Steiger filtering, which compared the effect size of a genetic instrument for exposure and outcome, suggested that some kidney function loci may affect kidney function through BP, such as *FGF5*, and some BP loci may affect BP through kidney function, such as *UMOD,* which expresses exclusively in the kidney.[143] These results provide insight into the potential pleiotropy underlying GWAS findings of these traits.

Our study has several strengths. First, we used a novel approach of combining GFR biomarkers to select genetic instruments that are more likely to be specific to kidney function to overcome measurement inaccuracies in kidney function contributed by GFR biomarkers. Second, we used summary statistics from large-scale GWAS. Lastly, to reduce the possibility of violating the assumptions of MR, we employed a range of techniques: evaluation of the association of index SNPs with potential confounders, use of Steiger filtering to reduce potential reverse causation driven by genetic instruments, and selecting a primary method that is robust to the presence of pleiotropy accompanied by sensitivity analysis with alternative methods.

Some limitations warrant mentioning. The MR approach uses genetic instruments to represent lifelong difference in exposure levels for estimating causal effects on an outcome. Developmental adaptation could alter the effect of the genetic instruments on the outcome.[144] The two-sample MR methods rely on GWAS summary statistics and assume a linear relationship between the exposure and the outcome. We did not evaluate potential non-linear relationship between kidney function and blood pressure or investigate potential mechanisms linking kidney function and blood pressure, such as sodium and potassium handling. In our primary analysis, the cohorts in CKDGen and UKB-ICBP had some overlap, which might lead to bias in the causal estimates.[145] However, our results using non-overlapping populations in exposure and outcome (CKDGen and UKB only) were similar to our primary analysis. The CKDGen populations included studies of CKD patients and children. These studies only made up a small proportion of the study population. Overall, the populations in the summary statistics for exposures and outcomes were similar.[146]

In summary, using a novel approach that combines GFR biomarkers for selecting genetic instruments for kidney function, we found that lower kidney function is causal to higher BP. This result suggests that preventing kidney function decline may reduce the public health burden of hypertension.

METHODS

Study design overview

We performed two-sample MR analyses to estimate the causal effects of kidney function on BP and vice versa. The primary kidney function trait was eGFRcr with BUN as a secondary trait. CKD, defined as eGFRcr<60 mL/min/1.73m^2, was a secondary outcome.[13] The primary BP trait was SBP with DBP as secondary.[147] Published GWAS summary statistics were obtained from the European-ancestry meta-analysis of the CKDGen Consortium for kidney function[13] and the UKB-ICBP for BP.[31] Genotypes in the GWAS were imputed using the Haplotype Reference Consortium (HRC)[148] or the 1000 Genomes Project reference panels.[149] All GWAS summary statistics assumed an additive genetic model.

Summary statistics of kidney function from the CKDGen Consortium

The meta-analysis of the GWAS of eGFRcr included 54 cohorts of European ancestry (*N*=567,460), largely adult population-based (**Table 1**). A small proportion of the participants were from cohorts of CKD patients, diabetes patients, or children (2.5%). The meta-analysis of the GWAS of BUN included 48 cohorts of European ancestry (*N*=243,031), and the analysis of CKD included 444,971 participants. eGFRcr was calculated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation[119] for adults and the Schwartz formula[150] for participants who were 18 years old or younger. BUN, the secondary kidney function trait, was derived as blood urea×2.8 mg/dl.[13] The phenotypes used in the GWAS of eGFRcr and BUN were the natural log transformed age- and sex-adjusted residuals of the traits. The basic characteristics of the CKDGen studies were summarized as weighted averages across studies with the sample size as the weight using summary data reported in Wuttke *et al*. 2019.[13]

Summary statistics of BP from the UKB-ICBP

Summary statistics of BP traits were obtained from the meta-analysis results of the UKB-ICBP.[31] The UKB is a population-based cohort.[30] SBP and DBP were calculated as the mean of two automated or manual BP measurements, except for participants with one BP measurement (*N*=413). The GWAS of SBP and DBP in UKB included 458,577 participants of European ancestry. The meta-analysis of the GWAS of SBP and DBP from ICBP included 77 cohorts of European ancestry (*N*=299,024).[147] In both UKB and ICBP, the values of SBP and DBP were adjusted for the use of BP lowering medications by adding 15 and 10 mmHg, respectively.[31,151] The characteristics of the ICBP studies were summarized as weighted averages across studies with the sample size as the weight using summary data reported in Evangelou *et al*. 2018.[31]

MR assumptions

Genetic instruments in MR studies rely on three assumptions: (i) the SNP must be associated with the exposure; (ii) the SNP is independent of confounders, i.e. other factors that can affect the exposure-outcome relationship; and (iii) the SNP must be associated with the outcome through the exposure only, i.e., no direct association due to horizontal pleiotropy.[144]

Selection of genetics instruments more likely to be related to kidney function

To ensure that the genetic instruments satisfied the first assumption with respect to kidney function, we selected index SNPs associated with multiple GFR biomarkers, so that they are more likely to be related to kidney function, the exposure of interest, rather than GFR

biomarkers. For the primary analysis using eGFRcr, we started with the index SNPs of the genome-wide significant loci of the European-ancestry meta-analysis of eGFRcr from the CKDGen Consortium.[13] We first evaluated the association of the index SNPs with potential confounders using the GWAS summary statistics from UKB for the following traits: prevalent diabetes, body mass index (BMI), triglycerides and high-density lipoprotein cholesterol (HDL-C) levels, smoking, and prevalent coronary heart disease (**Table S1**).[1,152,153] We removed index SNPs with genome-wide significant associations (5×10^{-8}) with the potential confounders listed above.[1,152,154]

Next we used genetic association information of BUN,[13] an alternative biomarker of kidney function, to select genetic instruments that were more likely to reflect kidney function as opposed to creatinine metabolism. This approach was similar to the approach in Wuttke *et al.* for prioritizing genetic loci most likely to be relevant for kidney function.[13] We required that the eGFRcr index SNPs to be associated with BUN at a Bonferroni-corrected significance ($P<0.05$ divided by the number of eGFRcr index SNPs) and in opposite direction since higher GFR would lead to lower BUN. To ensure independence among genetic instruments, we applied LD clumping[155] with a clumping window of 10 MB and an r^2 cutoff of 0.001 (default of the ld_clump function).[155] The matching of the effect allele of each SNP between the summary statistics of the exposure and the outcome was examined using the harmonise_data function, which removed SNPs that were palindromic or had possible strand mismatch. Finally, to reduce the possibility that a genetic instrument might affect the outcome independently of the exposure, we applied Steiger filtering to ensure that the association between a genetic instrument and the exposure was stronger than its association with the outcome.[156]

To select genetic instruments of BUN, the secondary kidney function trait, we started with index SNPs from the GWAS of BUN and followed similar procedure of selection described above. We used their association with eGFRcr for screening out those index SNPs that might only be related to metabolism of BUN but not to kidney function.

Selection of genetics instruments for BP

For the BP traits, we started with the index SNPs from genome-wide significant loci of SBP or DBP reported by the UKB-ICBP,[31] applied the same steps as described above for eGFRcr, without the alternative biomarker step. Briefly, we removed index SNPs that were associated with potential confounders listed above, removed correlated SNPs using the ld_clump function,[155] used the harmonise_data function to remove SNPs that were palindromic or had possible strand mismatch between the summary statistics of the exposure and outcome, and finally, we applied Steiger filtering[156].

Use of robust method to account for horizontal pleiotropy

Some methods for MR analysis can be heavily biased in the presence of direct association of SNP with the outcome that is not mediated by the exposure.[157] When the direct effects of genetic instrument on the outcomes and the exposures are correlated across different instruments due to the presence of unobserved confounders that may have heritable components, the bias can be severe.[130] To reduce the possibility that the genetic instruments might affect the outcome

independently of the exposure, in addition to the use of Steiger filtering,[156] we chose the weighted mode method, one of the most robust in the presence of horizontal pleiotropy,[130,136] as our primary MR method. In addition, we conducted sensitivity analysis using alternative methods that may be more powerful under various model assumptions (see Supplementary Methods section). Given our primary analyses were the causal effects of eGFRcr on SBP and vice versa, the significance level for MR analysis was set at $P<0.025$ (=0.05/2).

Units of causal effect estimates, variance explained by genetic instruments, and sensitivity analyses

For continuous exposures and outcomes, we estimated the causal effects of 1 SD difference of the exposure on the outcome. The SD of each trait was estimated based on data from population-based cohorts. The details are reported in Supplementary Methods.

Details of the calculation of exposure variance explained by the genetic instruments, power analysis, and sensitivity analysis are reported in the **Supplementary Methods and Results** section. All analyses were conducted using R (*version 3.5.3*), and the "TwoSampleMR" package was used for all MR analyses, except MRMix.

Table 4-1 Basic characteristics of the studies that contributed summary statistics of kidney function (CKDGen) and blood pressure (ICBP-UK Biobank).

	CKDGen	
Sample size	567,460	
Age (years), mean	50.1	
Male, %	48.2	
eGFRcr (mL/min/1.73m^2), median	91.4	
CKD, %	8.6	
	ICBP	**UK Biobank**
Sample size	299,024	458,577
Age (years), mean	54.9	56.8
Male, %	44.9	45.8
SBP (mmHg) , mean	134.3	141.1
DBP (mmHg) , mean	80.6	84.3

The basic characteristics of the studies participated in the CKDGen consortium were calculated as the weighted average of the characteristics of each study reported in Wuttke et al. Nat Genet 2019 with study-specific sample size as the weight.

The basic characteristics of the studies participated in the ICBP were calculated as the weighted average of the characteristics of each study with study-specific sample size as the weight. These study-specific characteristics and the UK Biobank characteristics were reported in Evangelou *et al*. Nat Genet 2018.

Abbreviation. CKDGen, Chronic Kidney Disease Genetics; ICBP, International Consortium for Blood Pressure; eGFRcr, estimated glomerular filtration rate calculated from serum creatinine; CKD, chronic kidney disease defined as eGFRcr < 60 mL/min/1.73m^2; SBP, systolic blood pressure; DBP, diastolic blood pressure.

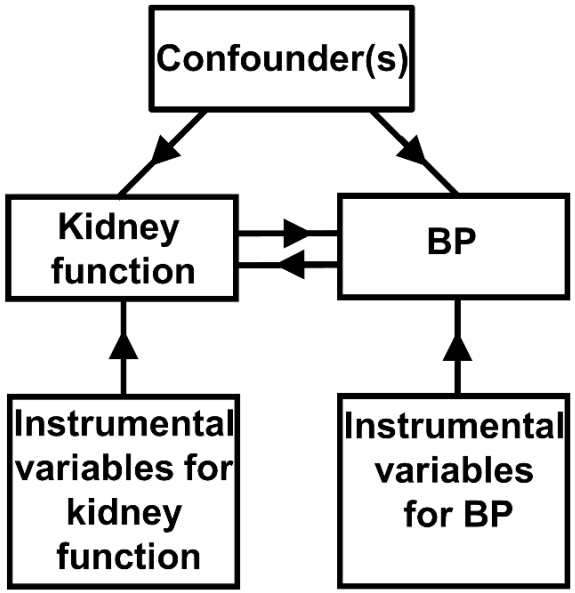


Figure 4-1. The hypothesized bi-directional relations between kidney function and blood pressure (BP) depicted in a directed acyclic graph (DAG) where the arrows represent causal relations (adapted from Davey Smith et al. 2014, PMID: 25064373). In our study, the primary kidney function measures were glomerular filtration rate (GFR) estimated from serum creatinine and not measured directly.

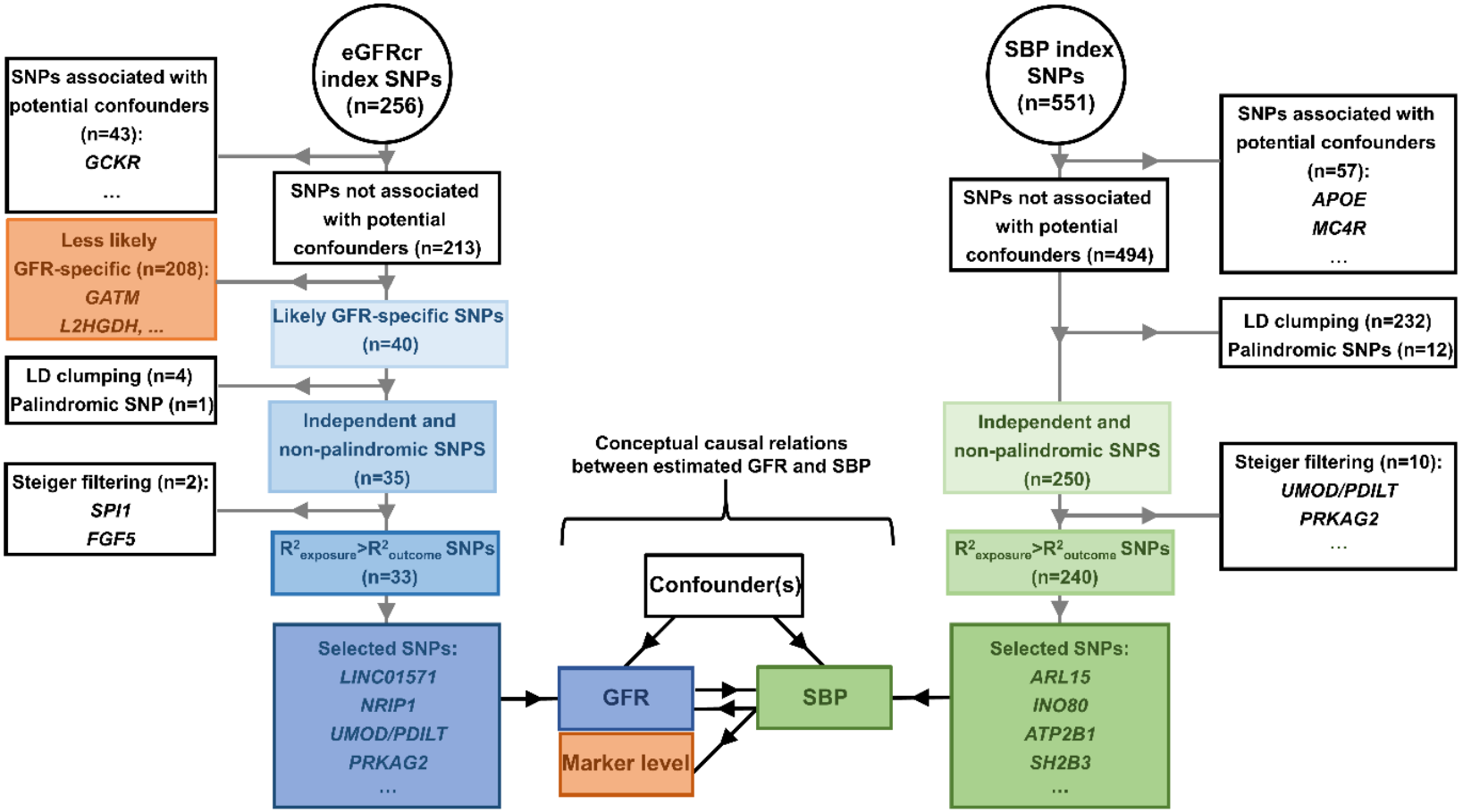


Figure 4-2 Selection of genetic instruments for GFR estimated from serum creatinine (eGFRcr) and systolic blood pressure (SBP), our primary traits. The centered graph represents the conceptual causal relations (black arrows) between estimated GFR and SBP, similar to the DAG in Figure 1A, except that estimated GFR levels were calculated from serum creatinine and therefore has two

parts: a GFR and a biomarker level components. The diagrams on the two sides with grey arrows represent the process of genetic instrument selection. Abbreviation: SNP, single nucleotide polymorphism.

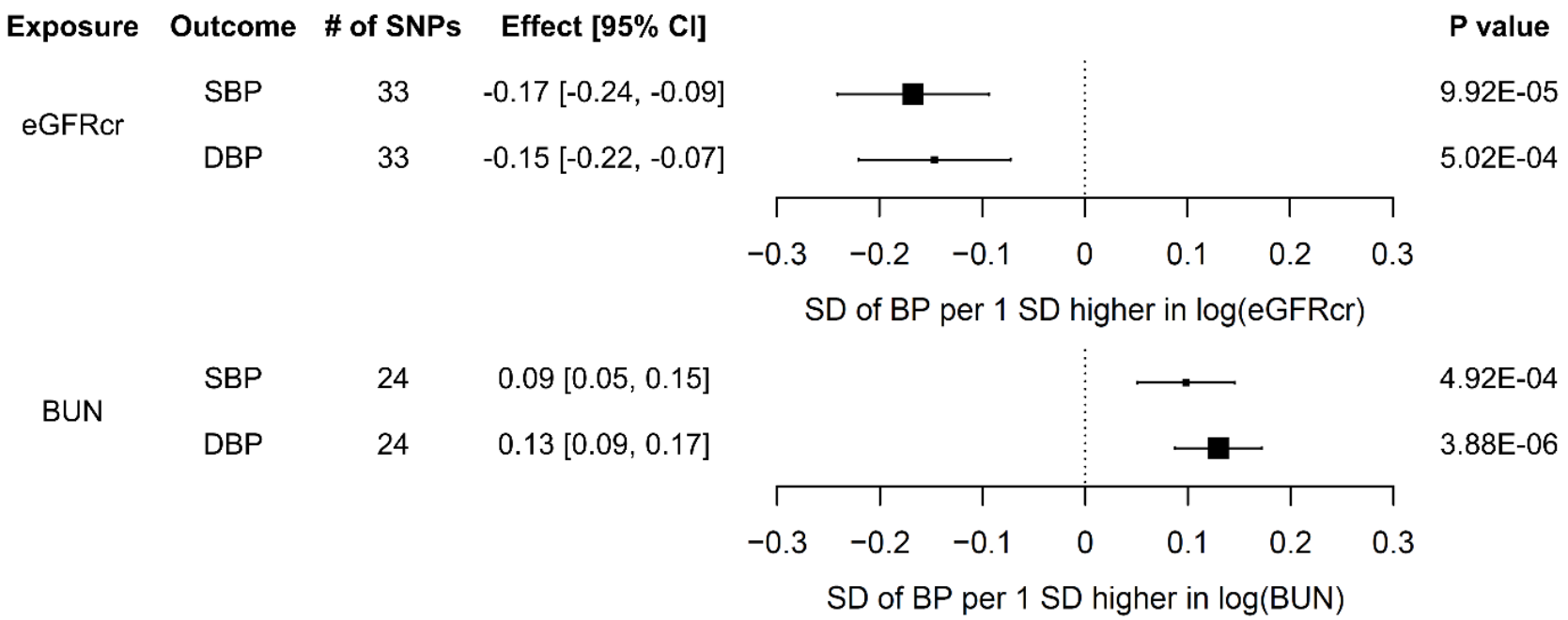


Figure 4-3 Causal effect estimates of log(eGFRcr) and log(BUN) on SBP and DBP using the weighted mode method.

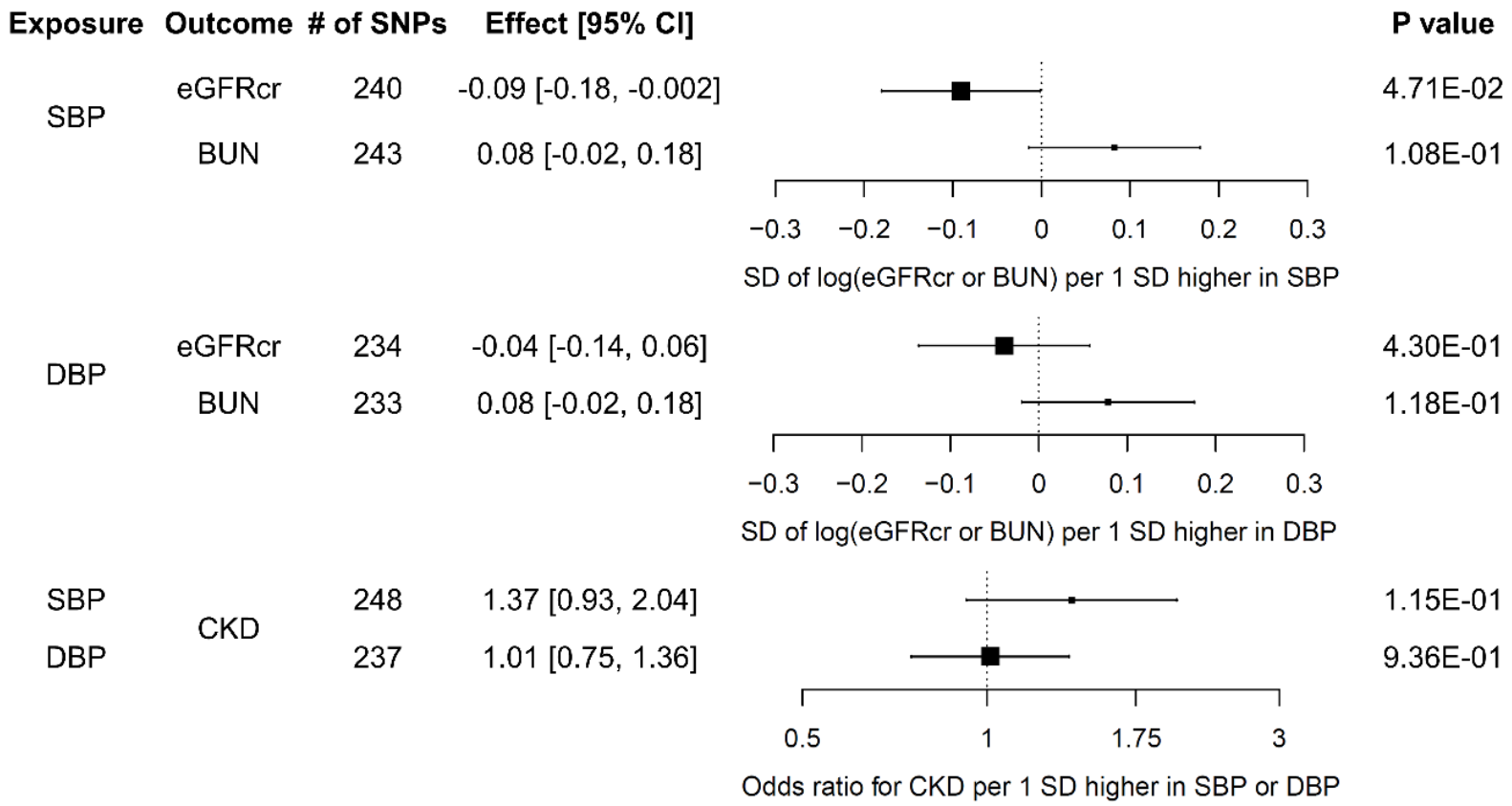


Figure 4-4 Causal effect estimates of SBP and DBP on log(eGFRcr), log(BUN) and CKD using the weighted model method.

Supplementary Materials Index (link: https://www.dropbox.com/sh/f9nlg2s775zdpqh/AAABxAJu-1iBw4j5UUUlIUqDa?dl=0)

Supplementary Methods and Results

Supplementary Table 1. Source and sample size of GWAS summary statistics of potential confounders.

Supplementary Table 2. Index SNPs significantly associated with potential confounders.

Supplementary Table 3. Association of 213 eGFRcr index SNPs with BUN in the meta-analysis of CKDGen European ancestry cohorts (eGFRcr N=567,460, BUN N=243,031).

Supplementary Table 4. SNPs excluded as genetic instruments due to having potential reverse causal direction (exposure r-squared ≤ outcome r-squared) based on Steiger filtering.

Supplementary Table 5. Association of the 33 genetic instruments of eGFRcr with urine sodium-to-creatinine ratio and urine potassium-to-creatinine ratio in large scale GWAS of UKB participants (N=327,613, Zanetti et al. 2019)

Supplementary Table 6. Association of 58 BUN index SNPs with eGFRcr in the meta-analysis of CKDGen European ancestry cohorts (BUN N=243,031, eGFRcr N=567,460).

Supplementary Table 7. Numbers of index SNPs retained after in each step of selecting genetic instruments.

Supplementary Table 8. Results of MR analysis of kidney function and BP by primary and secondary methods using CKDGen and UKB-ICBP summary statistics.

Supplementary Table 9. Results of MR-PRESSO for eGFRcr on SBP and SBP on eGFRcr using CKDGen and UKB-ICBP summary statistics.

Supplementary Table 10. Sensitivity analysis results using non-overlapping samples from CKDGen and UKB.

Supplementary Table 11. Sensitivity analysis results using eGFRcys as the alternative biomarker for kidney function for selecting GFR genetic instruments.

Supplementary Figure 1A. Regression lines of MR tests from eGFRcr on SBP.

Supplementary Figure 1B. Forest plot of single SNP from eGFRcr on SBP.

Supplementary Figure 2A. Regression lines of MR tests from eGFRcr on DBP.

Supplementary Figure 2B. Forest plot of single SNP from eGFRcr on DBP.

Supplementary Figure 3A. Regression lines of MR tests from BUN on SBP.

Supplementary Figure 3B. Forest plot of single SNP from BUN on SBP.

Supplementary Figure 4A. Regression lines of MR tests from BUN on DBP.

Supplementary Figure 4B. Forest plot of single SNP from BUN on DBP.

Supplementary Figure 5A. Regression lines of MR tests from SBP on eGFRcr.

Supplementary Figure 5B. Forest plot of single SNP from SBP on eGFRcr.

Supplementary Figure 6A. Regression lines of MR tests from SBP on CKD.

Supplementary Figure 6B. Forest plot of single SNP from SBP on CKD.

Supplementary Figure 7A. Regression lines of MR tests from SBP on BUN.

Supplementary Figure 7B. Forest plot of single SNP from SBP on BUN.

Supplementary Figure 8A. Regression lines of MR tests from DBP on eGFRcr.

Supplementary Figure 8B. Forest plot of single SNP from DBP on eGFRcr.

Supplementary Figure 9A. Regression lines of MR tests from DBP on CKD.

Supplementary Figure 9B. Forest plot of single SNP from DBP on CKD.

Supplementary Figure 10A. Regression lines of MR tests from DBP on BUN.

Supplementary Figure 10B. Forest plot of single SNP from DBP on BUN.

Chapter 5 Polygenic Risk Scores of Kidney Function, Circulating Proteome, and Incident Kidney Diseases: the Atherosclerosis Risk in Communities Study

Zhi Yu, BM, MS[1,2], Jin Jin, PhD[3], Adrienne Tin, PhD[1,4], Anna Kottgen, MD, MPH[1,5], Bing Yu, PhD[6], Jingsha Chen, MS[2], Christie M. Ballantyne, MD[7], Ron C. Hoogeveen, PhD[7], Dan E. Arking, PhD[8], Nilanjan Chatterjee, PhD[3,9], Morgan E. Grams, MD, PhD[2,9], Josef Coresh MD, PhD[1,2,9]

[1] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

[2] Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore, Maryland

[3] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

[4] Division of Nephrology, Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi

[5] Institute of Genetic Epidemiology, Department of Biometry, Epidemiology and Medical Bioinformatics, Faculty of Medicine and Medical Centre - University of Freiburg, Freiburg, Germany

[6] Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas

[7] Department of Medicine, Baylor College of Medicine, Houston, Texas

[8] McKusick-Nathans Department of Genetic Medicine, School of Medicine, Johns Hopkins University, Baltimore, Maryland

[9] Department of Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland

ABSTRACT

Genome-wide association studies (GWAS) have revealed numerous loci for kidney function (estimated glomerular filtration rate, eGFR). The relationship of polygenic predictors of eGFR, risk of incident adverse kidney outcomes, and the plasma proteome is not known. We developed a genome-wide polygenic risk score (PRS) using a weighted average of 1.2 million SNPs for eGFR using the LDpred algorithm, summary statistics generated by a European-ancestry (EA) meta-analysis of the CKDGen Consortium (N=558,423) and UK Biobank GWAS for eGFR (90% of the cohort; N=289,432), followed by best parameter selection using data from the remaining 10% of the UK Biobank (N=32,159). We then tested the association of the PRS among 8,886 EA participants in the Atherosclerosis Risk in Communities (ARIC) study (mean age: 54±6 years, 53% female) with incident chronic kidney disease (CKD), end stage kidney disease (ESKD), kidney failure (KF), and acute kidney injury (AKI). We also examined 4,877 plasma proteins measured at two time points (visit 3 (1993-95) and visit 5 (2011-13)) in relation to the PRS and compared associations between the proteome and eGFR itself. All models were adjusted for age, sex, center, and the first 10 principal components of ancestry. The developed PRS had an R^2 for eGFR of 0.078 in ARIC. Over 30 years of follow up, the number of incident CKD, ESKD, KF, and AKI were 2,959, 137, 470, and 1,723, respectively. The PRS showed significant associations with all outcomes: hazard ratios (95% CI) per 1 SD lower PRS were 1.33 (1.28, 1.39) , 1.20 (1.00, 1.42), 1.17 (1.06, 1.28), and 1.07 (1.02, 1.12) for incident CKD, ESKD, KF, and AKI respectively. The PRS was significantly associated (Bonferroni threshold $p<1.02 \times 10^{-5}$) with 108 proteins at both time points. The strongest associations were with cystatin-C (a marker of kidney function used in clinical practice), collagen alpha-1(XV) chain, and desmocollin-2. All significant associations were inversely correlated with the PRS, except that of

testican-2 and angiostatin. Correlations of proteins with eGFR were much stronger than those with the PRS. By incorporating variants across the entire genome, we demonstrated a link between the genetic basis of eGFR and a spectrum of incident kidney diseases and use implicate proteins across the plasma proteome that may reflect or lead to reduced kidney function.

INTRUDUCTION

Most kidney diseases are complex diseases with both genetic and environmental factors contributing to their risks. The heritability estimated by familiar studies are up to 30–75%.[158-161] Genome-wide associations studies (GWAS) have grown rapidly in the last decade and identified numerous loci for kidney function, which gave rise to increasing attention in using polygenic risk scores (PRS) to predict kidney diseases risks.[13-19] However, previous PRS provided limited risk stratification for adverse kidney outcomes such as end-stage kidney disease (ESKD).[13,18] Potential reasons include: small sample sizes of early GWAS, which might lead to imprecise estimation of the associations between individual variants and disease risk; limiting the PRS to genetic variants that reached genome-wide significance ($p < 5 \times 10^{-8}$); and a lack of deeply phenotyped data to identify cases.[14-19] With new data and methodologies, there is an opportunity to mitigate these limitations.

New methodologies for large scale proteomic measurement using aptamer technologies also provide an opportunity to assess the impact of genetic susceptibility to low kidney function on the plasma proteome.[21,22] The plasma proteome consists of thousands of secreted proteins that involve in numerous physiological and pathological processes, including transporting and signaling, metabolism, vascular function, and defense mechanisms.[25-27] As those proteins reflect the current state of human body from various aspects, the plasma proteome is a reservoir of important potential biomarkers. Although previous studies have demonstrated the heritability of plasma protein levels,[28] research in the plasma proteomic signals of genetic susceptibility of diseases, including kidney diseases, has been limited.[162-165] Also, for kidney diseases, as reduced

kidney function results in elevations of many proteins, it is particularly important to examine the balance of genetically predicted risk vs. secondary influences on proteomic signals.

Using large studies and new algorithms, we investigated the strength of associations of PRS for kidney function with incident kidney diseases over 30 years of follow-up in a deeply phenotyped community-based cohort. We included diseases strongly related to kidney function with evidence of a strong genetic basis, including chronic kidney disease (CKD), end-stage kidney disease (ESKD), and kidney failure,[158-161] as well as acute kidney injury (AKI).[166] We also examined 4,877 plasma proteins measured at two time points approximately 20 years apart in relation to both genetic susceptibility to low kidney function and the concurrent kidney function itself, in order to evaluate the strengths of proteomic associations with genetically predicted risk and physiological changes and how those associations change over time.

METHODS

Study Cohort

The Atherosclerosis Risk in Communities (ARIC) study is an ongoing longitudinal cohort of 15,792 45-60-year-old participants (55% female, 73% participants of European ancestry (EA)) recruited from four communities in the U.S.: Forsyth County, North Carolina; Jackson, Mississippi; suburbs of Minneapolis, Minnesota; and Washington County, Maryland at 1987-1989 (visit 1). Follow-up examinations were conducted approximately every three years: 1990-1992 (visit 2), 1993-1995 (visit 3), 1996-1998 (visit 4), more recently, in 2011-2013 (visit 5), in 2016-2017 (visit 6), and in 2018-2019 (visit 7).[167] Each study visit consisted of a clinical examination, blood and urine specimens collection, and filling out extensive questionnaires.

Proteomic levels were measured at visit 3 and visit 5. Our primary study population was restricted to 8,866 unrelated EA participants (**Supplemental Figure 1**), since so far most of the genomic studies with results available for use were conducted among EA. In the proteomic analysis of our study, 7,213 participants with valid proteomic measurements remained. In sensitivity analysis, we constructed another study population of 2,871 unrelated participants with African ancestry (AA). Study protocols were approved by the Institutional Review Boards and all study participants provided informed consent (including agreement for industry studies for SomaLogic sponsored proteomic quantification).

Genotyping

Participants were genotyped with the Affymetrix 6.0 DNA microarray (Affymetrix, Santa Clara, CA) with genotype calling performed using the Birdseed algorithm. Genotyping was performed on the Affymetrix 6.0 DNA microarray (Affymetrix, Santa Clara, CA) and analyzed with the Birdseed variant calling algorithm. Haplotype phasing was performed using ShapeIt (v1.r532).[168] Genotypes were imputed on the Michigan Server to the TOPMed reference panel.[169,170] A quality control was carried out prior to imputation: SNPs were included if they had call rate < 95%, Hardy-Weinberg equilibrium p-values < 0.0001, or minor allele frequencies (MAF) < 1%.[171] Individuals with cryptic relatedness defined as identity by state (IBS) distance generated from PLINK > 0.8 were also excluded.[155]

Assessing Kidney Function

Kidney function, measured as estimated glomerular filtration rate (eGFR), was assessed by measuring serum creatinine (at all visits excepted visit 7) and serum cystatin C (at all visits

excepted visit 1 and 7) using the 2009 Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) creatinine equation (eGFRcr) and 2012 CKD-EPI cystatin C equation (eGFRcys).[119,172] Cystatin C is an excellent marker for kidney function but is not as widely used as creatinine, which limits its use in genetics studies. Serum creatinine level was measured by the modified kinetic Jaffé method, standardized to the National Institute of Standards and Technology (NIST) standard, and calibrated to an isotope dilution mass spectrometry (*IDMS*)-traceable reference method.[56-58] Serum cystatin C level was measured by the turbidimetric method, and standardized and calibrated to the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) reference.[173] In the polygenic risk scores development, we used eGFRcr as the kidney function measurement since this has been the main trait with the largest samples size in GWAS meta-analysis.[13]

Polygenic Risk Score for Kidney Function

Polygenic risk scores aggregate genome-wide genetic variation into a single score that reflects individual's inherited disease risk. They are most commonly calculated by summing across SNPs associated with a given trait, weighted by their effect sizes from GWAS results of that trait.

For the PRS construction, we first conducted a GWAS for log(eGFRcr) using PLINK among 90% of unrelated EA participants in the UK Biobank (N=289,432; application ID 17712) using an additive genetic model adjusted for age and sex.[155] Details of the UK Biobank cohort has been described elsewhere.[174] Then we conducted a fixed-effects inverse variance weighted meta-analysis on the summary statistics from the UK Biobank GWAS and a meta-analysis by the

CKDGen Consortium of the GWAS of eGFRcr including up to 567,460 EA individuals.[13,175] As the CKDGen consortium included the ARIC study (N=9,037), we adjusted the effect sizes of SNPs by removing the ARIC participants ($\beta_{\text{corrected}} = \frac{\left(\frac{1}{\text{SE}^2_{\text{CKDGen}}} \cdot \beta_{\text{CKDGen}}\right) - \left(\frac{1}{\text{SE}^2_{\text{ARIC}}} \cdot \beta_{\text{ARIC}}\right)}{\frac{1}{\text{SE}^2_{\text{CKDGen}}} - \frac{1}{\text{SE}^2_{\text{ARIC}}}}$; $\text{SE}_{\text{corrected}} = \sqrt{\frac{1}{\frac{1}{\text{SE}^2_{\text{CKDGen}}} - \frac{1}{\text{SE}^2_{\text{ARIC}}}}}$).[176] We also used a sample of 489 unrelated EA individuals from phase 3 1000 Genomes as a linkage disequilibrium (LD) reference panel for the score construction step.[177] Approximately 1.2 million common (MAF ≥1%) variants in HapMap3 were kept for score construction, as suggested in Vilhjálmsson et al.[36,178] We computed PRS in three ways: LDpred, pruning and threshold (P+T), and a simple weighted combination of SNPs that reached genome-wide significance in our meta-analysis combining UK Biobank and CKDGen, a special case of P+T.

The primary PRS was calculated using the LDpred algorithm.[36] For this method, we created 5 candidate LDpred PRSs under different assumptions for the fraction of causal variants. This Bayesian approach utilizes GWAS summary statistics to compute the posterior mean effect sizes for the genetic variants by assuming a prior of the joint effect sizes and incorporating the LD structure of the reference population. Two parameters of the LDpred need to be set by the users. One is the LD radius, which is the number of variants being adjusted for at each side of a variant. We set it to 400 (which corresponds to $1.2\times 10^6/3{,}000$) based on Vilhjálmsson et al. The other parameter is the fraction of causal variants, ρ, which can be selected via parameter tuning on a separate dataset. Our tested ρ values were 1, 0.3, 0.1, 0.03, and 0.01, as suggested in Vilhjálmsson et al.[36]

We also implemented a second approach named pruning and thresholding. P+T scores were constructed with applying two filtering steps based on LD and *P* value.[179] The variants are first pruned to only keep variants that have absolute pairwise correlation weaker than a threshold, r^2, within certain genetic distance. The remaining variants are then filtered by removing the ones that have a *P* value larger than a pre-defined threshold of significance. We created 24 candidate P+T PRSs based on four r^2 levels (0.2, 0.4, 0.6, and 0.8) and six *P* values (5×10^{-8}, 5×10^{-6}, 5×10^{-4}, 0.05, 0.5, and 1). Finally, we created a "simple PRS" in a similar manner, using the most commonly used r^2 level, 0.1, and *P* value threshold, 5×10^{-8} (genome-wide significance).

For the PRS tuning, the 5 candidate LDpred PRSs, 24 candidate P+T PRSs, and one simple PRS were calculated in a tuning dataset of the remaining 10% unrelated EA participants in the UK Biobank (N=32,159). The best PRS of each approach was determined based on the proportion of the variance explained (R^2) of eGFRcr that can be explained by the PRS. Specifically, we fitted a linear regression model with eGFRcr being the outcome, each candidate PRS being the exposure, and age at baseline and sex as the covariates. The best LDpred PRS and P+T PRS, as well as the simple PRS were carried forward into subsequent analyses in an independent validation dataset.

PRS validation was conducted in the 8,866 unrelated EA participants in ARIC. The R^2 for eGFRcr by the best LDpred PRS, best P+T PRS, and simple PRS were calculated using the same approach with adjustment for the same covariates as in the tuning step. We compared the three PRSs with respect to number of SNPs included, phenotypic variance explained, and correlations with each other**.** In sensitivity analysis, we also directly implemented the PRS constructed and tuned on EA participants to the 2,871 unrelated AA participants in ARIC.

Assessing Incident Kidney Diseases

Four incident kidney diseases were included in our study as outcomes: chronic kidney disease (CKD), end-stage kidney disease (ESKD), kidney failure, and acute kidney injury (AKI). CKD was defined based on the following criteria: eGFR <60 mL/min/1.73 m^2 plus ≥30% eGFR decline during a follow-up visit comparing to baseline, ESKD cases identified through the direct linkage to the US Renal Data System (USRDS) registry, or International Classification of Diseases (ICD)-9/10-Clinical Modification (CM) codes **(Supplemental Table 1A)** representing CKD in any position of hospitalization or death records.[180] ESKD was defined as having kidney transplant or dialysis in the USRDS registry. Kidney failure was defined by hospitalization codes (**Supplemental Table 1B**). AKI was defined by hospitalization or death codes (ICD-9-CM code: 584.X or ICD-10-CM code: N17.x).[181]

Protein Measurements

Plasma proteins were measured in ARIC participants at visit 3 and visit 5 using the SOMAscan v.4 assay by SomaLogic. This platform uses Slow Off-rate Modified Aptamers (SOMAmers) to bind to targeted proteins and then uses DNA microarray to quantify them. SOMAscan v.4 includes 4,931 unique human proteins or protein complexes, with 95% of the proteins tagged by one modified aptamer and a total 5,211 modified aptamers. Protein measurements were reported as relative fluorescence units (RFUs).[21] There were no missing values in the proteomic data. Details of the quality control of the proteins were described elsewhere.[164] Previous studies of SOMAscan v.3 consisting of 4,001 aptamers have shown high precision of this assay in

quantifying proteins with median coefficient of variance (CV) of 4%-8%.[182-184] In the current study, all proteins were log2 transformed.

Assessing Covariates

Information on age, sex, center, and education level were assessed at baseline and current smoking status was assessed at all visits using an interviewer-administered questionnaires.[167] Body mass index (BMI) was calculated as weight (in kilograms) divided by the square of height (in meters), both of which were measured at all visits. Clinical factors included history of hypertension, diabetes, and coronary heart disease (CHD). Hypertension was defined at all visits as systolic blood pressure ≥140 mm Hg, diastolic blood pressure ≥ 90 mm Hg, or use of antihypertensive medication in the past 2 weeks. Diabetes was defined at all visits as fasting blood glucose ≥126 mg/dL, non-fasting glucose ≥200 mg/dL, self-reported doctor-diagnosed diabetes, or use of diabetes medication in the past 2 weeks. CHD was defined at all visits as prior myocardial infarction (MI) observed on ECG, self-reported doctor-diagnosed heart attack, or self-reported cardiovascular surgery or coronary angioplasty. Albumin to creatinine ratio (ACR) was calculated as urinary albumin divided by urinary creatinine, with albumin being measured by an immunoturbidimetric method and creatinine being measured by a modified kinetic Jaffé method.

Statistical Analysis

Baseline characteristics of the primary study population were examined. The R^2 for eGFRcr at all visits except for visit 7 by the LDpred PRS, P+T PRS, and simple PRS were calculated as $\frac{\text{Var(PRS)}\times Coefficient_{PRS}^2}{\text{Var(eGFRcr)}}$ in a linear regression model of eGFRcr adjusting for age at the

corresponded visit, sex, center, and first 10 genetic principal components (PCs). We also calculated the R^2 for eGFRcys at all visits except for visit 1 and visit 7. For comparison, we calculated the R^2 for eGFRcr and eGFRcys by the PRS using the same methods among the 2,871 AA participants as a sensitivity analysis.

We evaluated the association between PRS and incident kidney diseases. Using Cox proportional hazard models, we estimated hazard ratios (HR) and associated 95% confidence intervals (CI) of PRSs (per 1 SD lower PRS) for incident kidney diseases outcomes. We considered time at risk to start at visit 1 (1987–1989) and continue until the event of interest, death, loss to follow-up , or the end of follow-up (December 31, 2018). We evaluated three models: Model 1, which included age, sex, center, and first 10 genetic PCs; Model 2, which additionally included education, baseline BMI, baseline smoking status, baseline history of hypertension, diabetes, and CHD. In sensitivity analysis, we also evaluated additional adjustments for eGFRcr and for both eGFRcr and ACR. As ACR was first measured at visit 4, the time to event for this sensitivity analysis started at visit 4 and the baseline covariates were also assessed at this time. Time to incident kidney diseases was assessed among deciles of PRS using proportional hazard models and displayed using Kaplan-Meier survival curves.

To evaluate the association between PRS for kidney function and proteomic measurements, we conducted linear regression of LDpred PRS on 4,877 proteins measured at visit 3 and visit 5 adjusting for age, sex, center, and first 10 genetic PCs. These estimates reflect the difference in each log(2) transformed protein per normalized SD-unit higher in PRS for kidney function. Given that multiple statistical tests were performed, we utilized a Bonferroni adjusted P-value

threshold of 0.05/4,877 ≈ 1.2 x 10^{-5} to indicate evidence for significant associations. We identified proteins significantly associated with LDpred PRS at both visit 3 and visit 5. We then examined their correlations and associations with eGFRcr and eGFRcys at each visit through Pearson correlation matrix and linear regression of PRS on eGFRcr or eGFRcys with adjustment for age, sex, center, and first 10 genetic PCs. Then we made scatter plots of correlations between those proteins and eGFR at each visit against their Pearson correlations with PRS. Analyses used R version 3.6.2 software (The R Foundation), two-tailed P-values, as well as statistical significance level of $P < 0.05$ except for the identification of proteomic signals, which was $P < 1.02 \times 10^{-5}$.

RESULTS

Characteristics of Study Cohort

Our primary study population included 8,886 participants (mean age 54.3 years; 53% female). Around 40% of them received college or above education. At baseline, 25% were smokers; mean BMI was 27.0 kg/m^2; and the percentage of participants with prevalent hypertension, diabetes, and CHD were 26.7%, 8.6%, and 5.1% respectively. Over 30 years of follow up, the number of incident chronic kidney disease (CKD), end stage kidney disease (ESKD), kidney failure, and acute kidney injury (AKI) were 2,959, 137, 470, and 1,723, respectively (**Table 1**).

Characteristics of the Polygenic Risk Scores

LDpred PRS, P+T PRS, and simple PRS were all standardized to be approximately normally distributed in the population with zero-mean and unit-variance. The technical details of the three PRSs are summarized in **Supplemental Table 2** and described in detail elsewhere.[185] LDpred

PRS was highly correlated with P+T PRS with a Pearson correlation coefficient (r) of 0.84, but moderately correlated with the simple PRS (r=0.58, **Supplemental Figure 2).** The adjusted eGFRcr variance explained by the LDpred PRS for kidney function was relatively consistent across all visits over 30 years of follow-up, ranging from 6.8% to 10.3%. P+T PRS and simple PRS explained lower adjusted eGFRcr variance (P+T PRS: 5.4% to 8.8%; simple PRS: 4.5% to 6.6%). The adjusted eGFRcys variance was lower and its ranges for LDpred PRS, P+T PRS and simple PRS were 2.7% to 4.0%, 1.9% to 3.3, and 1.5% to 2.6%, respectively. Variance explained for eGFR based on creatinine and cystatin (eGFRcr-cys) was intermediate **(Supplemental Table 3).** As a comparison, directly applying the PRS trained and tuned on EA participants to AA participants led to substantial decrease of score performance, with the eGFRcr variances explained by the LDpred PRS, P+T PRS, and simple PRS ranging from 1.7% to 3.1%, 0.86% to 2.1%, and 0.91% to 2.3% respectively (**Supplemental Table 4**).

Associations between PRSs for Kidney Function and Incident Kidney Diseases

Categorizing the PRSs into deciles showed an incremental association with risk in Kaplan-Meier survival curves **(Figure 2).** In continuous analysis, we observed that the LDpred PRS for kidney function was strongly associated with all four incident kidney diseases: HRs (95% CI) per 1 SD-unit lower in LDpred PRS, indicating worse kidney function, were 1.33 (1.28, 1.39), 1.20 (1.00, 1.42), 1.17 (1.06, 1.28), and 1.07 (1.02, 1.12) for incident CKD, ESKD, kidney failure, and AKI respectively, after adjusting for age at baseline, sex, center, first 10 genetic PCs. Using P+T PRS and simple PRS, HRs for all incident kidney diseases were of smaller magnitude than using LDpred PRS and only statistically significant for CKD and kidney failure **(Table 2, Supplemental Figure 3).**

After adjustment for lifestyle and clinical risk factors (education, baseline BMI, baseline smoking status, and hypertension, diabetes, and CHD history at baseline) in Cox models, we observed limited changes in the risk estimates of the PRSs for all incident kidney diseases. However, risk estimates were substantially attenuated after additionally adjusting for the mediator eGFRcr **(Supplemental Table 5)**. Sensitivity analyses showed that additional adjustment for ACR made little difference **(Supplemental Table 6).**

Plasma Proteome as An Intermediate Trait

Using linear regression models adjusted for age, sex, center, and first 10 genetic PCs, we observed that 183 proteins were associated with LDpred PRS for kidney function at $P = 1.2\times10^{-5}$ level among 7,213 participants with valid proteomic measurements at visit 3, and 138 proteins among 3,666 participants at visit 5. Among those proteins, 108 were significant at both visits, which are 20 years apart. The strongest associations were with cystatin-C, collagen alpha-1(XV) chain, and desmocollin-2**.** Collagen alpha-1 (XV) chain exhibited strong and consistent associations with both eGFRcr and eGFRcys with a magnitude similar to that of cystatin with eGFRcr**.** For the 108 proteins consistently associated with LDpred PRS for kidney function, most of the associations were negative, indicating higher protein levels with lower kidney function. Testican 2 and angiostatin were the only two proteins with significant positive correlation to kidney function. The correlations with eGFRcr and eGFRcys measured at the corresponding visits were much stronger than those with the LDpred PRS, especially at visit 5 **(Figure 3, Supplemental Table 7,** median negative correlations at visit 5 of -0.0855, -0.4668, and -0.4697 with LDpred PRS, eGFRcr and eGFRcys with corresponding values at visit 3 of -0.0679, -0.2639

and -0.2820). This was also true for the positive correlations; the Pearson correlation coefficients with PRS, eGFRcr, and eGFRcys at visit3 were 0.100, 0.195, 0.197 for testican 2 and 0.067, 0.167, 0.258 for angiostatin respectively, and 0.103, 0.398, 0.433 for testican 2 and 0.095, 0.273, 0.344 for angiostatin respectively at visit 5.

DISCUSSION

In this community-based deeply phenotyped cohort of 8,866 middle-aged adults, we leveraged large studies to construct a range of polygenic risk scores for kidney function. A genome-wide score that included a weighted average of 1.2 million SNPs (LDpred PRS) showed the strongest association with kidney function while narrower risk scores (P+T PRS and simple PRS) showed weaker associations. The kidney function LDpred PRS was also associated with a range of plasma protein levels in mid-life and older age. However, these associations were weaker than the protein associations with eGFR itself. This may be due to many protein elevations being secondary to reduced kidney function. Genetic susceptibility to lower kidney function was associated with higher level of all significant proteins except testican 2 and angiostatin, which had lower levels at lower kidney function. Examination of future risk showed genetic susceptibility to lower kidney function was associated with higher risk of CKD, ESKD, kidney failure, and AKI. LDpred PRS showed the strongest association, which was mediated by eGFR and independent of all other measured clinical and lifestyle risk factors including albuminuria.

An advantage of PRS for kidney function is that it can be assessed at any time, well before the emergence of lifestyle and clinical risk factors, such as elevated BMI, hypertension, and diabetes. Our results demonstrate that, for a spectrum of kidney diseases, not only diseases with

established high heritability, but also entities like AKI whose genetic basis is less pronounced,[158-161,166] PRS can now identify individuals with higher genetic risk for over 30 years of follow up, suggesting its potential role in further research and clinical medicine.

During the last decade, GWAS studies demonstrated thousands of genetic loci associated with hundreds of phenotypes.[186] However, for most traits, the heritability explained by those SNPs (h^2_{gwas}) only explains a small portion of the estimated proportion of phenotypic variance due to additive genetic effects, i.e., narrow-sense heritability (h^2).[187] One of the proposed reasons for that was the existence of common causal variants of exceedingly low effect size which requires extremely large sample sizes to detect via GWAS.[188-192] Using large studies for discovery and algorithms that incorporate variants across the genome, our results showed a significant improvement in the performance of PRS compared with previous efforts for score development,[14-19] allowing us to explain more variance in kidney function and a link between the genetic basis of eGFR and a spectrum of incident kidney disease outcomes.

In addition to the improved algorithm of constructing PRS, two other factors were important for improving the prediction performance of PRS. In recent years, mega cohorts and global genetics consortia provided sufficient *power* for detecting loci that confer only small changes in disease risk, which was a key factor in improving accuracy of PRS for disease prediction. Another factor is accurate identification of disease. In our study, we benefited from linkage to the USRDS registry and algorithms developed based on best possible clinical and epidemiological evidence for identifying cases of incident kidney diseases, which also improved the power of our analysis.

Differences in plasma protein levels can provide clues for intermediate pathways between genetic susceptibility and disease. These alterations can happen as a result of the genetic susceptibility itself or secondary to other physiologic changes, including reduced eGFR. Our finding that many proteins were negatively associated with kidney function suggests that low renal clearance may result in relative protein accumulation. On the other hand, increased protein levels in the setting of kidney disease may reflect ongoing pathological processes, such as inflammation. Specific proteins are of interest. Collagen alpha-1 (XV) chain showed strong and consistent associations with PRS and both eGFRcr and eGFRcys with a magnitude similar to that of cystatin, the best marker for kidney function at current practice, which suggested its potential role as a marker for genetically predicted kidney function. Testican 2, one of two proteins that was lower with lower kidney function, forms a structural component of the extracellular matrix through covalently binding with glycosaminoglycans. [193] It is expressed in multiple tissues including the kidney and genetic variants in its gene, *SPOCK2*, have been strongly associated with bronchopulmonary dysplasia but the connection to kidney disease is largely unknown. Angiostatin, the other protein that was lower with lower kidney function, is a potent angiogenesis inhibitor generated through the proteolysis of plasminogen. Evidence also suggests its anti-inflammatory roles through hindering the recruitment of leukocyte[194] and the movement of neutrophil and macrophage.[195,196] Alterations in angiogenesis and inflammation have important roles in kidney disease pathophysiology.[197,198] Previous animal experiments demonstrated that angiostatin overexpression slowed the progression of renal disease after chronic kidney injury and its decrease expression accelerated the pathogenesis process of diabetic nephropathy[199,200] Observational studies have suggested elevated urinary angiostatin as potential biomarker of the disease severity and progression for IgA nephropathy and lupus nephritis.[201,202]

Our study had limitations. The PRS developed in our study were constructed, tuned, and validated in EA participants only, mirroring most genetic studies done so far. Since LD patterns, minor allele frequencies, effect sizes of common variants, and phenotypic features vary by ancestry, our PRS constructed based on GWAS results and LD structure of EA individuals will have poor disease prediction for individuals with other ancestries.[203,204] When directly applying our PRS trained and tuned on EA participants to AA participants, the phenotypic variance explained was four-fold lower. It is therefore necessary and important for future efforts to include more multi-ethnic participants in genomic studies and develop novel methods that appropriately adjust for the difference across ethnic groups. The PRS presented in this study were for eGFRcr which means they may include genetic influences of creatinine metabolism as well as kidney function. However, we included eGFRcys as an outcome to assess the extent to which associations were robust to the kidney function marker used. Finally, we focused on common genetic variants in calculating the PRS recognizing future work may include additional variants, including rare variants with larger effects.

In conclusion, our results show polygenic risk scores for kidney function are associated with future risk of incident kidney diseases, including CKD progression, end-stage kidney disease, kidney failure, and acute kidney injury, over 30 years of follow up in a community-based cohort. This association was independent of most risk factors, including albuminuria, but largely mediated through kidney function itself. A number of plasma protein levels were elevated among individuals with high genetic risk for low kidney function. The protein associations were much stronger with concurrent kidney function than with PRS, consistent with protein elevations

secondary to the progressive kidney disease may be the dominant influence instead of representing the primary link of kidney function genetic susceptibility to the proteome.

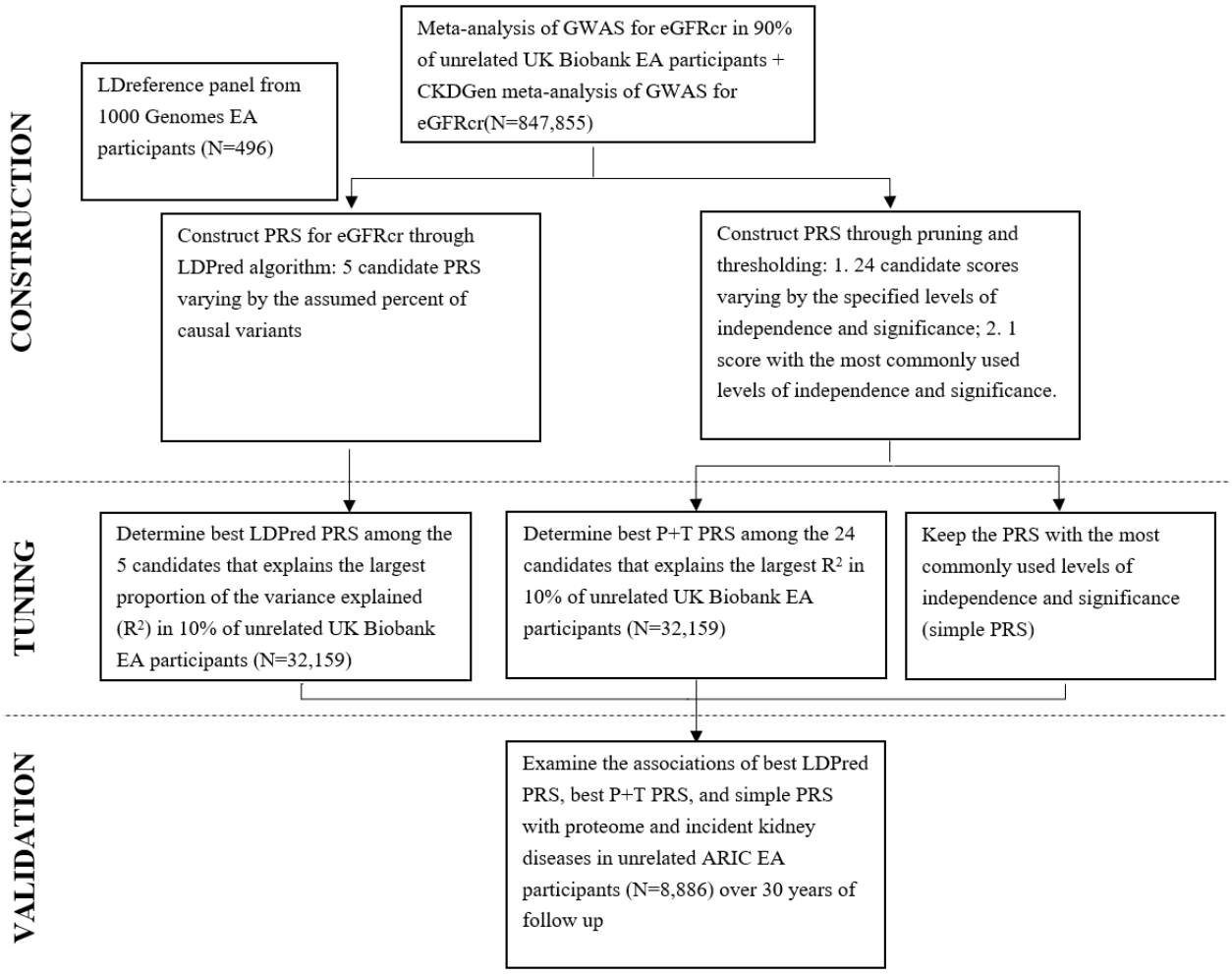


Figure 5-1 Study design and workflow.

Three polygenic risk scores (PRS) for kidney function measured as estimated glomerular filtration rate based on creatinine level (eGFRcr) were constructed by a European-ancestry (EA) meta-analysis of UK Biobank GWAS for eGFRcr (90% of the cohort) and a meta-analysis of GWAS for eGFRcr conducted by the CKDGen Consortium using LDpred algorithm, pruning and threshold (P+T), and a simple weighted combination of SNPs that reached genome-wide significance, followed by parameters tuned using data from the remaining 10% of UK Biobank EA participants, then tested for their associations with proteome and incident kidney diseases in the Atherosclerosis Risk in Communities (ARIC) study.

Table 5-1 Characteristics of the study population (N=8,886).

Characteristic[a]	
Age, years[b]	54.3 (5.7)
Female	4,708 (53.0)
Advanced education	3,513 (39.6)
Current smokers[b]	2,192 (24.7)
Body mass index, kg/m^2 [b]	27.0 (4.8)
History of hypertension[b]	2,363 (26.7)
History of diabetes[b]	764 (8.6)
History of CHD[b]	445 (5.1)
eGFRcr, mL/min/1.73 m^2[b]	99.6 (12.5)
Incident CKD	2959 (33.6)
Incident ESKD	137 (1.5)
Incident kidney failure	470 (5.3)
Incident AKI	1723 (19.4)

[a] Mean (standard deviation) for continuous variables and % (n) for categorical variables.

[b] Indicate baseline values

AKI, acute kidney injury; ARIC, Atherosclerosis Risk in Communities; CHD, coronary heart disease; CKD, Chronic kidney disease; eGFRcr, estimated glomerular filtration rate based on creatinine; ESKD, end stage kidney disease; PRS, polygenic risk score.

Table 5-2 Risk for incident kidney diseases according to polygenic risk scores of kidney function (N=8,886).

		Risk for incident kidney diseases per 1 SD lower in PRS					
		Ldpred PRS[a]		P+T PRS[a]		Simple PRS[a]	
		Hazard ratios (95% CI)	P value	Hazard ratios (95% CI)	P value	Hazard ratios (95% CI)	P value
Chronic kidney disease	**Model 1[b]**	1.33 (1.28, 1.39)	5.46E-50	1.28 (1.23, 1.32)	1.34E-37	1.23 (1.18, 1.27)	8.22E-28
	Model 2[b]	1.33 (1.28, 1.38)	8.91E-47	1.28 (1.23, 1.33)	3.01E-37	1.23 (1.19, 1.28)	8.83E-28
End stage kidney disease	**Model 1[b]**	1.20 (1.00, 1.42)	0.04	1.14 (0.96, 1.36)	0.13	1.05 (0.89, 1.25)	0.55
	Model 2[b]	1.21 (1.01, 1.45)	0.04	1.18 (0.98, 1.41)	0.08	1.09 (0.91, 1.29)	0.36
Kidney failure	**Model 1[b]**	1.17 (1.06, 1.28)	1.32E-03	1.12 (1.02, 1.23)	0.02	1.12 (1.03, 1.23)	0.01
	Model 2[b]	1.16 (1.06, 1.28)	2.24E-03	1.13 (1.03, 1.24)	0.01	1.13 (1.03, 1.24)	8.31E-03
Acute kidney injury	**Model 1[b]**	1.07 (1.02, 1.12)	9.77E-03	1.03 (0.98, 1.08)	0.21	1.00 (0.96, 1.05)	0.86
	Model 2[b]	1.06 (1.01, 1.12)	0.02	1.03 (0.98, 1.08)	0.27	1.01 (0.96, 1.06)	0.82

[a] LDPred PRS was constructed using LDPred algorithm, a Bayesian approach utilizes GWAS summary statistics to compute the posterior mean effect sizes for the genetic variants by assuming a prior of the joint effect sizes and incorporating the LD structure of the reference population. P+T PRS was constructed using 'pruning and thresholding (P+T)', which first prunes variants to only keep

those who have absolute pairwise correlation weaker than a threshold within certain genetic distance and then filtered variants that have a *P* value larger than a pre-defined threshold of significance. Simple score was constructed using the most commonly used level of absolute pairwise correlation for pruning and genome-wide significance level for thresholding.

[b] Model 1 adjusted for age at baseline, sex, center, and first 10 genetic principal components; Model 2 adjusted for all covariates in model 1 and education, baseline body mass index, baseline smoking status, baseline history of hypertension, diabetes, and coronary heart disease.

PRS: polygenic risk score.

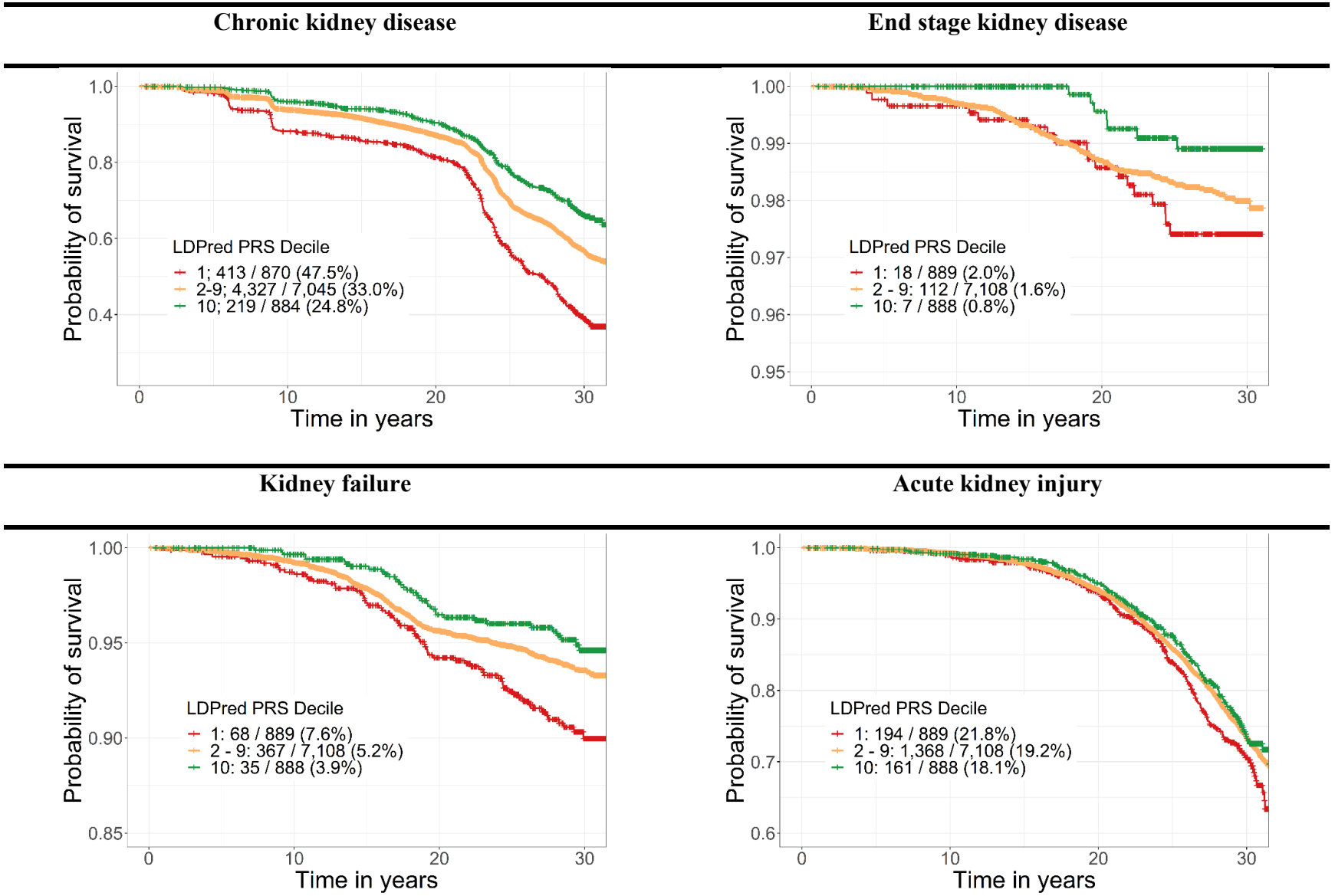
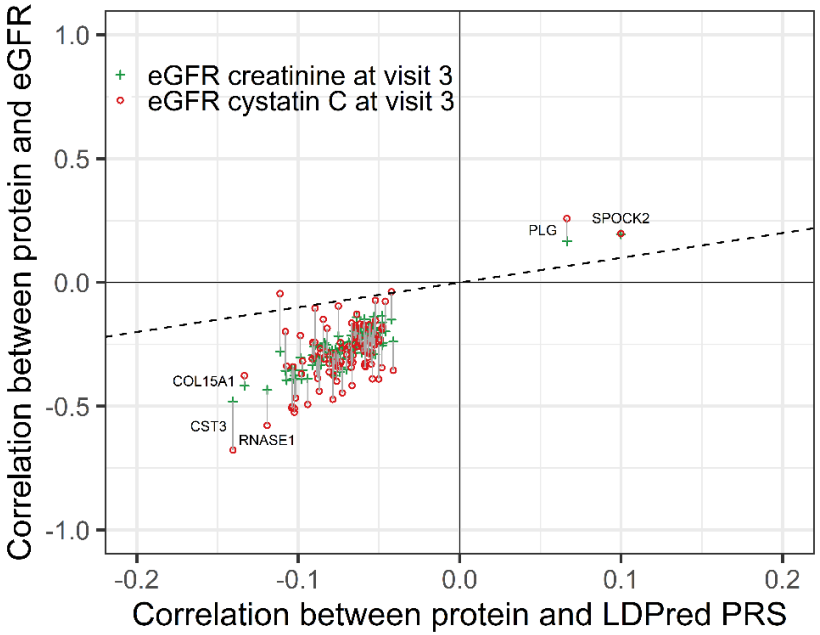
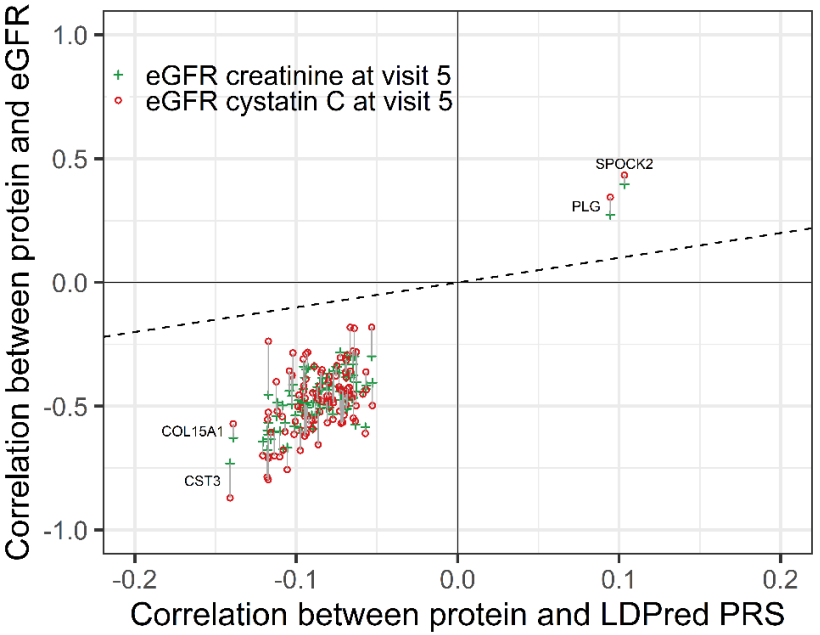


Figure 5-2 Association of deciles of LDPred polygenic risk score of kidney function with incident kidney diseases (N=8,886).

The LDPred polygenic risk score (PRS) for kidney function was categoried into three levels 1st decile, 2nd – 9th deciles, and 10th decile, and was examined for their unadjusted associations with incident kidney diseases over 30-year of follow-up. P-values of the score (log-rank) test were 3.43E-27, 0.10, 0.002, and 0.08 for incident chronic kidney disease, end stage kidney disease, kidney failure, and acute kidney injury respectively.



(A)



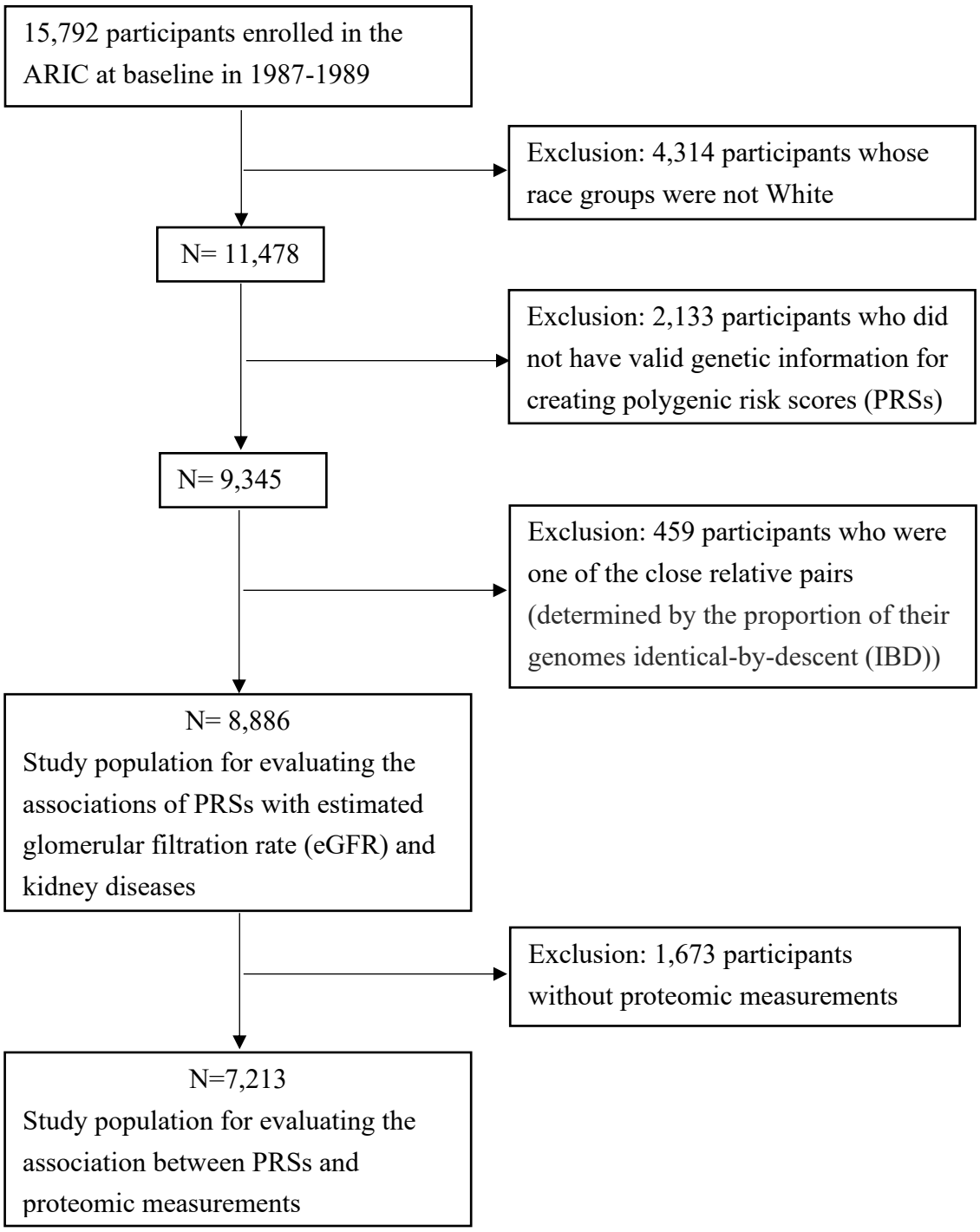
(B)

Figure 5-3 Scatter plots of correlations between protein and LDPred polygenic risk score for kidney function and correlation between protein and estimated glomerular filtration rate. Both protein measures and eGFR are visit specific (panel A – visit 3; panel B – visit 5).

A total of 108 proteins were identified as significantly (Bonferroni threshold $p < 1.02 \times 10^{-5}$) associated with LDPred PRS at both visit 3 and visit 5 through linear regression of LDPred PRS on 4,877 proteins adjusting for age at the corresponded visits, sex, center, and first 10 genetic principal components. Visit 3 (N=7,213) was conducted during 1993-1995 when the mean age of study population was 60.4 years and visit 5 (N=3,666) was conducted during 2011-2013 when the mean age of study population was 75.9 years. The dashed line in grey is the identity line.

COL15A1: collagen alpha-1(XV) chain; CST3: cystatin-C; PLG: angiostatin; RNASE1: ribonuclease pancreatic; SPOCK2: testican-2.

Supplementary Figure 5-1 Flow chart of subjects included in the study.



Supplemental Table 5-1. ICD-9/10 codes used for identifying chronic kidney disease and kidney failure.

(A) Chronic kidney disease (CKD)

ICD-9-code	Description	ICD-10-code
582	Chronic glomerulonephritis	N03
583	Nephritis and nephropathy	
585, 585.x where x≥3	Chronic kidney disease	N18, N18.x where x≥3
586	Renal failure	N19
587	Renal sclerosis	N26
588	Disorders resulting from impaired renal function	N25
403	Hypertensive chronic kidney disease	I12
404	Hypertensive heart and kidney disease	I13
593.9	Unspecified disorder of the kidney and ureter	
250.4	Diabetes with renal complications	E10.2, E11.2, E13.2
V42.0	Kidney replaced by transplant	Z94.0
55.6	Transplant of kidney	
996.81	Complications of transplanted kidney	

V45.1[a]	Renal dialysis status	Z99.2
V56 [a]	Admission for dialysis treatment or session	Z49
39.95[a]	Hemodialysis	
54.98[a]	Peritoneal dialysis	
	Encounter for adjustment and management of vascular access device	Z45.2

[a] Codes that are counted as incident CKD only if a concomitant acute kidney injury (AKI) code (ICD-9: 584.x, ICD-10: N17) is not present.

(B) Kidney failure

ICD-9-code	Description	ICD-10-code
V42.0	Kidney replaced by transplant	Z94.0
55.6	Transplant of kidney	
996.81	Complications of transplanted kidney	
V45.1[a]	Renal dialysis status	Z99.2
V56 [a]	Admission for dialysis treatment or session	Z49

39.95[a]	Hemodialysis	
54.98[a]	Peritoneal dialysis	
	Encounter for adjustment and management of vascular access device	Z45.2
585.5	Chronic kidney disease stage 5	N18.5
585.6	End stage renal disease	N18.6
586	Renal failure	N19
403.01	Hypertensive chronic kidney disease, malignant, with CKD 5 or ESRD	
403.91	Hypertensive chronic kidney disease, with CKD 5 or ESRD	I12.0

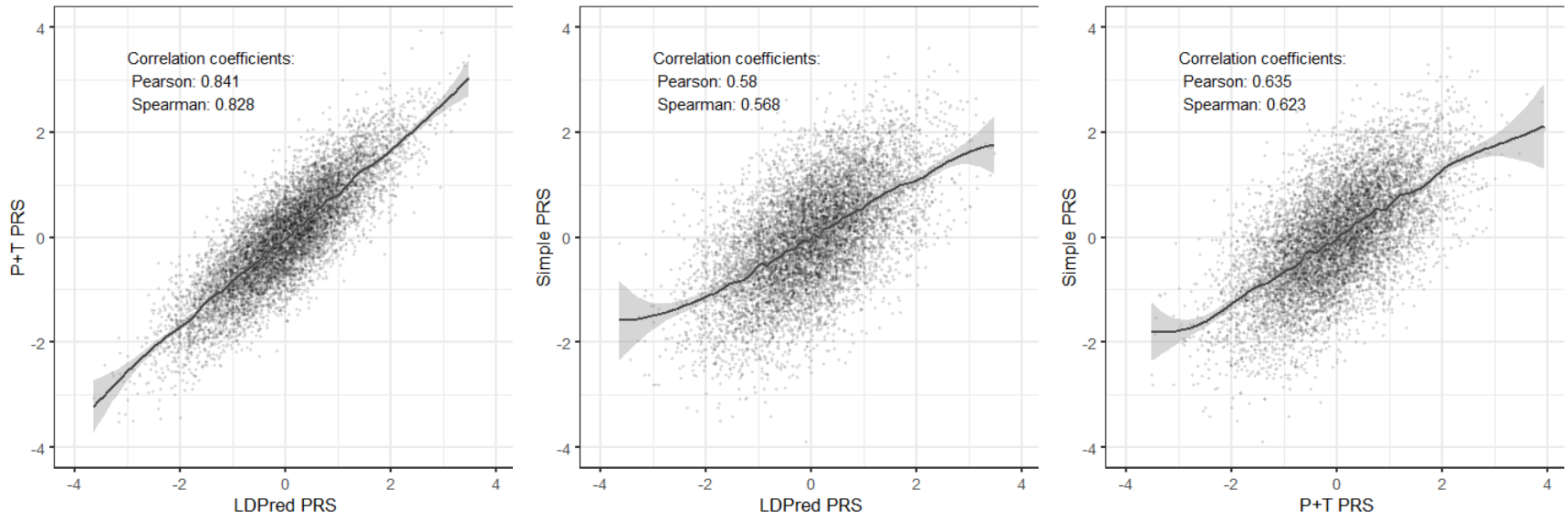
[a] Codes that are not counted as incident kidney failure if: 1) for hospitalizations, a concurrent AKI code is present; 2) for deaths, if a concurrent AKI code is present without a concurrent CKD code.

Supplemental Table 5-2. Technical details of polygenic risk scores.

	LDPred PRS	P+T PRS	Simple PRS
Features	Entire summary results of variants with rescaling weights based on LD structure, effect size, and estimated causal fraction	r^2 and P value thresholds to restrict variants	r^2 and genome-wide significant threshold to restrict variants
Settings	LD structure from 1000 Genome % of causal variants = 30%	$r^2 = 0.2$ P = 0.05	$r^2 = 0.1$ $P = 5 * 10^{-8}$
No. of candidate SNPs	~ 1.2 million	~ 1.2 million	~ 1.2 million
No. of included SNPs	~ 1.2 million	36,944	1,022
Rescaling weights	Yes	No	No

PRS: polygenic risk score.

P + T PRS[a] vs. LDPred PRS[a]	Simple PRS[a] vs. LDPred PRS	Simple PRS vs. P + T PRS



Supplemental Figure 5-2. Scatter plots of polygenic risk scores (PRS) with locally weighted smoothing (LOESS) regression line.

LDPred PRS was constructed using LDPred algorithm, a Bayesian approach utilizes GWAS summary statistics to compute the posterior mean effect sizes for the genetic variants by assuming a prior of the joint effect sizes and incorporating the LD structure of the reference population. P+T PRS was constructed using 'pruning and thresholding (P+T)', which first prunes variants to only keep those who have absolute pairwise correlation weaker than a threshold within certain genetic distance and then filtered variants that

have a *P* value larger than a pre-defined threshold of significance. Simple score was constructed using the most commonly used level of absolute pairwise correlation for pruning and genome-wide significance level for thresholding.

Supplemental Table 5-3. Adjusted proportion of the variance for estimated glomerular filtration rate (eGFR) explained by polygenic risk scores (PRS) (N = 8,886).

	eGFRcr			eGFRcys			eGFRcr-cys		
	LDPred PRS[a]	**P+T PRS[a]**	**Simple PRS[a]**	**LDPred PRS[a]**	**P+T PRS[a]**	**Simple PRS[a]**	**LDPred PRS[a]**	**P+T PRS[a]**	**Simple PRS[a]**
Visit 1 (1987-1989)	0.078[b]	0.061	0.051	-	-	-	-	-	-
Visit 2 (1990-1992)	0.082	0.071	0.055	0.033	0.027	0.022	0.067	0.057	0.044
Visit 3 (1993-1995)	0.103	0.088	0.066	0.040	0.033	0.026	0.079	0.066	0.051
Visit 4 (1996-1998)	0.084	0.07	0.048	0.032	0.025	0.022	0.067	0.054	0.042
Visit 5 (2011-2013)	0.068	0.054	0.045	0.027	0.019	0.015	0.048	0.037	0.030
Visit 6 (2016-2017)	0.073	0.062	0.051	0.032	0.027	0.021	0.054	0.046	0.037

[a] LDPred PRS was constructed using LDPred algorithm, a Bayesian approach utilizes GWAS summary statistics to compute the posterior mean effect sizes for the genetic variants by assuming a prior of the joint effect sizes and incorporating the LD structure of the reference population. P+T PRS was constructed using 'pruning and thresholding (P+T)', which first prunes variants to only keep

those who have absolute pairwise correlation weaker than a threshold within certain genetic distance and then filtered variants that have a *P* value larger than a pre-defined threshold of significance. Simple score was constructed using the most commonly used level of absolute pairwise correlation for pruning and genome-wide significance level for thresholding.

[b] Proportion of the variance for eGFR explained by PRS with adjusting for age at the corresponded visit, sex, center, and first 10 genetic principal components for all such values.

eGFRcr, estimated glomerular filtration rate based on creatinine; eGFRcr-cys, estimated glomerular filtration rate based on creatinine and cystatin; eGFRcysr, estimated glomerular filtration rate based on cystatin.

Supplemental Table 5-4. Adjusted proportion of the variance for estimated glomerular filtration rate (eGFR) explained by polygenic risk scores (PRS) among participants with African ancestry (N = 2,871).

	eGFRcr			eGFRcys			eGFRcr-cys		
	LDPred PRS[a]	**P+T PRS[a]**	**Simple PRS[a]**	**LDPred PRS[a]**	**P+T PRS[a]**	**Simple PRS[a]**	**LDPred PRS[a]**	**P+T PRS[a]**	**Simple PRS[a]**
Visit 1 (1987-1989)	0.017[b]	0.010	0.008	-	-	-	-	-	-
Visit 2 (1990-1992)	0.025	0.016	0.011	0.012	0.008	0.005	0.026	0.018	0.012
Visit 3 (1993-1995)	0.025	0.018	0.016	0.016	0.013	0.007	0.025	0.019	0.013
Visit 4 (1996-1998)	0.022	0.021	0.013	0.010	0.013	0.006	0.019	0.022	0.011
Visit 5 (2011-2013)	0.018	0.009	0.009	0.010	0.008	0.009	0.015	0.009	0.010
Visit 6 (2016-2017)	0.031	0.009	0.023	0.009	0.005	0.017	0.024	0.096	0.028

[a] LDPred PRS was constructed using LDPred algorithm, a Bayesian approach utilizes GWAS summary statistics to compute the posterior mean effect sizes for the genetic variants by assuming a prior of the joint effect sizes and incorporating the LD structure of the reference population. P+T PRS was constructed using 'pruning and thresholding (P+T)', which first prunes variants to only keep

those who have absolute pairwise correlation weaker than a threshold within certain genetic distance and then filtered variants that have a *P* value larger than a pre-defined threshold of significance. Simple score was constructed using the most commonly used level of absolute pairwise correlation for pruning and genome-wide significance level for thresholding.

[b] Proportion of the variance for eGFR explained by PRS with adjusting for age at the corresponded visit, sex, center, and first 10 genetic principal components for all such values.

eGFRcr, estimated glomerular filtration rate based on creatinine; eGFRcr-cys, estimated glomerular filtration rate based on creatinine and cystatin; eGFRcysr, estimated glomerular filtration rate based on cystatin.

Supplemental Table 5-5. Risk for incident kidney diseases adjusting for kidney fuction according to polygenic risk scores of kidney function (N=8,886).

	Risk for incident kidney diseases per 1 SD lower in PRS					
	Ldpred PRS[a]		P+T PRS[a]		Simple PRS[a]	
	Hazard ratios (95% CI)	P value	Hazard ratios (95% CI)	P value	Hazard ratios (95% CI)	P value
Chronic kidney disease[b]	1.21 (1.16, 1.26)	1.64E-19	1.17 (1.13, 1.22)	5.99E-15	1.13 (1.09, 1.18)	2.26E-10
End stage kidney disease[b]	0.98 (0.81, 1.18)	0.81	0.97 (0.81, 1.18)	0.79	0.91 (0.76, 1.08)	0.29
Kidney failure[b]	1.03 (0.93, 1.14)	0.57	1.01 (0.92, 1.12)	0.79	1.03 (0.93, 1.13)	0.59
Acute kidney injury[b]	1.03 (0.98, 1.08)	0.30	1.00 (0.95, 1.05)	0.93	0.98 (0.93, 1.03)	0.35

[a] LDPred PRS was constructed using LDPred algorithm, a Bayesian approach utilizes GWAS summary statistics to compute the posterior mean effect sizes for the genetic variants by assuming a prior of the joint effect sizes and incorporating the LD structure of the reference population. P+T PRS was constructed using 'pruning and thresholding (P+T)', which first prunes variants to only keep those who have absolute pairwise correlation weaker than a threshold within certain genetic distance and then filtered variants that have a *P* value larger than a pre-defined threshold of significance. Simple score was constructed using the most commonly used level of absolute pairwise correlation for pruning and genome-wide significance level for thresholding

[b] Model adjusted for age at baseline, sex, center, first 10 genetic principal components, education, baseline body mass index, baseline smoking status, baseline history of hypertension, diabetes, and coronary heart disease, and estimated glomerular filtration rate based on creatinine (eGFRcr) measured at baseline.

PRS: polygenic risk score.

Supplemental Table 5-6. Risk for incident kidney diseases according to polygenic risk scores of kidney function after visit 4 (N=6,719).

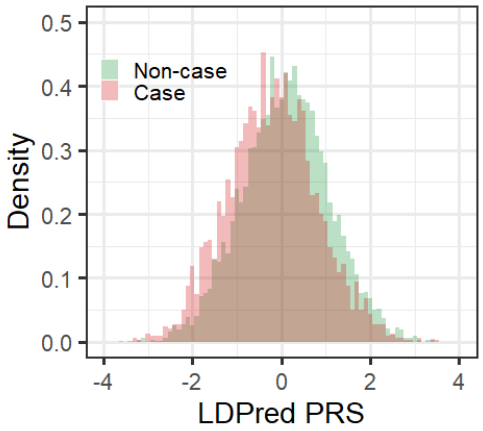
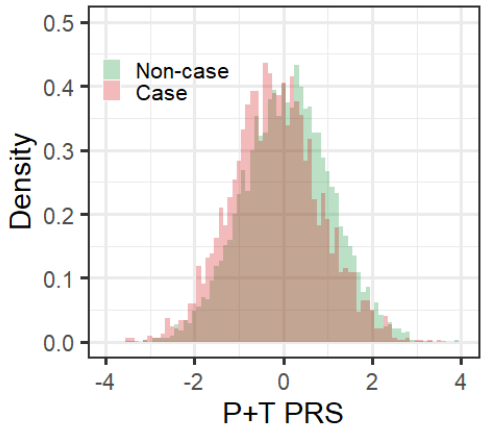
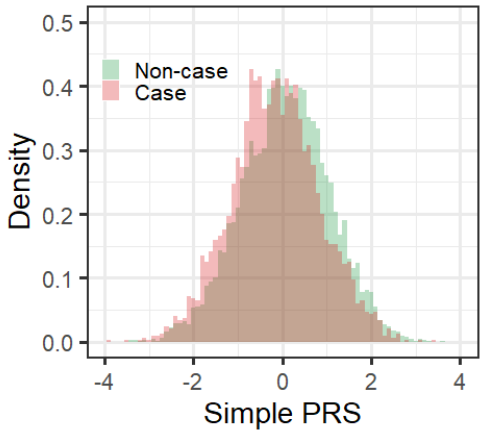
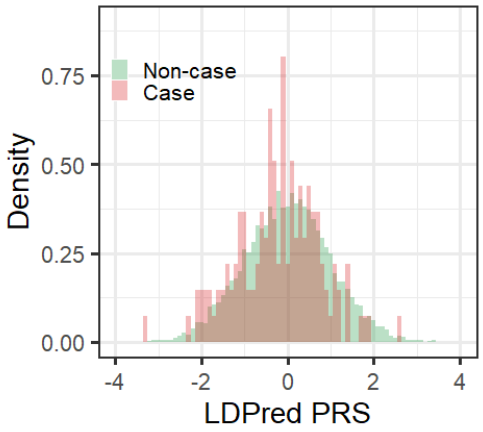
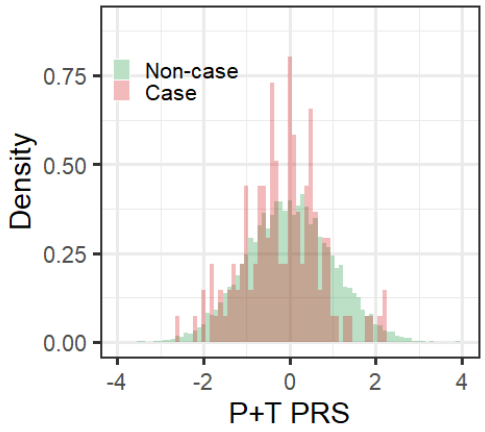
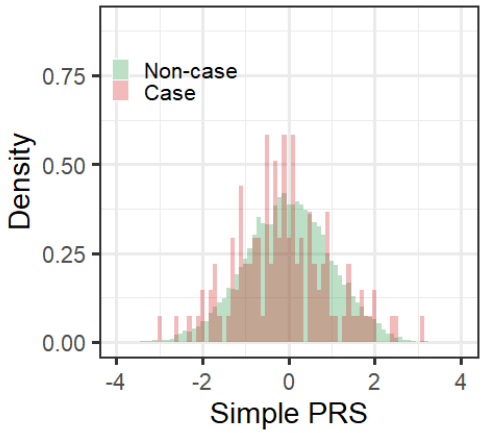
		Risk for incident kidney diseases per 1 SD lower in PRS					
		Ldpred PRS[a]		P+T PRS[a]		Simple PRS[a]	
		Hazard ratios (95% CI)	P value	Hazard ratios (95% CI)	P value	Hazard ratios (95% CI)	P value
Chronic kidney disease	**Model 1[b]**	1.25 (1.20, 1.31)	5.75E-22	1.21 (1.15, 1.26)	1.93E-16	1.16 (1.11, 1.22)	2.13E-11
	Model 2[b]	1.25 (1.19, 1.31)	1.90E-21	1.21 (1.15, 1.26)	6.22E-16	1.17 (1.12, 1.22)	3.64E-12
	Model 3[b]	1.16 (1.10, 1.21)	3.75E-09	1.12 (1.07, 1.18)	2.03E-06	1.10 (1.05, 1.15)	6.15E-05
	Model 4[b]	1.16 (1.11, 1.22)	1.25E-09	1.13 (1.07, 1.18)	1.05E-06	1.11 (1.06, 1.16)	1.55E-05
End stage kidney disease	**Model 1[b]**	1.17 (0.95, 1.43)	1.32E-01	1.08 (0.89, 1.32)	4.48E-01	0.99 (0.81, 1.2)	8.83E-01
	Model 2[b]	1.15 (0.94, 1.41)	1.80E-01	1.09 (0.88, 1.33)	4.34E-01	1.00 (0.82, 1.22)	9.99E-01
	Model 3[b]	0.78 (0.62, 0.97)	2.84E-02	0.78 (0.63, 0.96)	1.97E-02	0.79 (0.65, 0.97)	2.06E-02
	Model 4[b]	0.81 (0.65, 1.02)	7.80E-02	0.79 (0.64, 0.98)	3.16E-02	0.81 (0.66, 0.99)	4.38E-02
Kidney failure	**Model 1[b]**	1.19 (1.07, 1.33)	1.37E-03	1.15 (1.03, 1.27)	1.24E-02	1.12 (1.00, 1.24)	4.23E-02
	Model 2[b]	1.19 (1.07, 1.33)	1.98E-03	1.15 (1.03, 1.28)	1.46E-02	1.11 (1.00, 1.24)	4.69E-02
	Model 3[b]	0.96 (0.85, 1.08)	4.69E-01	0.94 (0.84, 1.06)	3.24E-01	0.96 (0.86, 1.07)	4.56E-01

	Model 4[b]	0.98 (0.87, 1.10)	7.06E-01	0.96 (0.86, 1.08)	4.89E-01	0.98 (0.88, 1.09)	7.16E-01
Acute kidney injury	**Model 1[b]**	1.07 (1.01, 1.13)	1.51E-02	1.04 (0.99, 1.10)	1.53E-01	1.01 (0.96, 1.07)	6.13E-01
	Model 2[b]	1.07 (1.02, 1.13)	1.27E-02	1.04 (0.99, 1.10)	1.38E-01	1.01 (0.96, 1.07)	6.46E-01
	Model 3[b]	1.00 (0.94, 1.06)	8.92E-01	0.97 (0.92, 1.03)	3.15E-01	0.96 (0.91, 1.01)	1.02E-01
	Model 4[b]	1.00 (0.94, 1.06)	9.69E-01	0.97 (0.92, 1.03)	3.56E-01	0.96 (0.91, 1.01)	1.43E-01

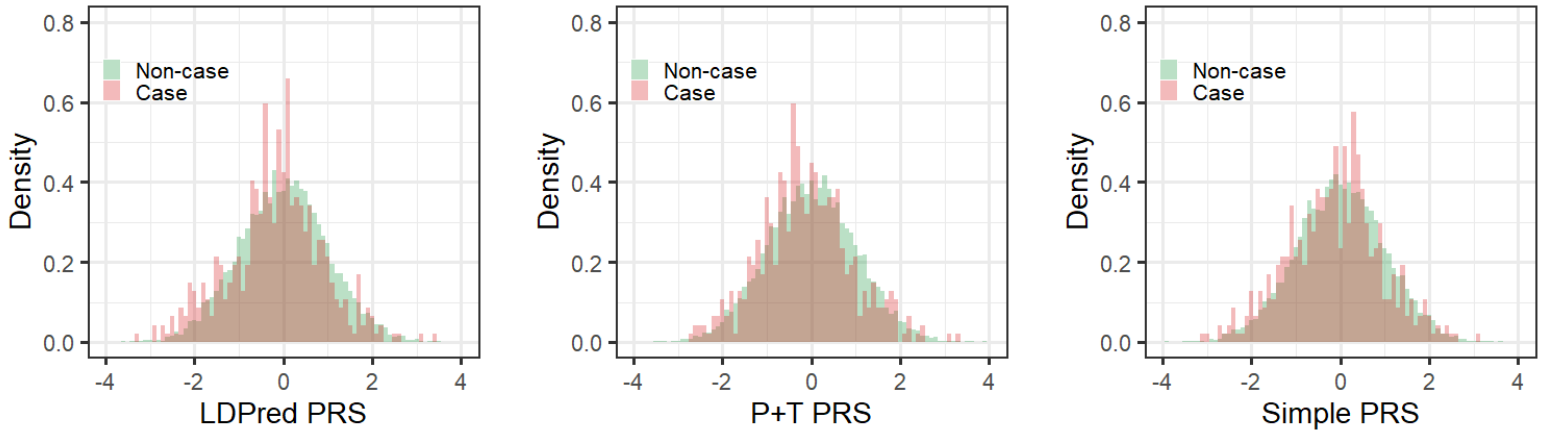
[a] LDPred PRS was constructed using LDPred algorithm, a Bayesian approach utilizes GWAS summary statistics to compute the posterior mean effect sizes for the genetic variants by assuming a prior of the joint effect sizes and incorporating the LD structure of the reference population. P+T PRS was constructed using 'pruning and thresholding (P+T)', which first prunes variants to only keep those who have absolute pairwise correlation weaker than a threshold within certain genetic distance and then filtered variants that have a *P* value larger than a pre-defined threshold of significance. Simple score was constructed using the most commonly used level of absolute pairwise correlation for pruning and genome-wide significance level for thresholding

[b] Model 1 adjusted for age at visit 4, sex, center, and first 10 genetic principal components; Model 2 adjusted for all covariates in model 1 and education, body mass index at visit 4, smoking status at visit 4, history of hypertension, diabetes, and coronary heart disease at visit 4; Model 3 adjusted for all covariates in model 2 and albumin-to-creatinine ratio (ACR) measured at visit 4; Model 4 adjusted for all covariates in model 3 and estimated glomerular filtration rate based on creatinine (eGFRcr) measured at visit 4.

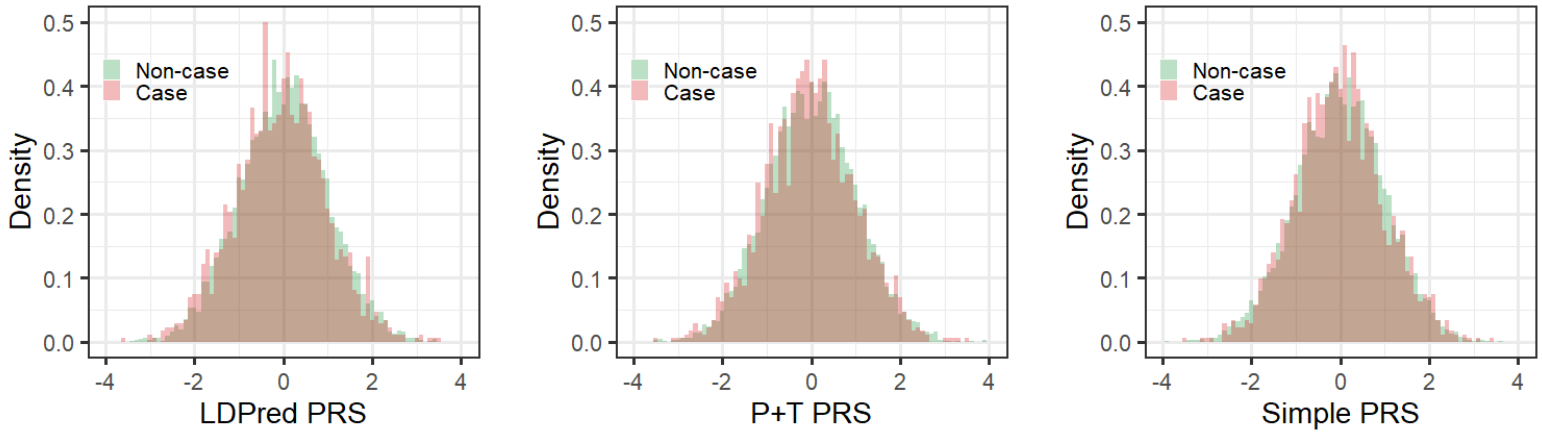
PRS: polygenic risk score

	LDPred PRS	P+T PRS	Simple PRS
Chronic kidney disease			
End stage kidney disease			

Kidney failure



Acute kidney injury



Supplemental Figure 5-3. Polygenic risk scores density distribution by incident kidney diseases status.

LDPred PRS was constructed using LDPred algorithm, a Bayesian approach utilizes GWAS summary statistics to compute the posterior mean effect sizes for the genetic variants by assuming a prior of the joint effect sizes and incorporating the LD structure of the reference population. P+T PRS was constructed using 'pruning and thresholding (P+T)', which first prunes variants to only keep those who have absolute pairwise correlation weaker than a threshold within certain genetic distance and then filtered variants that have a *P* value larger than a pre-defined threshold of significance. Simple score was constructed using the most commonly used level of absolute pairwise correlation for pruning and genome-wide significance level for thresholding. PRS: polygenic risk score

Supplemental Table 5-7. Associations of LDPred polygenic risk score for kidney function and estimated glomerular filtration rate with proteins that are significantly associated with LDPred polygenic risk score at both visit 3 (N=7,213) and visit 5 (N = 3,666)[a].

(A) LDPred PRS, eGFR measured at visit 3, and proteins measured at visit 3

Gene	Full name	Visit 3 Protein and LDPred PRS[b,c]			Visit 3 Protein and Visit 3 eGFRcr[c]			Visit 3 Protein and Visit 3 eGFRcys[c]		
		Correlation[e]	Beta (SE)	P	Correlation[e]	Beta (SE)	P	Correlation[e]	Beta (SE)	P
					Positive significant correlations					
SPOCK2	Testican-2	0.0999	0.0369 (0.0039)	6.96E-21	0.1951	0.0054 (0.0003)	8.77E-68	0.1972	0.0042 (0.0002)	9.18E-71
PLG	Angiostatin	0.0665	0.0153 (0.0028)	3.59E-08	0.1669	0.0021 (0.0002)	8.06E-22	0.2579	0.0033 (0.0002)	3.44E-86
Positive significant correlations, median[f]			0.0832			0.1810			0.2276	
					Negative significant correlations					
CST3	Cystatin-C	-0.1405	-0.0395 (0.003)	7.62E-39	-0.4811	-0.009 (0.0002)	0	-0.677	-0.0104 (0.0001)	0
COL15A1	Collagen alpha-1(XV) chain	-0.1334	-0.0351 (0.0029)	4.28E-34	-0.4171	-0.0084 (0.0002)	0	-0.3766	-0.0058 (0.0002)	2.50E-254
RNASE1	Ribonuclease pancreatic	-0.1193	-0.0707 (0.0064)	3.06E-28	-0.4341	-0.0166 (0.0005)	2.00E-253	-0.577	-0.0179 (0.0003)	0
DSC2	Desmocollin-2	-0.1031	-0.0305 (0.003)	2.07E-24	-0.3896	-0.0076 (0.0002)	4.40E-240	-0.3408	-0.0049 (0.0002)	1.20E-165
COL6A3	Collagen alpha-3(VI) chain	-0.104	-0.0299 (0.0029)	2.18E-24	-0.3501	-0.0064 (0.0002)	1.30E-185	-0.5068	-0.0073 (0.0002)	0
COL28A1	Collagen alpha-1(XXVIII) chain	-0.1024	-0.0321 (0.0032)	8.27E-24	-0.3534	-0.0071 (0.0002)	7.20E-181	-0.5101	-0.0081 (0.0002)	0
CD59	CD59 glycoprotein	-0.1072	-0.031 (0.0031)	1.09E-23	-0.3941	-0.0078 (0.0002)	1.80E-234	-0.3382	-0.0049 (0.0002)	1.70E-154
TMED10	Transmembrane emp24 domain-containing protein 10	-0.1024	-0.0359 (0.0036)	2.63E-23	-0.3768	-0.0087 (0.0003)	2.90E-213	-0.5245	-0.0095 (0.0002)	0
GM2A	Ganglioside GM2 activator	-0.1016	-0.0329 (0.0033)	1.05E-22	-0.3897	-0.0082 (0.0002)	3.40E-219	-0.4669	-0.0076 (0.0002)	0

TNFRSF1A	Tumor necrosis factor receptor superfamily member 1A	-0.1038	-0.0389 (0.004)	2.43E-22	-0.393	-0.0095 (0.0003)	4.10E-210	-0.5008	-0.0097 (0.0002)	0
ART3	Ecto-ADP-ribosyltransferase 3	-0.1113	-0.046 (0.0048)	1.26E-21	-0.2788	-0.0101 (0.0004)	2.30E-161	-0.045	-0.0015 (0.0003)	4.25E-07
CDNF	Cerebral dopamine neurotrophic factor	-0.1079	-0.0286 (0.0031)	1.54E-20	-0.3566	-0.0073 (0.0002)	4.10E-209	-0.1983	-0.0029 (0.0002)	6.99E-55
SELM	Selenoprotein M	-0.098	-0.0314 (0.0034)	7.35E-20	-0.3905	-0.0081 (0.0003)	2.40E-208	-0.3691	-0.0058 (0.0002)	5.70E-174
EFNB2	Ephrin-B2	-0.0973	-0.0318 (0.0036)	7.58E-19	-0.3532	-0.0082 (0.0003)	1.70E-195	-0.3181	-0.0056 (0.0002)	2.50E-151
B2M	Beta-2-microglobulin	-0.0941	-0.0274 (0.0032)	8.66E-18	-0.388	-0.0074 (0.0002)	1.50E-201	-0.4931	-0.0076 (0.0002)	0
MYOC	Myocilin	-0.0986	-0.0352 (0.0041)	2.21E-17	-0.3036	-0.0082 (0.0003)	1.40E-143	-0.2149	-0.0043 (0.0002)	5.08E-67
FABP4	Fatty acid-binding protein, adipocyte	-0.0879	-0.0522 (0.0062)	5.45E-17	-0.3044	-0.0124 (0.0005)	2.90E-145	-0.3874	-0.0121 (0.0004)	7.20E-238
LMAN2	Vesicular integral-membrane protein VIP36	-0.0884	-0.0248 (0.003)	9.20E-17	-0.3218	-0.0061 (0.0002)	5.20E-155	-0.3681	-0.0054 (0.0002)	1.40E-206
MB	Myoglobin	-0.0895	-0.0412 (0.005)	1.35E-16	-0.2509	-0.0093 (0.0004)	1.10E-127	-0.1043	-0.0029 (0.0003)	4.92E-22
IGFBP6	Insulin-like growth factor-binding protein 6	-0.0896	-0.0231 (0.0028)	2.04E-16	-0.3139	-0.0057 (0.0002)	6.10E-150	-0.2422	-0.0032 (0.0002)	8.14E-80
LCN2	Neutrophil gelatinase-associated lipocalin	-0.0901	-0.0365 (0.0045)	4.91E-16	-0.2589	-0.0074 (0.0003)	5.99E-98	-0.2976	-0.0068 (0.0003)	4.20E-139
HSPB6	Heat shock protein beta-6	-0.0913	-0.0384 (0.0047)	6.33E-16	-0.3341	-0.0091 (0.0004)	6.50E-134	-0.3086	-0.0063 (0.0003)	5.10E-107
MFAP2	Microfibrillar-associated protein 2	-0.0826	-0.0262 (0.0033)	2.19E-15	-0.2509	-0.0051 (0.0003)	2.52E-85	-0.28	-0.0043 (0.0002)	6.80E-104
FSTL3	Follistatin-related protein 3	-0.0726	-0.0248 (0.0032)	7.64E-15	-0.3423	-0.0065 (0.0002)	8.10E-154	-0.4465	-0.0067 (0.0002)	4.50E-286
RGMB	RGM domain family member B	-0.0845	-0.0212 (0.0027)	8.31E-15	-0.3045	-0.0058 (0.0002)	3.00E-165	-0.1494	-0.0021 (0.0002)	2.02E-36

ESAM	Endothelial cell-selective adhesion molecule	-0.091	-0.0254 (0.0033)	1.36E-14	-0.2935	-0.006 (0.0003)	2.90E-119	-0.2428	-0.0039 (0.0002)	6.59E-86
CFD	Complement factor D	-0.0792	-0.0256 (0.0033)	1.45E-14	-0.2749	-0.0057 (0.0003)	1.10E-107	-0.2996	-0.0046 (0.0002)	2.00E-119
JAM2	Junctional adhesion molecule B	-0.0863	-0.0202 (0.0026)	2.38E-14	-0.2633	-0.0043 (0.0002)	2.81E-94	-0.2662	-0.0034 (0.0002)	2.20E-99
NBL1	Neuroblastoma suppressor of tumorigenicity 1	-0.0781	-0.0311 (0.0041)	3.11E-14	-0.3766	-0.0095 (0.0003)	2.90E-201	-0.3767	-0.0073 (0.0002)	7.00E-199
SMOC2	SPARC-related modular calcium-binding protein 2	-0.082	-0.0272 (0.0036)	4.14E-14	-0.2827	-0.0066 (0.0003)	3.10E-121	-0.1853	-0.0031 (0.0002)	1.82E-44
TNFRSF1B	Tumor necrosis factor receptor superfamily member 1B	-0.0785	-0.0286 (0.0038)	4.52E-14	-0.3007	-0.0069 (0.0003)	5.30E-120	-0.4723	-0.0088 (0.0002)	0
GAS1	Growth arrest-specific protein 1	-0.086	-0.0256 (0.0034)	7.70E-14	-0.3326	-0.0069 (0.0003)	2.00E-146	-0.2843	-0.0043 (0.0002)	3.87E-94
COL18A1	Endostatin	-0.0729	-0.0262 (0.0035)	1.40E-13	-0.2665	-0.0061 (0.0003)	3.70E-108	-0.2652	-0.0045 (0.0002)	4.13E-98
UNC5B	Netrin receptor UNC5B	-0.0771	-0.0257 (0.0035)	1.80E-13	-0.3227	-0.0066 (0.0003)	1.60E-132	-0.3293	-0.0052 (0.0002)	2.60E-134
PPIC	Peptidyl-prolyl cis-trans isomerase C	-0.0771	-0.0262 (0.0036)	1.85E-13	-0.2719	-0.0063 (0.0003)	7.40E-114	-0.3251	-0.0059 (0.0002)	2.10E-168
RNASE6	Ribonuclease K6	-0.0762	-0.0367 (0.005)	2.36E-13	-0.2812	-0.0085 (0.0004)	1.40E-105	-0.3993	-0.0097 (0.0003)	2.00E-237
GABARAP	Gamma-aminobutyric acid receptor-associated protein	-0.0869	-0.0195 (0.0027)	2.87E-13	-0.3085	-0.005 (0.0002)	7.90E-126	-0.4391	-0.0058 (0.0001)	6.80E-308
UNC5C	Netrin receptor UNC5C	-0.0739	-0.0243 (0.0034)	4.32E-13	-0.247	-0.0049 (0.0003)	3.57E-78	-0.243	-0.0036 (0.0002)	7.68E-71
PTGDS	Prostaglandin-H2 D-isomerase	-0.0742	-0.0323 (0.0045)	6.61E-13	-0.3607	-0.0093 (0.0003)	5.10E-157	-0.3399	-0.0065 (0.0003)	9.80E-131
DNAJB12	DnaJ homolog subfamily B member 12	-0.0807	-0.0313 (0.0044)	7.55E-13	-0.319	-0.0086 (0.0003)	3.10E-143	-0.3283	-0.0068 (0.0003)	4.40E-152

WFDC1	WAP four-disulfide core domain protein 1	-0.0777	-0.0298 (0.0042)	7.86E-13	-0.3137	-0.0077 (0.0003)	1.70E-125	-0.3575	-0.007 (0.0002)	8.00E-177
CRABP2	Cellular retinoic acid-binding protein 2	-0.0809	-0.0327 (0.0046)	8.24E-13	-0.28	-0.0083 (0.0003)	1.80E-119	-0.3613	-0.0084 (0.0003)	1.30E-211
CCL14	C-C motif chemokine 14	-0.0735	-0.0391 (0.0055)	9.77E-13	-0.2645	-0.0092 (0.0004)	5.30E-104	-0.3223	-0.0085 (0.0003)	1.70E-149
MIA	Melanoma-derived growth regulatory protein	-0.0751	-0.026 (0.0036)	9.93E-13	-0.216	-0.0053 (0.0003)	2.00E-77	-0.0955	-0.0015 (0.0002)	4.48E-12
RARRES2	Retinoic acid receptor responder protein 2	-0.0671	-0.0266 (0.0038)	1.69E-12	-0.2127	-0.0054 (0.0003)	2.32E-73	-0.3435	-0.0067 (0.0002)	2.00E-200
EFNA5	Ephrin-A5	-0.0778	-0.0211 (0.003)	2.39E-12	-0.3063	-0.0058 (0.0002)	1.60E-135	-0.2857	-0.004 (0.0002)	5.20E-109
ATOX1	Copper transport protein ATOX1	-0.0839	-0.0348 (0.005)	5.50E-12	-0.2437	-0.0073 (0.0004)	1.26E-75	-0.3091	-0.008 (0.0003)	2.80E-155
UNC5B	Netrin receptor UNC5B	-0.0757	-0.0252 (0.0037)	6.46E-12	-0.289	-0.0062 (0.0003)	2.40E-104	-0.3195	-0.0053 (0.0002)	2.20E-130
EFNA4	Ephrin-A4	-0.0751	-0.0196 (0.0029)	1.06E-11	-0.2856	-0.0052 (0.0002)	4.20E-119	-0.3333	-0.0046 (0.0002)	2.20E-164
TAGLN	Transgelin	-0.07	-0.0279 (0.0041)	1.50E-11	-0.3513	-0.0075 (0.0003)	2.80E-118	-0.2919	-0.0041 (0.0002)	1.86E-59
TWSG1	Twisted gastrulation protein homolog 1	-0.0663	-0.0154 (0.0023)	1.51E-11	-0.2736	-0.0038 (0.0002)	1.00E-101	-0.2726	-0.0029 (0.0001)	5.80E-101
CD46	Membrane cofactor protein	-0.0688	-0.0181 (0.0027)	2.16E-11	-0.2421	-0.0043 (0.0002)	4.39E-90	-0.2577	-0.0034 (0.0002)	2.76E-97
TNFRSF1B	Tumor necrosis factor receptor superfamily member 1B	-0.0666	-0.0256 (0.0038)	2.19E-11	-0.2628	-0.006 (0.0003)	2.29E-90	-0.417	-0.0078 (0.0002)	3.70E-267
AIF1L	Allograft inflammatory factor 1-like	-0.0737	-0.0213 (0.0032)	3.97E-11	-0.3092	-0.0054 (0.0002)	3.60E-102	-0.3205	-0.0044 (0.0002)	6.70E-113
CD55	Complement decay-accelerating factor	-0.0669	-0.0183 (0.0028)	8.47E-11	-0.2368	-0.0043 (0.0002)	2.38E-85	-0.1639	-0.0021 (0.0002)	6.61E-35

NPDC1	Neural proliferation differentiation and control protein 1	-0.0654	-0.0233 (0.0036)	9.86E-11	-0.1954	-0.0044 (0.0003)	1.04E-53	-0.2183	-0.0037 (0.0002)	1.74E-64
VWC2	Brorin	-0.0685	-0.0242 (0.0038)	1.13E-10	-0.2693	-0.0064 (0.0003)	1.10E-105	-0.2584	-0.0046 (0.0002)	6.11E-90
CALCOCO2	Calcium-binding and coiled-coil domain-containing protein 2	-0.0637	-0.0191 (0.003)	2.12E-10	-0.1397	-0.0032 (0.0002)	2.06E-40	-0.1282	-0.0022 (0.0002)	1.99E-33
EPHA2	Ephrin type-A receptor 2	-0.0673	-0.0237 (0.0037)	2.24E-10	-0.2754	-0.0059 (0.0003)	7.89E-91	-0.3183	-0.0054 (0.0002)	2.70E-128
EPHB6	Ephrin type-B receptor 6	-0.0618	-0.0168 (0.0027)	3.81E-10	-0.2274	-0.0044 (0.0002)	1.51E-98	-0.168	-0.0024 (0.0002)	2.98E-49
ROR2	Tyrosine-protein kinase transmembrane receptor ROR2	-0.0661	-0.0239 (0.0038)	3.89E-10	-0.286	-0.0066 (0.0003)	5.70E-110	-0.2946	-0.0052 (0.0002)	1.90E-113
WFDC2	WAP four-disulfide core domain protein 2	-0.054	-0.0305 (0.0049)	5.69E-10	-0.2893	-0.0078 (0.0004)	7.60E-91	-0.3899	-0.0087 (0.0003)	2.10E-194
FABP3	Fatty acid-binding protein, heart	-0.0562	-0.0381 (0.0063)	1.89E-09	-0.1851	-0.0095 (0.0005)	1.34E-81	-0.2872	-0.0106 (0.0004)	4.00E-173
DLK1	Protein delta homolog 1	-0.0645	-0.0402 (0.0067)	2.24E-09	-0.2262	-0.0093 (0.0005)	5.42E-70	-0.1857	-0.0056 (0.0004)	5.76E-43
CD300C	CMRF35-like molecule 6	-0.0591	-0.0223 (0.0037)	2.59E-09	-0.148	-0.0035 (0.0003)	5.21E-32	-0.2778	-0.0053 (0.0002)	4.50E-121
EPHB4	Ephrin type-B receptor 4	-0.0635	-0.0215 (0.0036)	2.84E-09	-0.2543	-0.0056 (0.0003)	4.14E-86	-0.2586	-0.0044 (0.0002)	2.74E-89
DLK1	Protein delta homolog 1	-0.0647	-0.0374 (0.0063)	3.05E-09	-0.2185	-0.0085 (0.0005)	2.53E-66	-0.1743	-0.005 (0.0004)	3.03E-38
EFNA2	Ephrin-A2	-0.0662	-0.0194 (0.0033)	5.06E-09	-0.2839	-0.0058 (0.0003)	5.90E-112	-0.3221	-0.0051 (0.0002)	5.40E-145
LRP10	Low-density lipoprotein receptor-related protein 10	-0.0699	-0.016 (0.0027)	5.89E-09	-0.3042	-0.005 (0.0002)	1.70E-121	-0.269	-0.0033 (0.0002)	3.97E-89
IGFBP4	Insulin-like growth factor-binding protein 4	-0.0531	-0.0204 (0.0035)	1.02E-08	-0.1375	-0.0034 (0.0003)	1.79E-34	-0.2318	-0.0044 (0.0002)	1.62E-95

SERPINF1	Pigment epithelium-derived factor	-0.0612	-0.0126 (0.0022)	1.23E-08	-0.207	-0.0031 (0.0002)	1.49E-73	-0.2505	-0.003 (0.0001)	1.10E-113
MXRA7	Matrix-remodeling-associated protein 7	-0.0606	-0.0207 (0.0037)	1.70E-08	-0.2429	-0.0056 (0.0003)	1.92E-83	-0.234	-0.004 (0.0002)	1.97E-73
IL18BP	Interleukin-18-binding protein	-0.0585	-0.0225 (0.004)	1.82E-08	-0.21	-0.0044 (0.0003)	1.99E-45	-0.3404	-0.0062 (0.0002)	4.60E-152
CAPG	Macrophage-capping protein	-0.0575	-0.0335 (0.006)	2.14E-08	-0.1979	-0.0065 (0.0005)	4.18E-43	-0.2807	-0.0076 (0.0004)	5.60E-100
TNFRSF19	Tumor necrosis factor receptor superfamily member 19	-0.0604	-0.0232 (0.0042)	3.37E-08	-0.2841	-0.0074 (0.0003)	6.90E-111	-0.2617	-0.0051 (0.0003)	2.14E-89
VIT	Vitrin	-0.0651	-0.0222 (0.004)	3.44E-08	-0.2417	-0.0057 (0.0003)	7.07E-74	-0.2456	-0.0046 (0.0002)	4.26E-81
CPLX2	Complexin-2	-0.0561	-0.0249 (0.0045)	3.59E-08	-0.2839	-0.0081 (0.0003)	8.50E-117	-0.1717	-0.0031 (0.0003)	6.49E-29
RETN	Resistin	-0.0623	-0.0268 (0.0049)	4.85E-08	-0.2112	-0.0068 (0.0004)	1.81E-69	-0.2422	-0.006 (0.0003)	8.50E-91
ASGR1	Asialoglycoprotein receptor 1	-0.0567	-0.0184 (0.0034)	7.36E-08	-0.2175	-0.0048 (0.0003)	1.34E-70	-0.3015	-0.0052 (0.0002)	6.60E-142
TNFRSF21	Tumor necrosis factor receptor superfamily member 21	-0.0538	-0.0151 (0.0028)	8.82E-08	-0.1901	-0.0036 (0.0002)	2.12E-58	-0.2484	-0.0035 (0.0002)	6.94E-96
DCLK1	Serine/threonine-protein kinase DCLK1	-0.0634	-0.022 (0.0041)	9.64E-08	-0.1991	-0.0054 (0.0003)	6.46E-63	-0.2091	-0.0044 (0.0002)	3.35E-69
PXDN	Peroxidasin homolog	-0.0586	-0.0493 (0.0093)	1.09E-07	-0.2027	-0.011 (0.0007)	4.03E-51	-0.3277	-0.0149 (0.0005)	1.50E-160
EPHA1	Ephrin type-A receptor 1	-0.0521	-0.0314 (0.0059)	1.21E-07	-0.1813	-0.0066 (0.0005)	1.39E-44	-0.2125	-0.006 (0.0004)	5.37E-62
PI3	Elafin	-0.0493	-0.0319 (0.0061)	1.46E-07	-0.2037	-0.008 (0.0005)	6.52E-63	-0.2305	-0.0071 (0.0004)	2.58E-84
SRL	Sarcalumenin	-0.0595	-0.0252 (0.0048)	1.65E-07	-0.1838	-0.0056 (0.0004)	2.50E-49	-0.1708	-0.0039 (0.0003)	1.96E-41
IGFLR1	IGF-like family receptor 1	-0.05	-0.0278 (0.0053)	1.97E-07	-0.252	-0.0079 (0.0004)	7.49E-80	-0.3905	-0.01 (0.0003)	1.30E-222
DLK2	Protein delta homolog 2	-0.0525	-0.0227 (0.0044)	2.12E-07	-0.2905	-0.0078 (0.0003)	7.20E-117	-0.2236	-0.0044 (0.0003)	2.07E-62
VASN	Vasorin	-0.0643	-0.0138 (0.0027)	2.35E-07	-0.1882	-0.0033 (0.0002)	7.55E-57	-0.2238	-0.0031 (0.0002)	1.13E-85
CLMP	CXADR-like membrane protein	-0.05	-0.0228 (0.0045)	4.19E-07	-0.2243	-0.0067 (0.0003)	2.64E-82	-0.2388	-0.0053 (0.0003)	7.24E-87

NRXN3	Neurexin-3-beta	-0.0521	-0.0204 (0.0041)	4.75E-07	-0.1902	-0.0044 (0.0003)	1.22E-43	-0.154	-0.0024 (0.0002)	1.16E-22
XXYLT1	Xyloside xylosyltransferase 1	-0.0576	-0.0172 (0.0034)	4.86E-07	-0.2024	-0.0043 (0.0003)	7.71E-59	-0.207	-0.0033 (0.0002)	8.24E-59
MAP2K2	Dual specificity mitogen-activated protein kinase kinase 2	-0.0528	-0.0298 (0.006)	5.61E-07	-0.2286	-0.008 (0.0005)	3.14E-66	-0.2525	-0.0069 (0.0004)	1.28E-81
DDOST	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	-0.0571	-0.0163 (0.0033)	6.69E-07	-0.2293	-0.0046 (0.0003)	1.12E-70	-0.2301	-0.0036 (0.0002)	1.33E-72
B4GALT1	Beta-1,4-galactosyltransferase 1	-0.0562	-0.0118 (0.0024)	8.96E-07	-0.2378	-0.0036 (0.0002)	1.01E-81	-0.1896	-0.002 (0.0001)	7.78E-45
TREM1	Triggering receptor expressed on myeloid cells 1	-0.0547	-0.0218 (0.0045)	9.58E-07	-0.2204	-0.0054 (0.0003)	2.07E-52	-0.3302	-0.0068 (0.0003)	3.10E-144
MANSC1	MANSC domain-containing protein 1	-0.0482	-0.0184 (0.0038)	1.05E-06	-0.1609	-0.0039 (0.0003)	3.03E-39	-0.1886	-0.0035 (0.0002)	1.15E-52
IL15RA	Interleukin-15 receptor subunit alpha	-0.0481	-0.0213 (0.0044)	1.06E-06	-0.2545	-0.0061 (0.0003)	7.24E-72	-0.3445	-0.0069 (0.0003)	5.00E-154
CD93	Complement component C1q receptor	-0.0483	-0.0154 (0.0032)	1.09E-06	-0.1341	-0.0033 (0.0002)	1.77E-39	-0.1857	-0.0034 (0.0002)	2.93E-69
EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1	-0.0411	-0.0141 (0.0029)	1.11E-06	-0.2385	-0.0034 (0.0002)	3.41E-49	-0.3549	-0.0044 (0.0002)	2.20E-142
AMBP	Alpha-1-microglobulin	-0.0542	-0.0154 (0.0032)	1.20E-06	-0.2113	-0.004 (0.0002)	4.13E-57	-0.2833	-0.0043 (0.0002)	7.60E-114
TMPO	Lamina-associated polypeptide 2, isoforms beta/gamma	-0.0588	-0.0253 (0.0053)	1.92E-06	-0.2184	-0.0066 (0.0004)	1.38E-56	-0.336	-0.0088 (0.0003)	1.00E-171
NEGR1	Neuronal growth regulator 1	-0.0423	-0.0113 (0.0024)	3.64E-06	-0.1503	-0.0025 (0.0002)	1.51E-38	-0.0375	-0.0002 (0.0001)	0.253267

PENK	Proenkephalin-A	-0.0555	-0.0379 (0.0082)	4.19E-06	-0.1669	-0.007 (0.0006)	4.52E-27	-0.2399	-0.0086 (0.0005)	2.08E-67
SPINK7	Serine protease inhibitor Kazal-type 7	-0.0521	-0.0222 (0.0049)	5.39E-06	-0.1937	-0.0068 (0.0004)	1.47E-70	-0.0726	-0.002 (0.0003)	2.67E-11
SMOC1	SPARC-related modular calcium-binding protein 1	-0.0477	-0.0118 (0.0026)	7.27E-06	-0.2433	-0.0035 (0.0002)	7.45E-65	-0.1798	-0.0017 (0.0002)	4.78E-26
CST2	Cystatin-SA	-0.0461	-0.0337 (0.0076)	9.71E-06	-0.198	-0.0096 (0.0006)	9.57E-58	-0.0767	-0.0022 (0.0005)	1.47E-06
Negative significant correlations, median[f]			-0.0679			-0.2639			-0.2820	

(B) LDPred PRS, eGFR measured at visit 5, and proteins measured at visit 5

Gene	Full name	Visit 5 Protein and LDPred PRS[b,d]			Visit 5 Protein and Visit 5 eGFRcr[d]			Visit 5 Protein and Visit 5 eGFRcys[d]		
		Correlation	Beta (SE)	P	Correlation	Beta (SE)	P	Correlation	Beta (SE)	P
					Positive significant correlations					
SPOCK2	Testican-2	0.1033	0.0411 (0.0066)	7.20E-10	0.3976	0.0096 (0.0004)	6.00E-122	0.4333	0.009 (0.0003)	9.85E-147
PLG	Angiostatin	0.0946	0.0192 (0.0037)	1.66E-07	0.2734	0.0033 (0.0002)	7.74E-48	0.3442	0.0037 (0.0002)	7.82E-79
Positive significant correlations, median[f]			0.0989			0.3355			0.3887	
					Negative significant correlations					
COL15A1	Collagen alpha-1(XV) chain	-0.139	-0.0456 (0.0048)	2.18E-21	-0.6275	-0.0112 (0.0002)	0.00E+00	-0.5708	-0.0091 (0.0002)	1.93E-313
CST3	Cystatin-C	-0.141	-0.0535 (0.0056)	4.59E-21	-0.7304	-0.015 (0.0003)	0.00E+00	-0.8703	-0.0159 (0.0002)	0.00E+00
DSC2	Desmocollin-2	-0.1157	-0.0493 (0.0058)	3.72E-17	-0.6337	-0.0133 (0.0003)	0.00E+00	-0.6052	-0.0112 (0.0003)	8.89E-323
CD59	CD59 glycoprotein	-0.1177	-0.0433 (0.0051)	5.41E-17	-0.5965	-0.0111 (0.0003)	1.64E-304	-0.5533	-0.0091 (0.0002)	1.11E-266
TNFRSF1A	Tumor necrosis factor receptor superfamily member 1A	-0.1173	-0.0563 (0.0067)	9.16E-17	-0.6328	-0.0151 (0.0003)	0.00E+00	-0.71	-0.0151 (0.0003)	0.00E+00

GM2A	Ganglioside GM2 activator	-0.1207	-0.0542 (0.0065)	1.34E-16	-0.6421	-0.015 (0.0003)	0.00E+00	-0.6991	-0.0145 (0.0003)	0.00E+00
RNASE1	Ribonuclease pancreatic	-0.118	-0.104 (0.0126)	2.51E-16	-0.6764	-0.0303 (0.0006)	0.00E+00	-0.7862	-0.0313 (0.0004)	0.00E+00
COL6A3	Collagen alpha-3(VI) chain	-0.1173	-0.0483 (0.0059)	3.33E-16	-0.6145	-0.0129 (0.0003)	5.23E-317	-0.7037	-0.0132 (0.0002)	0.00E+00
TMED10	Transmembrane emp24 domain-containing protein 10	-0.1173	-0.0596 (0.0073)	4.59E-16	-0.6933	-0.018 (0.0003)	0.00E+00	-0.7963	-0.0183 (0.0003)	0.00E+00
ART3	Ecto-ADP-ribosyltransferase 3	-0.1173	-0.0564 (0.007)	7.09E-16	-0.4533	-0.0128 (0.0004)	4.54E-207	-0.2375	-0.0062 (0.0004)	3.15E-60
EFNB2	Ephrin-B2	-0.1172	-0.0469 (0.0058)	8.70E-16	-0.5655	-0.0118 (0.0003)	4.93E-262	-0.5251	-0.0097 (0.0003)	7.08E-230
COL28A1	Collagen alpha-1(XXVIII) chain	-0.1136	-0.0492 (0.0062)	2.47E-15	-0.6049	-0.0133 (0.0003)	3.74E-300	-0.7006	-0.0138 (0.0002)	0.00E+00
NBL1	Neuroblastoma suppressor of tumorigenicity 1	-0.1082	-0.0692 (0.0091)	3.25E-14	-0.6745	-0.0218 (0.0004)	0.00E+00	-0.6762	-0.019 (0.0004)	0.00E+00
MFAP2	Microfibrillar-associated protein 2	-0.1118	-0.0406 (0.0054)	6.35E-14	-0.4859	-0.009 (0.0003)	2.45E-169	-0.5207	-0.0086 (0.0003)	1.09E-206
PXDN	Peroxidasin homolog	-0.1102	-0.0894 (0.0119)	8.25E-14	-0.6003	-0.0255 (0.0006)	9.65E-299	-0.7044	-0.0265 (0.0005)	0.00E+00
CD55	Complement decay-accelerating factor	-0.1042	-0.0329 (0.0044)	1.03E-13	-0.438	-0.0068 (0.0003)	5.54E-143	-0.3575	-0.0049 (0.0002)	3.66E-96
TWSG1	Twisted gastrulation protein homolog 1	-0.1069	-0.0295 (0.004)	2.32E-13	-0.5667	-0.008 (0.0002)	8.36E-250	-0.6029	-0.0074 (0.0002)	4.31E-298
FSTL3	Follistatin-related protein 3	-0.0975	-0.0404 (0.0055)	2.43E-13	-0.5862	-0.0112 (0.0003)	1.06E-263	-0.6781	-0.0115 (0.0002)	0.00E+00
LMAN2	Vesicular integral-membrane protein VIP36	-0.1013	-0.0407 (0.0056)	2.86E-13	-0.5818	-0.0116 (0.0003)	2.60E-282	-0.6133	-0.0109 (0.0002)	0.00E+00
ROR2	Tyrosine-protein kinase transmembrane receptor ROR2	-0.1006	-0.0481 (0.0066)	4.42E-13	-0.5376	-0.0123 (0.0004)	3.76E-213	-0.5614	-0.0113 (0.0003)	6.21E-247
B2M	Beta-2-microglobulin	-0.1056	-0.0452 (0.0062)	4.69E-13	-0.6675	-0.0147 (0.0003)	0.00E+00	-0.7559	-0.0147 (0.0002)	0.00E+00
MB	Myoglobin	-0.102	-0.0543 (0.0075)	6.77E-13	-0.413	-0.0126 (0.0004)	1.27E-169	-0.2849	-0.008 (0.0004)	2.07E-86

LCN2	Neutrophil gelatinase-associated lipocalin	-0.1087	-0.0556 (0.0077)	8.19E-13	-0.4967	-0.0139 (0.0004)	1.12E-196	-0.5418	-0.0135 (0.0004)	2.24E-258
RGMB	RGM domain family member B	-0.1026	-0.0306 (0.0043)	8.48E-13	-0.492	-0.0075 (0.0002)	2.41E-192	-0.3779	-0.0051 (0.0002)	4.14E-110
CDNF	Cerebral dopamine neurotrophic factor	-0.1123	-0.0385 (0.0054)	1.10E-12	-0.5404	-0.0107 (0.0003)	4.47E-248	-0.4012	-0.0069 (0.0003)	3.14E-127
SELM	Selenoprotein M	-0.0965	-0.0423 (0.0059)	1.13E-12	-0.6084	-0.0126 (0.0003)	1.41E-295	-0.5979	-0.0109 (0.0003)	3.42E-292
SRL	Sarcalumenin	-0.1026	-0.0455 (0.0065)	3.01E-12	-0.458	-0.0108 (0.0004)	6.11E-167	-0.3738	-0.0078 (0.0003)	3.17E-111
IGFBP6	Insulin-like growth factor-binding protein 6	-0.0848	-0.0269 (0.0038)	3.15E-12	-0.4328	-0.0062 (0.0002)	1.11E-154	-0.3647	-0.0047 (0.0002)	8.52E-118
EPHA2	Ephrin type-A receptor 2	-0.0931	-0.043 (0.0064)	2.64E-11	-0.5039	-0.011 (0.0004)	4.60E-179	-0.539	-0.0105 (0.0003)	7.15E-222
TNFRSF1B	Tumor necrosis factor receptor superfamily member 1B	-0.0937	-0.0481 (0.0072)	3.21E-11	-0.4956	-0.0125 (0.0004)	4.63E-183	-0.609	-0.0138 (0.0003)	2.77E-322
EFNA5	Ephrin-A5	-0.0966	-0.0348 (0.0052)	3.30E-11	-0.5222	-0.0096 (0.0003)	7.39E-207	-0.5068	-0.0082 (0.0003)	9.21E-199
CD46	Membrane cofactor protein	-0.0989	-0.0289 (0.0043)	3.32E-11	-0.4917	-0.0074 (0.0002)	9.14E-176	-0.5017	-0.0066 (0.0002)	3.77E-186
COL18A1	Endostatin	-0.0921	-0.033 (0.0051)	1.07E-10	-0.5416	-0.0096 (0.0003)	3.16E-220	-0.5732	-0.0089 (0.0002)	1.16E-256
EFNA4	Ephrin-A4	-0.0904	-0.033 (0.0051)	1.16E-10	-0.5355	-0.0097 (0.0003)	8.30E-225	-0.5568	-0.009 (0.0002)	2.98E-262
TNFRSF1B	Tumor necrosis factor receptor superfamily member 1B	-0.0864	-0.0464 (0.0072)	1.71E-10	-0.5315	-0.0135 (0.0004)	2.83E-216	-0.6549	-0.0149 (0.0003)	0.00E+00
WFDC1	WAP four-disulfide core domain protein 1	-0.0947	-0.0421 (0.0066)	1.81E-10	-0.5379	-0.0123 (0.0004)	1.08E-216	-0.5405	-0.0108 (0.0003)	3.02E-224
UNC5B	Netrin receptor UNC5B	-0.0855	-0.0352 (0.0055)	2.30E-10	-0.5099	-0.0097 (0.0003)	1.73E-187	-0.5222	-0.0087 (0.0003)	1.15E-205
RETN	Resistin	-0.094	-0.0449 (0.0071)	2.36E-10	-0.3516	-0.0084 (0.0004)	2.02E-82	-0.3901	-0.0085 (0.0004)	1.30E-112
TNFRSF21	Tumor necrosis factor receptor superfamily member 21	-0.0856	-0.028 (0.0044)	2.57E-10	-0.4081	-0.0063 (0.0003)	3.85E-120	-0.419	-0.0057 (0.0002)	5.35E-134
MYOC	Myocilin	-0.0957	-0.0406 (0.0064)	3.24E-10	-0.4303	-0.0103 (0.0004)	4.10E-155	-0.3091	-0.0066 (0.0003)	1.50E-80

CFD	Complement factor D	-0.08	-0.0293 (0.0046)	3.47E-10	-0.4101	-0.0066 (0.0003)	2.02E-118	-0.4065	-0.0059 (0.0002)	1.72E-124
PTGDS	Prostaglandin-H2 D-isomerase	-0.0898	-0.0407 (0.0065)	4.88E-10	-0.5897	-0.0134 (0.0003)	3.74E-273	-0.5922	-0.0119 (0.0003)	1.74E-285
CD300C	CMRF35-like molecule 6	-0.0945	-0.0392 (0.0063)	5.02E-10	-0.3838	-0.0083 (0.0004)	7.48E-103	-0.4669	-0.009 (0.0003)	6.25E-167
FABP4	Fatty acid-binding protein, adipocyte	-0.0807	-0.0524 (0.0084)	6.21E-10	-0.4937	-0.0156 (0.0005)	9.64E-212	-0.5664	-0.0156 (0.0004)	5.28E-296
TNFRSF19	Tumor necrosis factor receptor superfamily member 19	-0.0898	-0.0458 (0.0074)	7.25E-10	-0.4861	-0.0126 (0.0004)	4.04E-175	-0.4458	-0.0101 (0.0004)	3.27E-147
ESAM	Endothelial cell-selective adhesion molecule	-0.0956	-0.0296 (0.0048)	8.90E-10	-0.4823	-0.0082 (0.0003)	7.00E-178	-0.4412	-0.0065 (0.0002)	1.19E-147
UNC5C	Netrin receptor UNC5C	-0.0802	-0.0343 (0.0056)	1.04E-09	-0.3872	-0.0073 (0.0003)	2.58E-100	-0.4023	-0.0068 (0.0003)	1.00E-114
SEMG2	Protein delta homolog 1	-0.0939	-0.0594 (0.0097)	1.04E-09	-0.3515	-0.0122 (0.0006)	7.05E-92	-0.2881	-0.0087 (0.0005)	2.54E-61
DLK1	Protein delta homolog 1	-0.0929	-0.0579 (0.0096)	1.91E-09	-0.3443	-0.0118 (0.0006)	7.20E-88	-0.2819	-0.0084 (0.0005)	1.20E-58
TREM1	Triggering receptor expressed on myeloid cells 1	-0.0869	-0.0421 (0.007)	2.09E-09	-0.4217	-0.0097 (0.0004)	4.48E-113	-0.468	-0.0098 (0.0003)	1.03E-155
DCLK1	Serine/threonine-protein kinase DCLK1	-0.0877	-0.0405 (0.0068)	2.35E-09	-0.4399	-0.0106 (0.0004)	1.78E-148	-0.4479	-0.0096 (0.0003)	6.01E-163
GAS1	Growth arrest-specific protein 1	-0.0863	-0.0317 (0.0053)	2.63E-09	-0.4827	-0.0087 (0.0003)	1.22E-160	-0.4712	-0.0074 (0.0003)	1.02E-153
JAM2	Junctional adhesion molecule B	-0.0984	-0.0236 (0.004)	2.91E-09	-0.4757	-0.0068 (0.0002)	4.07E-180	-0.455	-0.0056 (0.0002)	1.03E-160
EPHB6	Ephrin type-B receptor 6	-0.0811	-0.0272 (0.0046)	2.95E-09	-0.5061	-0.0079 (0.0003)	4.87E-185	-0.4816	-0.0065 (0.0002)	2.32E-162
MXRA7	Matrix-remodeling-associated protein 7	-0.0794	-0.0377 (0.0065)	7.36E-09	-0.5043	-0.0114 (0.0004)	1.20E-188	-0.4854	-0.0096 (0.0003)	1.52E-177
PPIC	Peptidyl-prolyl cis-trans isomerase C	-0.0794	-0.0344 (0.0059)	8.20E-09	-0.4335	-0.0093 (0.0003)	1.57E-147	-0.4598	-0.0089 (0.0003)	1.81E-182
MANSC1	MANSC domain-containing protein 1	-0.0796	-0.0312 (0.0054)	8.40E-09	-0.3686	-0.0067 (0.0003)	3.27E-89	-0.3798	-0.0062 (0.0003)	7.80E-103

CCL14	C-C motif chemokine 14	-0.0881	-0.0531 (0.0092)	8.74E-09	-0.4783	-0.0153 (0.0005)	7.84E-168	-0.555	-0.0159 (0.0004)	3.89E-251
DNAJB12	DnaJ homolog subfamily B member 12	-0.0882	-0.0389 (0.0068)	1.07E-08	-0.5364	-0.0128 (0.0004)	5.81E-224	-0.5541	-0.0117 (0.0003)	1.16E-252
EFNA2	Ephrin-A2	-0.0772	-0.0278 (0.0049)	1.69E-08	-0.5312	-0.0091 (0.0003)	1.35E-216	-0.5529	-0.0084 (0.0002)	3.15E-251
CRABP2	Cellular retinoic acid-binding protein 2	-0.081	-0.0414 (0.0074)	2.21E-08	-0.4069	-0.0112 (0.0004)	8.88E-138	-0.4583	-0.0114 (0.0004)	1.33E-193
HSPB6	Heat shock protein beta-6	-0.0692	-0.041 (0.0073)	2.35E-08	-0.5076	-0.0128 (0.0004)	6.51E-186	-0.4937	-0.0109 (0.0004)	2.83E-180
VASN	Vasorin	-0.0888	-0.0205 (0.0037)	2.45E-08	-0.3395	-0.0046 (0.0002)	2.03E-92	-0.3392	-0.0041 (0.0002)	3.50E-98
B4GALT1	Beta-1,4-galactosyltransferase 1	-0.0729	-0.0214 (0.0038)	2.73E-08	-0.4407	-0.0058 (0.0002)	8.02E-138	-0.4314	-0.0051 (0.0002)	9.76E-140
GABARAP	Gamma-aminobutyric acid receptor-associated protein	-0.0947	-0.0244 (0.0044)	2.98E-08	-0.4869	-0.0075 (0.0002)	1.44E-177	-0.6207	-0.0083 (0.0002)	4.98E-315
IL15RA	Interleukin-15 receptor subunit alpha	-0.073	-0.0383 (0.0069)	3.51E-08	-0.4396	-0.0102 (0.0004)	2.04E-129	-0.52	-0.011 (0.0003)	3.42E-207
EPHB4	Ephrin type-B receptor 4	-0.072	-0.0302 (0.0055)	4.62E-08	-0.4415	-0.0084 (0.0003)	4.51E-140	-0.4348	-0.0074 (0.0003)	1.18E-144
UNC5B	Netrin receptor UNC5B	-0.0794	-0.0301 (0.0056)	7.48E-08	-0.4751	-0.0091 (0.0003)	3.90E-159	-0.5033	-0.0085 (0.0003)	5.09E-189
AMBP	Alpha-1-microglobulin	-0.084	-0.0224 (0.0042)	7.86E-08	-0.4258	-0.0063 (0.0002)	3.53E-136	-0.4805	-0.0063 (0.0002)	1.43E-189
TAGLN	Transgelin	-0.0633	-0.0354 (0.0066)	7.98E-08	-0.5736	-0.0126 (0.0004)	2.15E-230	-0.5601	-0.0105 (0.0003)	3.27E-210
AIF1L	Allograft inflammatory factor 1-like	-0.0669	-0.0273 (0.0051)	8.28E-08	-0.4596	-0.0078 (0.0003)	1.84E-142	-0.4507	-0.0068 (0.0003)	2.82E-140
XXYLT1	Xyloside xylosyltransferase 1	-0.0748	-0.0267 (0.005)	8.78E-08	-0.341	-0.0058 (0.0003)	8.09E-78	-0.3364	-0.0051 (0.0003)	2.00E-79
SERPINF1	Pigment epithelium-derived factor	-0.0695	-0.0176 (0.0033)	1.36E-07	-0.3319	-0.0043 (0.0002)	3.14E-95	-0.3629	-0.0042 (0.0002)	2.14E-126
MAP2K2	Dual specificity mitogen-activated protein kinase kinase 2	-0.084	-0.0445 (0.0084)	1.41E-07	-0.3862	-0.0117 (0.0005)	5.85E-113	-0.3532	-0.0089 (0.0004)	9.25E-87
CAPG	Macrophage-capping protein	-0.0774	-0.0452 (0.0086)	1.46E-07	-0.4178	-0.0114 (0.0005)	2.10E-103	-0.4728	-0.0113 (0.0004)	3.13E-137

EPHA1	Ephrin type-A receptor 1	-0.0728	-0.0445 (0.0085)	1.99E-07	-0.283	-0.0088 (0.0005)	5.36E-61	-0.3036	-0.0085 (0.0004)	1.19E-75
IGFLR1	IGF-like family receptor 1	-0.0722	-0.0455 (0.0088)	2.44E-07	-0.4744	-0.0143 (0.0005)	5.14E-160	-0.5687	-0.0153 (0.0004)	9.54E-259
EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1	-0.0712	-0.0243 (0.0047)	2.53E-07	-0.4354	-0.0063 (0.0003)	2.16E-106	-0.5645	-0.0075 (0.0002)	6.01E-209
DLK2	Protein delta homolog 2	-0.0682	-0.0356 (0.0069)	2.69E-07	-0.4989	-0.012 (0.0004)	9.44E-185	-0.4489	-0.0094 (0.0003)	1.84E-148
IGFBP4	Insulin-like growth factor-binding protein 4	-0.0735	-0.0249 (0.0049)	3.33E-07	-0.3655	-0.0066 (0.0003)	3.38E-109	-0.4215	-0.0067 (0.0002)	1.18E-151
VIT	Vitrin	-0.0664	-0.0302 (0.0059)	3.91E-07	-0.3833	-0.008 (0.0004)	2.63E-105	-0.3584	-0.0067 (0.0003)	1.30E-98
PENK	Proenkephalin-A	-0.0753	-0.0432 (0.0085)	4.12E-07	-0.3396	-0.0091 (0.0005)	1.13E-66	-0.3779	-0.0089 (0.0004)	5.96E-85
LRP10	Low-density lipoprotein receptor-related protein 10	-0.0682	-0.0225 (0.0045)	4.76E-07	-0.4505	-0.0069 (0.0003)	5.60E-142	-0.429	-0.0058 (0.0002)	1.05E-133
CD93	Complement component C1q receptor	-0.0684	-0.0236 (0.0047)	5.91E-07	-0.2957	-0.0051 (0.0003)	4.17E-67	-0.2939	-0.0047 (0.0002)	8.05E-76
IL18BP	Interleukin-18-binding protein	-0.0645	-0.029 (0.0058)	6.38E-07	-0.3747	-0.0072 (0.0003)	1.79E-88	-0.4523	-0.008 (0.0003)	6.75E-150
NPDC1	Neural proliferation differentiation and control protein 1	-0.0654	-0.0295 (0.0059)	6.87E-07	-0.4681	-0.0097 (0.0003)	1.42E-162	-0.4648	-0.0083 (0.0003)	1.17E-158
WFDC2	WAP four-disulfide core domain protein 2	-0.0572	-0.0349 (0.007)	7.11E-07	-0.5846	-0.0141 (0.0004)	1.56E-260	-0.6098	-0.0131 (0.0003)	2.05E-304
VWC2	Brorin	-0.0691	-0.027 (0.0055)	8.57E-07	-0.4736	-0.0088 (0.0003)	1.70E-157	-0.4511	-0.0072 (0.0003)	3.25E-139
RARRES2	Retinoic acid receptor responder protein 2	-0.0704	-0.0284 (0.0058)	9.13E-07	-0.4542	-0.0096 (0.0003)	3.38E-167	-0.5357	-0.01 (0.0003)	4.81E-252
RNASE6	Ribonuclease K6	-0.0647	-0.0459 (0.0093)	9.40E-07	-0.4655	-0.0148 (0.0005)	4.54E-152	-0.5481	-0.0156 (0.0004)	2.94E-236
PI3	Elafin	-0.0566	-0.0496 (0.0102)	1.17E-06	-0.4154	-0.0155 (0.0006)	1.64E-138	-0.4343	-0.0148 (0.0005)	5.46E-171

CALCOCO2	Calcium-binding and coiled-coil domain-containing protein 2	-0.069	-0.0224 (0.0046)	1.28E-06	-0.3562	-0.006 (0.0003)	4.21E-98	-0.314	-0.0047 (0.0002)	2.57E-80
SPINK7	Serine protease inhibitor Kazal-type 7	-0.0628	-0.0351 (0.0072)	1.29E-06	-0.4028	-0.0105 (0.0004)	4.92E-125	-0.2811	-0.0064 (0.0004)	1.35E-60
MIA	Melanoma-derived growth regulatory protein	-0.0641	-0.0246 (0.0052)	2.28E-06	-0.2978	-0.0054 (0.0003)	1.65E-61	-0.1855	-0.0029 (0.0003)	2.70E-24
ASGR1	Asialoglycoprotein receptor 1	-0.0629	-0.0284 (0.0061)	2.79E-06	-0.4406	-0.0099 (0.0003)	1.62E-160	-0.4976	-0.01 (0.0003)	2.31E-227
CLMP	CXADR-like membrane protein	-0.0587	-0.034 (0.0073)	2.80E-06	-0.4428	-0.0116 (0.0004)	1.82E-155	-0.4516	-0.0104 (0.0004)	1.13E-165
CPLX2	Complexin-2	-0.0687	-0.0356 (0.0076)	3.05E-06	-0.511	-0.0135 (0.0004)	1.62E-192	-0.4491	-0.0096 (0.0004)	4.77E-125
ATOX1	Copper transport protein ATOX1	-0.0954	-0.0389 (0.0083)	3.15E-06	-0.3392	-0.0104 (0.0005)	2.25E-91	-0.4176	-0.0106 (0.0004)	2.52E-128
FABP3	Fatty acid-binding protein, heart	-0.0529	-0.0435 (0.0094)	3.93E-06	-0.4051	-0.0157 (0.0005)	1.78E-170	-0.4974	-0.0166 (0.0004)	2.90E-268
SMOC1	SPARC-related modular calcium-binding protein 1	-0.0671	-0.0238 (0.0052)	4.60E-06	-0.4404	-0.0074 (0.0003)	2.96E-120	-0.4244	-0.0061 (0.0003)	1.95E-109
TMPO	Lamina-associated polypeptide 2, isoforms beta/gamma	-0.0768	-0.039 (0.0085)	5.20E-06	-0.3591	-0.0108 (0.0005)	4.41E-94	-0.4851	-0.0127 (0.0004)	5.88E-179
DDOST	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	-0.0691	-0.025 (0.0055)	5.34E-06	-0.383	-0.0072 (0.0003)	5.18E-102	-0.3842	-0.0064 (0.0003)	1.92E-106
CST2	Cystatin-SA	-0.0666	-0.0537 (0.0119)	6.29E-06	-0.2963	-0.0122 (0.0007)	8.93E-61	-0.1812	-0.0062 (0.0006)	1.24E-21
NRXN3	Neurexin-3-beta	-0.0649	-0.0265 (0.0059)	6.71E-06	-0.3296	-0.0064 (0.0004)	2.00E-69	-0.277	-0.0045 (0.0003)	8.94E-46
SMOC2	SPARC-related modular calcium-binding protein 2	-0.0568	-0.0272 (0.0061)	7.97E-06	-0.4313	-0.0088 (0.0004)	9.48E-126	-0.3614	-0.0063 (0.0003)	3.47E-82
NEGR1	Neuronal growth regulator 1	-0.0533	-0.0172 (0.0039)	8.57E-06	-0.2982	-0.0038 (0.0002)	3.86E-56	-0.1808	-0.0018 (0.0002)	1.09E-17

Negative significant correlations, median[f]	-0.0855	-0.4668	-0.4697

[a] Proteins were identified through linear regression of LDPred PRS for kidney function on 4,877 proteins measured at visit 3 and visit 5 with adjusting for age at the corresponded visits, sex, center, and first 10 genetic principal components. A total of 108 proteins were significantly associated with LDPred PRS at both visit 3 and visit 5. The threshold of significance was Bonferroni corrected: $p = 1.02 \times 10^{-2}$. Visit 3 was conducted during 1993-1995 when the mean age of study population was 60.4 years and visit 5 was conducted during 2011-2013 when the mean age of study population was 75.9 years.

[b] LDPred PRS was constructed using LDPred algorithm, a Bayesian approach utilizes GWAS summary statistics to compute the posterior mean effect sizes for the genetic variants by assuming a prior of the joint effect sizes and incorporating the LD structure of the reference population.

[c] Linear regression of LDPred PRS, eGFRcr measured at visit 3, and eGFRcys measured at visit 3 on proteins measured at visit 3, with adjusting for age at visit 3, sex, center, and first 10 genetic principal components.

[d] Linear regression of LDPred PRS, eGFRcr measured at visit 5, and eGFRcys measured at visit 5 on proteins measured at visit 5, with adjusting for age at visit 5, sex, center, and first 10 genetic principal components.

[e] P values of Wilcoxon signed rank test for the comparison between correlations of proteins with LDPred PRS and with eGFR were 2.58E-19 for eGFRcr and 1.15E-18 for eGFRcys at visit 3, and were 3.90E-19 for eGFRcr and 6.99E-19 for eGFRcys at visit 5.

[f] P values of Wilcoxon signed rank test for the comparison between visit 3 and visit 5 correlations of proteins with LDPred PRS was 1.03E-13, and that of proteins with eGFR were 8.71E-19 for eGFRcr and 2.13E-18 for eGFRcys.

eGFRcr: estimated glomerular filtration rate based on creatinine; eGFRcys: estimated glomerular filtration rate based on cystatin C; PRS: polygenic risk score

Chapter 6 Conclusions

As a summary, in Chapter 2 and 3, we examined the associations between potential risk factors, hypertension and obesity, and kidney function trajectory, CKD, and ESKD using longitudinal analysis. We observed that individuals with hypertension had significantly greater decline in kidney function than those without hypertension, and the average 30-year predicted probabilities of developing CKD at all stages was higher among those with hypertension.[205] We also found that obesity status was a risk factor for future decline in kidney function and development of ESKD in black and white women with less consistent associations among men. In Chapter 4, we dig deeper into the causation between hypertension and kidney function. Not only did we demonstrate strong causal effects of lower kidney function on higher blood pressure, we also learned how Mendelian randomization (MR) can be altered when biomarkers are not perfect measurements of underlying biology and may contain genetic influences of the marker itself (e.g., creatinine) separate from the underlying physiology (e.g. reduced kidney function).[206] Interestingly, using multiple markers allowed us to triangulate on the subset of genes that are likely to reflect kidney function susceptibility. In Chapter 5, by incorporating variants across the genome, we demonstrated the link between genetic basis of kidney function measured as a PRS and a spectrum of incident kidney diseases, which is undetected by previous studies. Protein associations were stronger with eGFR than its PRS consistent with many protein elevations being secondary to the reduced kidney function. In sum, with using multiple types of data and methods, this doctoral thesis examined multiple aspects of kidney function for better understanding kidney function. It can provide scientific evidence from multiple facets -

association, to causation, to prediction - for guiding effective and efficient prevention strategies of CKD.

The implications of this doctoral thesis can also expand to public health and epidemiology in general. We build a three-step strategy for studying an outcome of interest, with each step leveraging the strengths of certain epidemiological methods. Starting with association, we benefited the solid methods and interpretations, as well as large amount of data of the classical risk factor analyses. With which, we were able to identify a target that worth further investigation. At the second step, causation, we went further and used causal inference and resources on genetics to evaluate the causal directions of risk factors and disease. At the last step, prediction, we utilized knowledge and methods from genetic and molecular epidemiology to conduct robust risk stratification and evaluate how physiological environment influence that risk stratification. Together, we were able to establish a comprehensive view of the outcome of interests, kidney function.

Future directions of this work could focus on extending the trajectory work to span the full range of ages from childhood to adolescence, young adulthood, and then midlife and older age. It is unclear how strongly abnormalities at older age such as more rapid kidney function decline and proteinuria are heralded by changes early in life. Blood pressure and body mass index increases can start at very early ages, and it may be that early prevention would result in gains through the entire lifespan. Integration of time-dependent covariates when lag times and reverse causality exists is a challenge that requires attention. The research can also be expanded to other CKD risk factors, including socioeconomic status. Likewise, the work can examine pathways by which

risk factors likely have their effect. For example, socioeconomic disparities likely contribute to hypertension, obesity, hyperglycemia, diabetes and poorer control. Understanding these pathways and proposing strategies for reducing risk and eliminating disparities is important.

With the emergence of mega cohorts, such as the UK Biobank (UKB) and the Million Veteran Program (MVP), and the rapid growth of global-wide consortiums, such as the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium and the Population Architecture Using Genomics and Epidemiology (PAGE), the power for detecting loci conferring small changes in disease risk has been greatly increased. MR and PRS, methods that aggregate genetic influences of many common genetic variants, will be more and more important in the causal inference and risk prediction of common complex diseases, including kidney diseases. We do want to point out that both our MR and PRS studies were conducted among individuals of European ancestry, the group in which most genetic studies have been undertaken to date. Because allele frequencies, linkage disequilibrium patterns, and effect sizes of common polymorphisms vary with ancestry, our findings in these two studies cannot be generalized to other ethnic groups. We would like to emphasize the need to assemble diverse cohorts for genetic discovery. And we are very glad to see that growing number of cohorts and consortiums, such as the MVP and the PAGE Consortium, are moving towards that direction. In addition to greater generalizability, studying a diverse range of populations allows for a better understanding of shared causal mechanisms. For example, linkage disequilibrium can inform identification of potentially causal variants to study in vitro and in animal systems with the goal of improved design for therapeutic agents. Addressing causation will also benefit from a synthesis of data across different study designs. Epidemiology focuses on human studies and its observations can

be tested and illuminated by animal and in vitro experimental studies. Once an firm understanding of a potential treatment for reducing risk is in place and adequate safety data is collected, human experimental studies are a useful final step for testing strategies to reduce the risk of CKD and its progression.

Driven by technological advances, larger amount of omics data on more layers at more time points has been generated, which poses opportunities for better understanding the biology of kidney function and challenges for their integration. Machine learning approaches to mining multi-omics data hold great promises in overcoming the challenges in integrating heterogenous and temporal data. Complex models which integrate multi-omics data from genetics to epigenetics, expression, proteomics, physiologic and eventually disease risk can be assembled from datasets such as the ones we have studied. Such models will yield insights into biology and the design of therapeutics. However, it is still uncertain how much risk is predictable vs. unpredictable due to chaos theory's implications for large biological systems; there are limits to prediction, even if we know all components of a system and its operating rules. At present, I am working on applying a novel continuous-time, random forest method for survival analysis, RF-SLAM, to predict sudden cardiac death using the longitudinal data in ARIC. Time dependent modeling holds promise for improving our understanding of the dynamics of disease. Genetics are powerful time invariant elements of biology but their expression over time is both fascinating and complex. I plan to extend the application of this method and other machine learning approaches to integrate multiple omics as well as environmental factors such as lifestyle and clinical covariates for better studying kidney function and diseases. Overall, we must combine improved understanding of biology to advance therapeutics, treatment and prevention,

combining new tools with proven strategies for vascular disease prevention such as eating a healthy diet, exercising, avoiding smoking and excess weight which are powerful drivers of our complex biology and vascular disease risk.

References

1. Eckardt KU, Coresh J, Devuyst O, et al. Evolving importance of kidney disease: from subspecialty to global health burden. Lancet 2013;382:158-69.

2. Baumeister SE, Boger CA, Kramer BK, et al. Effect of chronic kidney disease and comorbid conditions on health care costs: A 10-year observational study in a general population. American journal of nephrology 2010;31:222-9.

3. Mills KT, Bundy JD, Kelly TN, et al. Global Disparities of Hypertension Prevalence and Control: A Systematic Analysis of Population-Based Studies From 90 Countries. Circulation 2016;134:441-50.

4. Astor BC, Matsushita K, Gansevoort RT, et al. Lower estimated glomerular filtration rate and higher albuminuria are associated with mortality and end-stage renal disease. A collaborative meta-analysis of kidney disease population cohorts. Kidney international 2011;79:1331-40.

5. Gansevoort RT, Matsushita K, van der Velde M, et al. Lower estimated GFR and higher albuminuria are associated with adverse kidney outcomes. A collaborative meta-analysis of general and high-risk population cohorts. Kidney international 2011;80:93-104.

6. Matsushita K, Coresh J, Sang Y, et al. Estimated glomerular filtration rate and albuminuria for prediction of cardiovascular outcomes: a collaborative meta-analysis of individual participant data. The lancet Diabetes & endocrinology 2015;3:514-25.

7. Whelton PK, Klag MJ. Hypertension as a risk factor for renal disease. Review of clinical and epidemiological evidence. Hypertension (Dallas, Tex : 1979) 1989;13:I19-27.

8. Anderson AH, Yang W, Townsend RR, et al. Time-updated systolic blood pressure and the progression of chronic kidney disease: a cohort study. Annals of internal medicine 2015;162:258-65.

9. Yu Z, Rebholz CM, Wong E, et al. Association Between Hypertension and Kidney Function Decline: The Atherosclerosis Risk in Communities (ARIC) Study. American journal of kidney diseases : the official journal of the National Kidney Foundation 2019.

10. Judson GL, Rubinsky AD, Shlipak MG, et al. Longitudinal Blood Pressure Changes and Kidney Function Decline in Persons Without Chronic Kidney Disease: Findings From the MESA Study. American journal of hypertension 2018;31:600-8.

11. Beddhu S, Rocco MV, Toto R, et al. Effects of Intensive Systolic Blood Pressure Control on Kidney and Cardiovascular Outcomes in Persons Without Kidney Disease: A Secondary Analysis of a Randomized Trial. Ann Intern Med 2017;167:375-83.

12. Xie X, Atkins E, Lv J, et al. Effects of intensive blood pressure lowering on cardiovascular and renal outcomes: updated systematic review and meta-analysis. Lancet 2016;387:435-43.

13. Wuttke M, Li Y, Li M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. Nature genetics 2019;51:957-72.

14. Pani A, Bragg-Gresham J, Masala M, et al. Prevalence of CKD and its relationship to eGFR-related genetic loci and clinical risk factors in the SardiNIA study cohort. Journal of the American Society of Nephrology : JASN 2014;25:1533-44.

15. Ma J, Yang Q, Hwang SJ, Fox CS, Chu AY. Genetic risk score and risk of stage 3 chronic kidney disease. BMC nephrology 2017;18:32.

16. Thio CHL, van der Most PJ, Nolte IM, et al. Evaluation of a genetic risk score based on creatinine-estimated glomerular filtration rate and its association with kidney outcomes. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association 2018;33:1757-64.

17. Yun S, Han M, Kim HJ, et al. Genetic risk score raises the risk of incidence of chronic kidney disease in Korean general population-based cohort. Clinical and experimental nephrology 2019;23:995-1003.

18. Hellwege JN, Velez Edwards DR, Giri A, et al. Mapping eGFR loci to the renal transcriptome and phenome in the VA Million Veteran Program. Nature communications 2019;10:3842.

19. Fujii R, Hishida A, Nakatochi M, et al. Association of genetic risk score and chronic kidney disease in a Japanese population. Nephrology (Carlton, Vic) 2019;24:670-3.

20. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 2014;23:R89-98.

21. Rohloff JC, Gelinas AD, Jarvis TC, et al. Nucleic Acid Ligands With Protein-like Side Chains: Modified Aptamers and Their Use as Diagnostic and Therapeutic Agents. Molecular therapy Nucleic acids 2014;3:e201.

22. Tin A, Yu B, Ma J, et al. Reproducibility and Variability of Protein Analytes Measured Using a Multiplexed Modified Aptamer Assay. The journal of applied laboratory medicine 2019;4:30-9.

23. Marshall J, Bowden P, Schmit JC, Betsou F. Creation of a federated database of blood proteins: a powerful new tool for finding and characterizing biomarkers in serum. Clin Proteomics 2014;11:3.

24. Nanjappa V, Thomas JK, Marimuthu A, et al. Plasma Proteome Database as a resource for proteomics research: 2014 update. Nucleic Acids Res 2014;42:D959-65.

25. Stastna M, Van Eyk JE. Secreted proteins as a fundamental source for biomarker discovery. Proteomics 2012;12:722-35.

26. Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. Science (New York, NY) 2015;347:1260419.

27. Schwenk JM, Omenn GS, Sun Z, et al. The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. Journal of proteome research 2017;16:4299-310.

28. Liu Y, Buil A, Collins BC, et al. Quantitative variability of 342 plasma proteins in a human twin population. Molecular systems biology 2015;11:786.

29. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. American journal of epidemiology 1989;129:687-702.

30. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12:e1001779.

31. Evangelou E, Warren HR, Mosen-Ansorena D, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet 2018;50:1412-25.

32. Yu Z, Rebholz CM, Wong E, et al. Association Between Hypertension and Kidney Function Decline: The Atherosclerosis Risk in Communities (ARIC) Study. American journal of kidney diseases : the official journal of the National Kidney Foundation 2019;74:310-9.

33. McCallum W, Ku E, Sarnak MJ. More Evidence to Suggest a Relation of Blood Pressure to Long-term Progression of Kidney Disease: Is It Causal? American journal of kidney diseases : the official journal of the National Kidney Foundation 2019.

34. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. Hum Mol Genet 2018;27:R195-R208.

35. Qi G, Chatterjee N. A Comprehensive Evaluation of Methods for Mendelian Randomization Using Realistic Simulations of Genome-wide Association Studies. bioRxiv 2019:702787.

36. Vilhjálmsson BJ, Yang J, Finucane HK, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. American journal of human genetics 2015;97:576-92.

37. Lim SS, Vos T, Flaxman AD, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet (London, England) 2012;380:2224-60.

38. Black HR. The burden of cardiovascular disease: following the link from hypertension to myocardial infarction and heart failure. Am J Hypertens 2003;16:4s-6s.

39. Muntner P, Carey RM, Gidding S, et al. Potential U.S. Population Impact of the 2017 ACC/AHA High Blood Pressure Guideline. Journal of the American College of Cardiology 2018;71:109-18.

40. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults:

Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2018;71:1269-324.

41. Stephens-Shields AJ, Spieker AJ, Anderson A, et al. Blood pressure and the risk of chronic kidney disease progression using multistate marginal structural models in the CRIC Study. Statistics in medicine 2017;36:4167-81.

42. Inker LA, Shafi T, Okparavero A, et al. Effects of Race and Sex on Measured GFR: The Multi-Ethnic Study of Atherosclerosis. American journal of kidney diseases : the official journal of the National Kidney Foundation 2016;68:743-51.

43. Tedla FM, Brar A, Browne R, Brown C. Hypertension in chronic kidney disease: navigating the evidence. International journal of hypertension 2011;2011:132405.

44. Cole MB, Wright B, Wilson IB, Galarraga O, Trivedi AN. Longitudinal Analysis of Racial/Ethnic Trends in Quality Outcomes in Community Health Centers, 2009-2014. Journal of general internal medicine 2018;33:906-13.

45. Howard G, Prineas R, Moy C, et al. Racial and geographic differences in awareness, treatment, and control of hypertension: the REasons for Geographic And Racial Differences in Stroke study. Stroke 2006;37:1171-8.

46. Lackland DT. Racial differences in hypertension: implications for high blood pressure management. The American journal of the medical sciences 2014;348:135-8.

47. Hypertension among adults in the United States, 2009–2010. NCHS data brief no. 107. Hyattsville, MD: US Department of Health and Human Services, CDC, National Center for Health Statistics, 2012. at http://www.cdc.gov/nchs/data/databriefs/db107.htm.)

48. Saran R, Li Y, Robinson B, et al. US Renal Data System 2015 Annual Data Report: Epidemiology of Kidney Disease in the United States. Am J Kidney Dis 2016;67:A7-8.

49. Muntner P, Newsome B, Kramer H, et al. Racial differences in the incidence of chronic kidney disease. Clin J Am Soc Nephrol 2012;7:101-7.

50. Peralta CA, Katz R, DeBoer I, et al. Racial and ethnic differences in kidney function decline among persons without chronic kidney disease. J Am Soc Nephrol 2011;22:1327-34.

51. Grams ME, Chow EK, Segev DL, Coresh J. Lifetime incidence of CKD stages 3-5 in the United States. Am J Kidney Dis 2013;62:245-52.

52. Klag MJ, Whelton PK, Randall BL, Neaton JD, Brancati FL, Stamler J. End-stage renal disease in African-American and white men. 16-year MRFIT findings. JAMA 1997;277:1293-8.

53. Young BA, Maynard C, Boyko EJ. Racial differences in diabetic nephropathy, cardiovascular disease, and mortality in a national population of veterans. Diabetes Care 2003;26:2392-9.

54. Grams ME, Rebholz CM, Chen Y, et al. Race, APOL1 Risk, and eGFR Decline in the General Population. J Am Soc Nephrol 2016.

55. Parsa A, Kao WH, Xie D, et al. APOL1 risk variants, race, and progression of chronic kidney disease. N Engl J Med 2013;369:2183-96.

56. Eckfeldt JH, Chambless LE, Shen YL. Short-term, within-person variability in clinical chemistry test results. Experience from the Atherosclerosis Risk in Communities Study. Archives of pathology & laboratory medicine 1994;118:496-500.

57. Coresh J, Astor BC, McQuillan G, et al. Calibration and random variation of the serum creatinine assay as critical elements of using equations to estimate glomerular filtration rate. Am J Kidney Dis 2002;39:920-9.

58. Parrinello CM, Grams ME, Couper D, et al. Recalibration of blood analytes over 25 years in the atherosclerosis risk in communities study: impact of recalibration on chronic kidney disease prevalence and incidence. Clinical chemistry 2015;61:938-47.

59. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. Annals of internal medicine 2009;150:604-12.

60. Warnick GR, Benderson J, Albers JJ. Dextran sulfate-Mg2+ precipitation procedure for quantitation of high-density-lipoprotein cholesterol. Clinical chemistry 1982;28:1379-88.

61. Robinson GK. That BLUP is a Good Thing: The Estimation of Random Effects. Statistical Science 1991;6:15-32.

62. Kidney Disease: Improving Global Outcomes (KDIGO). KDIGO clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney International Supplements 2013;3:1-150.

63. Lash JP, Go AS, Appel LJ, et al. Chronic Renal Insufficiency Cohort (CRIC) Study: baseline characteristics and associations with kidney function. Clin J Am Soc Nephrol 2009;4:1302-11.

64. Anderson AH, Yang W, Townsend RR, et al. Time-updated systolic blood pressure and the progression of chronic kidney disease: a cohort study. Annals of internal medicine 2015;162:258-65.

65. Zoppini G, Targher G, Chonchol M, et al. Predictors of estimated GFR decline in patients with type 2 diabetes and preserved kidney function. Clin J Am Soc Nephrol 2012;7:401-8.

66. Gulley SP, Rasch EK, Chan L. Difference, disparity, and disability: a comparison of health, insurance coverage, and health service use on the basis of race/ethnicity among US adults with disabilities, 2006-2008. Medical care 2014;52:S9-16.

67. Redmond N, Baer HJ, Hicks LS. Health behaviors and racial disparity in blood pressure control in the national health and nutrition examination survey. Hypertension 2011;57:383-9.

68. Hsu CY, Lin F, Vittinghoff E, Shlipak MG. Racial differences in the progression from chronic renal insufficiency to end-stage renal disease in the United States. Journal of the American Society of Nephrology : JASN 2003;14:2902-7.

69. Grams ME, Sang Y, Levey AS, et al. Kidney-Failure Risk Projection for the Living Kidney-Donor Candidate. The New England journal of medicine 2016;374:411-21.

70. Crews DC, Liu Y, Boulware LE. Disparities in the burden, outcomes, and care of chronic kidney disease. Curr Opin Nephrol Hypertens 2014;23:298-305.

71. Hall YN. Racial and ethnic disparities in end stage renal disease: access failure. Clin J Am Soc Nephrol 2012;7:196-8.

72. Kao WH, Klag MJ, Meoni LA, et al. MYH9 is associated with nondiabetic end-stage renal disease in African Americans. Nat Genet 2008;40:1185-92.

73. Keane WF, Eknoyan G. Proteinuria, albuminuria, risk, assessment, detection, elimination (PARADE): a position paper of the National Kidney Foundation. American journal of kidney diseases : the official journal of the National Kidney Foundation 1999;33:1004-10.

74. Yoshioka T, Rennke HG, Salant DJ, Deen WM, Ichikawa I. Role of abnormally high transmural pressure in the permselectivity defect of glomerular capillary wall: a study in early passive Heymann nephritis. Circulation research 1987;61:531-8.

75. Walser M, Drew HH, LaFrance ND. Creatinine measurements often yielded false estimates of progression in chronic renal failure. Kidney international 1988;34:412-8.

76. Turin TC, Coresh J, Tonelli M, et al. Short-term change in kidney function and risk of end-stage renal disease. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association 2012;27:3835-43.

77. Turin TC, Coresh J, Tonelli M, et al. Change in the estimated glomerular filtration rate over time and risk of all-cause mortality. Kidney international 2013;83:684-91.

78. Matsushita K, Selvin E, Bash LD, Franceschini N, Astor BC, Coresh J. Change in estimated GFR associates with coronary heart disease and mortality. Journal of the American Society of Nephrology : JASN 2009;20:2617-24.

79. Herrington WG, Smith M, Bankhead C, et al. Body-mass index and risk of advanced chronic kidney disease: Prospective analyses from a primary care cohort of 1.4 million adults in England. PloS one 2017;12:e0173515.

80. Kopple JD. Obesity and chronic kidney disease. Journal of renal nutrition : the official journal of the Council on Renal Nutrition of the National Kidney Foundation 2010;20:S29-30.

81. Grubbs V, Lin F, Vittinghoff E, et al. Body mass index and early kidney function decline in young adults: a longitudinal analysis of the CARDIA (Coronary Artery Risk Development in Young Adults) study. American journal of kidney diseases : the official journal of the National Kidney Foundation 2014;63:590-7.

82. Ahima RS, Lazar MA. Physiology. The health risk of obesity--better metrics imperative. Science (New York, NY) 2013;341:856-8.

83. Mahadevan S, Ali I. Is body mass index a good indicator of obesity? International Journal of Diabetes in Developing Countries 2016;36:140-2.

84. Ortega FB, Sui X, Lavie CJ, Blair SN. Body Mass Index, the Most Widely Used But Also Widely Criticized Index: Would a Criterion Standard Measure of Total Body Fat Be a Better Predictor of Cardiovascular Disease Mortality? Mayo Clin Proc 2016;91:443-55.

85. Komura H, Nomura I, Kitamura K, Kuwasako K, Kato J. Gender difference in relationship between body mass index and development of chronic kidney disease. BMC research notes 2013;6:463.

86. Cobo G, Hecking M, Port FK, et al. Sex and gender differences in chronic kidney disease: progression to end-stage renal disease and haemodialysis. Clinical science (London, England : 1979) 2016;130:1147-63.

87. Lee DH, Keum N, Hu FB, et al. Development and validation of anthropometric prediction equations for lean body mass, fat mass and percent fat in adults using the National Health and Nutrition Examination Survey (NHANES) 1999-2006. The British journal of nutrition 2017;118:858-66.

88. Srikanthan P, Seeman TE, Karlamangla AS. Waist-hip-ratio as a predictor of all-cause mortality in high-functioning older adults. Ann Epidemiol 2009;19:724-31.

89. Song M, Giovannucci E. Estimating the Influence of Obesity on Cancer Risk: Stratification by Smoking Is Critical. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2016;34:3237-9.

90. Flegal KM, Kit BK, Orpana H, Graubard BI. Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. Jama 2013;309:71-82.

91. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association 1999;94:496-509.

92. Huang WH, Chen CY, Lin JL, Lin-Tan DT, Hsu CW, Yen TH. High body mass index reduces glomerular filtration rate decline in type II diabetes mellitus patients with stage 3 or 4 chronic kidney disease. Medicine 2014;93:e41.

93. Burckhardt P, Nagin DS, Padman R. Multi-Trajectory Models of Chronic Kidney Disease Progression. AMIA Annual Symposium proceedings AMIA Symposium 2016;2016:1737-46.

94. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. Jama 2006;295:1549-55.

95. Parsa A, Kao WH, Xie D, et al. APOL1 risk variants, race, and progression of chronic kidney disease. The New England journal of medicine 2013;369:2183-96.

96. 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. Annals of internal medicine 1999;130:461-70.

97. Liu X, Foster MC, Tighiouart H, et al. Non-GFR Determinants of Low-Molecular-Weight Serum Protein Filtration Markers in CKD. American journal of kidney diseases : the official journal of the National Kidney Foundation 2016;68:892-900.

98. Rule AD, Bailey KR, Schwartz GL, Khosla S, Lieske JC, Melton LJ, 3rd. For estimating creatinine clearance measuring muscle mass gives better results than those based on demographics. Kidney international 2009;75:1071-8.

99. LeBlanc ES, Rizzo JH, Pedula KL, et al. Long-Term Weight Trajectory and Risk of Hip Fracture, Falls, Impaired Physical Function, and Death. Journal of the American Geriatrics Society 2018;66:1972-9.

100. Cobb LK, McAdams-DeMarco MA, Gudzune KA, et al. Changes in Body Mass Index and Obesity Risk in Married Couples Over 25 Years: The ARIC Cohort Study. American journal of epidemiology 2016;183:435-43.

101. Wei GS, Coady SA, Reis JP, et al. Duration and Degree of Weight Gain and Incident Diabetes in Younger Versus Middle-Aged Black and White Adults: ARIC, CARDIA, and the Framingham Heart Study. Diabetes care 2015;38:2042-9.

102. Stevens J, Tyroler HA, Cai J, et al. Body weight change and carotid artery wall thickness. The Atherosclerosis Risk in Communities (ARIC) Study. American journal of epidemiology 1998;147:563-73.

103. Ndumele CE, Cobb L, Lazo M, et al. Weight History and Subclinical Myocardial Damage. Clinical chemistry 2018;64:201-9.

104. Rissanen PM, Laakkonen EI, Suntioinen S, Penttila IM, Uusitupa MI. The nutritional status of Finnish home-living elderly people and the relationship between energy intake and chronic diseases. Age and ageing 1996;25:133-8.

105. Hedlund J. Community-acquired pneumonia requiring hospitalisation. Factors of importance for the short-and long term prognosis. Scandinavian journal of infectious diseases Supplementum 1995;97:1-60.

106. Tosteson AN, Gottlieb DJ, Radley DC, Fisher ES, Melton LJ, 3rd. Excess mortality following hip fracture: the role of underlying health status. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 2007;18:1463-72.

107. Baumgartner RN, Stauber PM, Koehler KM, Romero L, Garry PJ. Associations of fat and muscle masses with bone mineral in elderly men and women. The American journal of clinical nutrition 1996;63:365-72.

108. Lee DH, Keum N, Hu FB, et al. Comparison of the association of predicted fat mass, body mass index, and other obesity indicators with type 2 diabetes risk: two large prospective studies in US men and women. European journal of epidemiology 2018;33:1113-23.

109. Lee DH, Keum N, Hu FB, et al. Predicted lean body mass, fat mass, and all cause and cause specific mortality in men: prospective US cohort study. BMJ (Clinical research ed) 2018;362:k2575.

110. Curtis JP, Selter JG, Wang Y, et al. The obesity paradox: body mass index and outcomes in patients with heart failure. Archives of internal medicine 2005;165:55-61.

111. Tobias DK, Pan A, Jackson CL, et al. Body-mass index and mortality among adults with incident type 2 diabetes. The New England journal of medicine 2014;370:233-44.

112. Tonelli M, Wiebe N, Culleton B, et al. Chronic kidney disease and mortality risk: a systematic review. J Am Soc Nephrol 2006;17:2034-47.

113. Matsushita K, van der Velde M, Astor BC, et al. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. Lancet (London, England) 2010;375:2073-81.

114. Forouzanfar MH, Liu P, Roth GA, et al. Global Burden of Hypertension and Systolic Blood Pressure of at Least 110 to 115 mm Hg, 1990-2015. JAMA 2017;317:165-82.

115. Huang M, Matsushita K, Sang Y, Ballew SH, Astor BC, Coresh J. Association of kidney function and albuminuria with prevalent and incident hypertension: the Atherosclerosis Risk in Communities (ARIC) study. Am J Kidney Dis 2015;65:58-66.

116. Kestenbaum B, Rudser KD, de Boer IH, et al. Differences in kidney function and incident hypertension: the multi-ethnic study of atherosclerosis. Ann Intern Med 2008;148:501-8.

117. Liu HM, Hu Q, Zhang Q, et al. Causal Effects of Genetically Predicted Cardiovascular Risk Factors on Chronic Kidney Disease: A Two-Sample Mendelian Randomization Study. Frontiers in genetics 2019;10:415.

118. Morris AP, Le TH, Wu H, et al. Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. Nat Commun 2019;10:29.

119. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. Annals of internal medicine 2009;150:604-12.

120. Item CB, Stockler-Ipsiroglu S, Stromberger C, et al. Arginine:glycine amidinotransferase deficiency: the third inborn error of creatine metabolism in humans. Am J Hum Genet 2001;69:1127-33.

121. Zanetti D, Rao A, Gustafsson S, Assimes TL, Montgomery SB, Ingelsson E. Identification of 22 novel loci associated with urinary biomarkers of albumin, sodium, and potassium excretion. Kidney Int 2019;95:1197-208.

122. Olives B, Martial S, Mattei MG, et al. Molecular characterization of a new urea transporter in the human kidney. FEBS Lett 1996;386:156-60.

123. Smith CP, Fenton RA. Genomic organization of the mammalian SLC14a2 urea transporter genes. J Membr Biol 2006;212:109-17.

124. Vaes B, Beke E, Truyers C, et al. The correlation between blood pressure and kidney function decline in older people: a registry-based cohort study. BMJ Open 2015;5:e007571.

125. Lindeman RD, Tobin JD, Shock NW. Association between blood pressure and the rate of decline in renal function with age. Kidney Int 1984;26:861-8.

126. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013;37:658-65.

127. Schmidt AF, Dudbridge F. Mendelian randomization with Egger pleiotropy correction and weakly informative Bayesian priors. Int J Epidemiol 2018;47:1217-28.

128. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol 2016;40:304-14.

129. Qi G, Chatterjee N. Mendelian Randomization Analysis Using Mixture Models (MRMix) for Genetic Effect-Size-Distribution Leads to Robust Estimation of Causal Effects. bioRxiv 2018:367821.

130. Qi G, Chatterjee N. A Comprehensive Evaluation of Methods for Mendelian Randomization Using Realistic Simulations and an Analysis of 38 Biomarkers for Risk of Type-2 Diabetes. bioRxiv 2020:702787.

131. Levey AS, Inker LA. GFR as the "Gold Standard": Estimated, Measured, and True. American Journal of Kidney Diseases 2016;67:9-12.

132. Wadei HM, Textor SC. The role of the kidney in regulating arterial blood pressure. Nat Rev Nephrol 2012;8:602-9.

133. Heymann J, Winkler CA, Hoek M, Susztak K, Kopp JB. Therapeutics for APOL1 nephropathies: putting out the fire in the podocyte. Nephrol Dial Transplant 2017;32:i65-i70.

134. Kopp JB. Rethinking hypertensive kidney disease: arterionephrosclerosis as a genetic, metabolic, and inflammatory disorder. Curr Opin Nephrol Hypertens 2013;22:266-72.

135. Haas ME, Aragam KG, Emdin CA, et al. Genetic Association of Albuminuria with Cardiometabolic Disease and Blood Pressure. Am J Hum Genet 2018;103:461-73.

136. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. International journal of epidemiology 2017;46:1985-98.

137. Carlstrom M, Wilcox CS, Arendshorst WJ. Renal autoregulation in health and disease. Physiol Rev 2015;95:405-511.

138. Levey AS, Gansevoort RT, Coresh J, et al. Change in Albuminuria and GFR as End Points for Clinical Trials in Early Stages of CKD: A Scientific Workshop Sponsored by the National Kidney Foundation in Collaboration With the US Food and Drug Administration and European Medicines Agency. Am J Kidney Dis 2020;75:84-104.

139. Stengel B. Hypertension and glomerular function decline: the chicken or the egg? Kidney Int 2016;90:254-6.

140. Chang TI, Sarnak MJ. Intensive Blood Pressure Targets and Kidney Disease. Clinical journal of the American Society of Nephrology : CJASN 2018;13:1575-7.

141. Tsai WC, Wu HY, Peng YS, et al. Association of Intensive Blood Pressure Control and Kidney Disease Progression in Nondiabetic Patients With Chronic Kidney Disease: A Systematic Review and Meta-analysis. JAMA internal medicine 2017;177:792-9.

142. Grassi G, Mancia G, Nilsson PM. Specific Blood Pressure Targets for Patients With Diabetic Nephropathy? Diabetes care 2016;39 Suppl 2:S228-33.

143. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. Science 2015;347:1260419.

144. Zheng J, Baird D, Borges MC, et al. Recent Developments in Mendelian Randomization Studies. Curr Epidemiol Rep 2017;4:330-45.

145. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genetic epidemiology 2016;40:597-608.

146. Hartwig FP, Davies NM, Hemani G, Davey Smith G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. International Journal of Epidemiology 2017;45:1717-26.

147. Ehret GB, Ferreira T, Chasman DI, et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. Nature genetics 2016;48:1171-84.

148. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 2016;48:1279-83.

149. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56-65.

150. Schwartz GJ, Schneider MF, Maier PS, et al. Improved equations estimating GFR in children with chronic kidney disease using an immunonephelometric determination of cystatin C. Kidney international 2012;82:445-53.

151. Tobin MD, Sheehan NA, Scurrah KJ, Burton PR. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. Statistics in medicine 2005;24:2911-35.

152. Hallan SI, Orth SR. Smoking is a risk factor in the progression to kidney failure. Kidney Int 2011;80:516-23.

153. GWAS Round 2 Results. 2018. (Accessed July, 2019, at http://www.nealelab.is/uk-biobank/.)

154. Orth SR, Hallan SI. Smoking: a risk factor for progression of chronic kidney disease and for cardiovascular morbidity and mortality in renal patients--absence of evidence or evidence of absence? Clin J Am Soc Nephrol 2008;3:226-36.

155. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics 2007;81:559-75.

156. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLoS Genet 2017;13:e1007081.

157. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiology (Cambridge, Mass) 2017;28:30-42.

158. Satko SG, Freedman BI. The familial clustering of renal disease and related phenotypes. The Medical clinics of North America 2005;89:447-56.

159. O'Seaghdha CM, Fox CS. Genome-wide association studies of chronic kidney disease: what have we learned? Nature reviews Nephrology 2011;8:89-99.

160. Wu HH, Kuo CF, Li IJ, et al. Family Aggregation and Heritability of ESRD in Taiwan: A Population-Based Study. American journal of kidney diseases : the official journal of the National Kidney Foundation 2017;70:619-26.

161. Akrawi DS, PirouziFard M, Fjellstedt E, Sundquist J, Sundquist K, Zöller B. Heritability of End-Stage Renal Disease: A Swedish Adoption Study. Nephron 2018;138:157-65.

162. Suhre K, Arnold M, Bhagwat AM, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. Nature communications 2017;8:14357.

163. de Vries PS, Yu B, Feofanova EV, et al. Whole-genome sequencing study of serum peptide levels: the Atherosclerosis Risk in Communities study. Human molecular genetics 2017;26:3442-50.

164. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. Nature 2018;558:73-9.

165. Emilsson V, Ilkov M, Lamb JR, et al. Co-regulatory networks of human serum proteins link genetics to disease. Science (New York, NY) 2018;361:769-73.

166. Stafford-Smith M, Li YJ, Mathew JP, et al. Genome-wide association study of acute kidney injury after coronary bypass graft surgery identifies susceptibility loci. Kidney international 2015;88:823-32.

167. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. American journal of epidemiology 1989;129:687-702.

168. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nature methods 2011;9:179-81.

169. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics 2009;5:e1000529.

170. Kowalski MH, Qian H, Hou Z, et al. Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. PLoS genetics 2019;15:e1008500.

171. Ellinor PT, Lunetta KL, Albert CM, et al. Meta-analysis identifies six new susceptibility loci for atrial fibrillation. Nature genetics 2012;44:670-5.

172. Inker LA, Eckfeldt J, Levey AS, et al. Expressing the CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration) cystatin C equations for estimating GFR with standardized serum cystatin C values. American journal of kidney diseases : the official journal of the National Kidney Foundation 2011;58:682-4.

173. Grubb A, Blirup-Jensen S, Lindström V, Schmidt C, Althaus H, Zegers I. First certified reference material for cystatin C in human serum ERM-DA471/IFCC. Clinical chemistry and laboratory medicine 2010;48:1619-21.

174. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562:203-9.

175. Köttgen A, Pattaro C. The CKDGen Consortium: ten years of insights into the genetic basis of kidney function. Kidney international 2020;97:236-42.

176. Nolte IM, van der Most PJ, Alizadeh BZ, et al. Missing heritability: is the gap closing? An analysis of 32 complex traits in the Lifelines Cohort Study. European journal of human genetics : EJHG 2017;25:877-85.

177. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. Nature 2015;526:68-74.

178. Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. Nature 2010;467:52-8.

179. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;4:7.

180. Grams ME, Rebholz CM, McMahon B, et al. Identification of incident CKD stage 3 in research studies. American journal of kidney diseases : the official journal of the National Kidney Foundation 2014;64:214-21.

181. Grams ME, Waikar SS, MacMahon B, Whelton S, Ballew SH, Coresh J. Performance and limitations of administrative data in the identification of AKI. Clinical journal of the American Society of Nephrology : CJASN 2014;9:682-9.

182. Ganz P, Heidecker B, Hveem K, et al. Development and Validation of a Protein-Based Risk Score for Cardiovascular Outcomes Among Patients With Stable Coronary Heart Disease. JAMA 2016;315:2532-41.

183. Ngo D, Sinha S, Shen D, et al. Aptamer-Based Proteomic Profiling Reveals Novel Candidate Biomarkers and Pathways in Cardiovascular Disease. Circulation 2016;134:270-85.

184. Candia J, Cheung F, Kotliarov Y, et al. Assessment of Variability in the SOMAscan Assay. Sci Rep 2017;7:14248.

185. Aragam KG, Natarajan P. Polygenic Scores to Assess Atherosclerotic Cardiovascular Disease Risk: Clinical Perspectives and Basic Implications. Circulation research 2020;126:1159-77.

186. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences of the United States of America 2009;106:9362-7.

187. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. American journal of human genetics 2012;90:7-24.

188. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. Nature reviews Genetics 2008;9:255-66.

189. Loh PR, Bhatia G, Gusev A, et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. Nature genetics 2015;47:1385-92.

190. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 2017;169:1177-86.

191. Zeng J, de Vlaming R, Wu Y, et al. Signatures of negative selection in the genetic architecture of human complex traits. Nature genetics 2018;50:746-53.

192. Zhang W, Bojorquez-Gomez A, Velez DO, et al. A global transcriptional network connecting noncoding mutations to changes in tumor gene expression. Nature genetics 2018;50:613-20.

193. Nakada M, Miyamori H, Yamashita J, Sato H. Testican 2 abrogates inhibition of membrane-type matrix metalloproteinases by other testican family proteins. Cancer research 2003;63:3364-9.

194. Chavakis T, Athanasopoulos A, Rhee JS, et al. Angiostatin is a novel anti-inflammatory factor by inhibiting leukocyte recruitment. Blood 2005;105:1036-43.

195. Benelli R, Morini M, Carrozzino F, et al. Neutrophils as a key cellular target for angiostatin: implications for regulation of angiogenesis and inflammation. FASEB journal : official publication of the Federation of American Societies for Experimental Biology 2002;16:267-9.

196. Perri SR, Annabi B, Galipeau J. Angiostatin inhibits monocyte/macrophage migration via disruption of actin cytoskeleton. FASEB journal : official publication of the Federation of American Societies for Experimental Biology 2007;21:3928-36.

197. Kang DH, Kanellis J, Hugo C, et al. Role of the microvascular endothelium in progressive renal disease. Journal of the American Society of Nephrology : JASN 2002;13:806-16.

198. Stenvinkel P, Ketteler M, Johnson RJ, et al. IL-10, IL-6, and TNF-alpha: central factors in the altered cytokine network of uremia--the good, the bad, and the ugly. Kidney international 2005;67:1216-33.

199. Mu W, Long DA, Ouyang X, et al. Angiostatin overexpression is associated with an improvement in chronic kidney injury by an anti-inflammatory mechanism. American journal of physiology Renal physiology 2009;296:F145-52.

200. Zhang SX, Wang JJ, Lu K, Mott R, Longeras R, Ma JX. Therapeutic potential of angiostatin in diabetic nephropathy. Journal of the American Society of Nephrology : JASN 2006;17:475-86.

201. Xia YY, Bu R, Cai GY, et al. Urinary angiostatin: a novel biomarker of kidney disease associated with disease severity and progression. BMC nephrology 2019;20:118.

202. Wu T, Du Y, Han J, et al. Urinary angiostatin--a novel putative marker of renal pathology chronicity in lupus nephritis. Molecular & cellular proteomics : MCP 2013;12:1170-9.

203. Wojcik GL, Graff M, Nishimura KK, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature 2019;570:514-8.

204. Martin AR, Gignoux CR, Walters RK, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. American journal of human genetics 2017;100:635-49.

205. Yu Z, Rebholz CM, Wong E, et al. Association Between Hypertension and Kidney Function Decline: The Atherosclerosis Risk in Communities (ARIC) Study. American journal of kidney diseases : the official journal of the National Kidney Foundation 2019;74:310-9.

206. Yu Z, Coresh J, Qi G, et al. A bidirectional Mendelian randomization study supports causal effects of kidney function on blood pressure. Kidney international 2020.

Curriculum Vitae

ZHI YU

615 N. Wolfe Street, Baltimore, MD 21205 zyu33@jhmi.edu

EDUCATION AND TRAINING

Johns Hopkins Bloomberg School of Public Health, Baltimore, US **Expected Aug 2020**

Doctor of Philosophy (PhD) in Epidemiology

Advisors: Josef Coresh, MD, PhD, Nilanjan Chatterjee, PhD, Adrienne Tin, PhD

Thesis: Integrating multi-omics approaches and human biology for understanding kidney function

Master of Health Science (MHS) in Biostatistics

Harvard T.H. Chan School of Public Health, Boston, US **May 2015**

Master of Science (MS) in Epidemiology

Guangdong Hospital of Chinese Medicine, Guangzhou, China **May 2012 - Jun 2013**

Medical Intern

Hong Kong Baptist University, Hong Kong, Hong Kong SAR **Jun 2013**

Bachelor of Medicine (BM) in Chinese Medicine

Bachelor of Science (BS) in Biomedical Sciences

PEER-REVIEWED JOURNAL PUBLICATIONS

1. Yu, Z., J. Coresh, G. Qi, M. Grams, E. Boerwinkle, H. Snieder, A. Teumer, C. Pattaro, A. Köttgen, N. Chatterjee, and A. Tin. 2020. "A bidirectional Mendelian randomization study supports causal effects of kidney function on blood pressure." *Kidney Int*. doi: 10.1016/j.kint.2020.04.044.
2. Marshall, A. A., A. Zaccardelli, Z. Yu, M. G. Prado, X. Liu, R. Miller Kroouze, S. S. Kalia, R. C. Green, N. A. Triedman, B. Lu, K. D. Deane, M. D. Iversen, E. W. Karlson, and J. A. Sparks. 2019. "Effect of communicating personalized rheumatoid arthritis risk on concern for developing RA: A randomized controlled trial." *Patient Educ Couns* 102 (5):976-983. doi: 10.1016/j.pec.2018.12.011.
3. Solomon, D. H., Z. Yu, J. N. Katz, A. Bitton, C. Corrigan, L. Fraenkel, L. R. Harrold, J. S. Smolen, E. Losina, and B. Lu. 2019. "Adverse Events and Resource Use Before and After Treat-to-Target in Rheumatoid Arthritis: A Post Hoc Analysis of a Randomized Controlled Trial." *Arthritis Care Res (Hoboken)* 71 (9):1243-1248. doi: 10.1002/acr.23755.
4. Teumer, A., Y. Li, S. Ghasemi, B. P. Prins, M. Wuttke, T. Hermle, A. Giri, K. B. Sieber, C. Qiu, H. Kirsten, A. Tin, A. Y. Chu, N. Bansal, M. F. Feitosa, L. Wang, J. F. Chai, M. Cocca, C. Fuchsberger, M. Gorski, A. Hoppmann, K. Horn, M. Li, J. Marten, D. Noce, T. Nutile, S. Sedaghat, G. Sveinbjornsson, B. O. Tayo, P. J. van der Most, Y. Xu, Z. Yu, L. Gerstner, J. Ärnlöv, S. J. L. Bakker, D. Baptista, M. L. Biggs, E. Boerwinkle, H. Brenner, R. Burkhardt, R. J. Carroll, M. L. Chee, M. L. Chee, M. Chen, C. Y. Cheng, J. P. Cook, J. Coresh, T. Corre, J. Danesh, M. H. de Borst, A. De Grandi, R. de Mutsert, A. P. J. de Vries, F. Degenhardt, K. Dittrich, J. Divers, K. U. Eckardt, G. Ehret, K. Endlich, J. F. Felix, O. H. Franco, A. Franke, B. I. Freedman, S. Freitag-Wolf, R. T. Gansevoort, V.

Giedraitis, M. Gögele, F. Grundner-Culemann, D. F. Gudbjartsson, V. Gudnason, P. Hamet, T. B. Harris, A. A. Hicks, H. Holm, V. H. X. Foo, S. J. Hwang, M. A. Ikram, E. Ingelsson, V. W. V. Jaddoe, J. Jakobsdottir, N. S. Josyula, B. Jung, M. Kähönen, C. C. Khor, W. Kiess, W. Koenig, A. Körner, P. Kovacs, H. Kramer, B. K. Krämer, F. Kronenberg, L. A. Lange, C. D. Langefeld, J. J. Lee, T. Lehtimäki, W. Lieb, S. C. Lim, L. Lind, C. M. Lindgren, J. Liu, M. Loeffler, L. P. Lyytikäinen, A. Mahajan, J. C. Maranville, D. Mascalzoni, B. McMullen, C. Meisinger, T. Meitinger, K. Miliku, D. O. Mook-Kanamori, M. Müller-Nurasyid, J. C. Mychaleckyj, M. Nauck, K. Nikus, B. Ning, R. Noordam, J. O. Connell, I. Olafsson, N. D. Palmer, A. Peters, A. I. Podgornaia, B. Ponte, T. Poulain, P. P. Pramstaller, T. J. Rabelink, L. M. Raffield, D. F. Reilly, R. Rettig, M. Rheinberger, K. M. Rice, F. Rivadeneira, H. Runz, K. A. Ryan, C. Sabanayagam, K. U. Saum, B. Schöttker, C. M. Shaffer, Y. Shi, A. V. Smith, K. Strauch, M. Stumvoll, B. B. Sun, S. Szymczak, E. S. Tai, N. Y. Q. Tan, K. D. Taylor, A. Teren, Y. C. Tham, J. Thiery, C. H. L. Thio, H. Thomsen, U. Thorsteinsdottir, A. Tönjes, J. Tremblay, A. G. Uitterlinden, P. van der Harst, N. Verweij, S. Vogelezang, U. Völker, M. Waldenberger, C. Wang, O. D. Wilson, C. Wong, T. Y. Wong, Q. Yang, M. Yasuda, S. Akilesh, M. Bochud, C. A. Böger, O. Devuyst, T. L. Edwards, K. Ho, A. P. Morris, A. Parsa, S. A. Pendergrass, B. M. Psaty, J. I. Rotter, K. Stefansson, J. G. Wilson, K. Susztak, H. Snieder, I. M. Heid, M. Scholz, A. S. Butterworth, A. M. Hung, C. Pattaro, and A. Köttgen. 2019. "Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria." *Nat Commun* 10 (1):4130. doi: 10.1038/s41467-019-11576-0.

5. Tin, A., J. Marten, V. L. Halperin Kuhns, Y. Li, M. Wuttke, H. Kirsten, K. B. Sieber, C. Qiu, M. Gorski, Z. Yu, A. Giri, G. Sveinbjornsson, M. Li, A. Y. Chu, A. Hoppmann, L. J. O'Connor, B. Prins, T. Nutile, D. Noce, M. Akiyama, M. Cocca, S. Ghasemi, P. J. van der Most, K. Horn, Y. Xu, C. Fuchsberger, S. Sedaghat, S. Afaq, N. Amin, J. Ärnlöv, S. J. L. Bakker, N. Bansal, D. Baptista, S. Bergmann, M. L. Biggs, G. Biino, E. Boerwinkle, E. P. Bottinger, T. S. Boutin, M. Brumat, R. Burkhardt, E. Campana, A. Campbell, H. Campbell, R. J. Carroll, E. Catamo, J. C. Chambers, M. Ciullo, M. P. Concas, J. Coresh, T. Corre, D. Cusi, S. C. Felicita, M. H. de Borst, A. De Grandi, R. de Mutsert, A. P. J. de Vries, G. Delgado, A. Demirkan, O. Devuyst, K. Dittrich, K. U. Eckardt, G. Ehret, K. Endlich, M. K. Evans, R. T. Gansevoort, P. Gasparini, V. Giedraitis, C. Gieger, G. Girotto, M. Gögele, S. D. Gordon, D. F. Gudbjartsson, V. Gudnason, T. Haller, P. Hamet, T. B. Harris, C. Hayward, A. A. Hicks, E. Hofer, H. Holm, W. Huang, N. Hutri-Kähönen, S. J. Hwang, M. A. Ikram, R. M. Lewis, E. Ingelsson, J. Jakobsdottir, I. Jonsdottir, H. Jonsson, P. K. Joshi, N. S. Josyula, B. Jung, M. Kähönen, Y. Kamatani, M. Kanai, S. M. Kerr, W. Kiess, M. E. Kleber, W. Koenig, J. S. Kooner, A. Körner, P. Kovacs, B. K. Krämer, F. Kronenberg, M. Kubo, B. Kühnel, M. La Bianca, L. A. Lange, B. Lehne, T. Lehtimäki, J. Liu, M. Loeffler, R. J. F. Loos, L. P. Lyytikäinen, R. Magi, A. Mahajan, N. G. Martin, W. März, D. Mascalzoni, K. Matsuda, C. Meisinger, T. Meitinger, A. Metspalu, Y. Milaneschi, C. J. O'Donnell, O. D. Wilson, J. M. Gaziano, P. P. Mishra, K. L. Mohlke, N. Mononen, G. W. Montgomery, D. O. Mook-Kanamori, M. Müller-Nurasyid, G. N. Nadkarni, M. A. Nalls, M. Nauck, K. Nikus, B. Ning, I. M. Nolte, R. Noordam, J. R. O'Connell, I. Olafsson, S. Padmanabhan, Bwjh Penninx, T. Perls, A. Peters, M. Pirastu, N. Pirastu, G. Pistis, O. Polasek, B. Ponte, D. J. Porteous, T. Poulain,

M. H. Preuss, T. J. Rabelink, L. M. Raffield, O. T. Raitakari, R. Rettig, M. Rheinberger, K. M. Rice, F. Rizzi, A. Robino, I. Rudan, A. Krajcoviechova, R. Cifkova, R. Rueedi, D. Ruggiero, K. A. Ryan, Y. Saba, E. Salvi, H. Schmidt, R. Schmidt, C. M. Shaffer, A. V. Smith, B. H. Smith, C. N. Spracklen, K. Strauch, M. Stumvoll, P. Sulem, S. M. Tajuddin, A. Teren, J. Thiery, C. H. L. Thio, U. Thorsteinsdottir, D. Toniolo, A. Tönjes, J. Tremblay, A. G. Uitterlinden, S. Vaccargiu, P. van der Harst, C. M. van Duijn, N. Verweij, U. Völker, P. Vollenweider, G. Waeber, M. Waldenberger, J. B. Whitfield, S. H. Wild, J. F. Wilson, Q. Yang, W. Zhang, A. B. Zonderman, M. Bochud, J. G. Wilson, S. A. Pendergrass, K. Ho, A. Parsa, P. P. Pramstaller, B. M. Psaty, C. A. Böger, H. Snieder, A. S. Butterworth, Y. Okada, T. L. Edwards, K. Stefansson, K. Susztak, M. Scholz, I. M. Heid, A. M. Hung, A. Teumer, C. Pattaro, O. M. Woodward, V. Vitart, and A. Köttgen. 2019. "Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels." *Nat Genet* 51 (10):1459-1474. doi: 10.1038/s41588-019-0504-x.

6. Wuttke, M., Y. Li, M. Li, K. B. Sieber, M. F. Feitosa, M. Gorski, A. Tin, L. Wang, A. Y. Chu, A. Hoppmann, H. Kirsten, A. Giri, J. F. Chai, G. Sveinbjornsson, B. O. Tayo, T. Nutile, C. Fuchsberger, J. Marten, M. Cocca, S. Ghasemi, Y. Xu, K. Horn, D. Noce, P. J. van der Most, S. Sedaghat, Z. Yu, M. Akiyama, S. Afaq, T. S. Ahluwalia, P. Almgren, N. Amin, J. Ärnlöv, S. J. L. Bakker, N. Bansal, D. Baptista, S. Bergmann, M. L. Biggs, G. Biino, M. Boehnke, E. Boerwinkle, M. Boissel, E. P. Bottinger, T. S. Boutin, H. Brenner, M. Brumat, R. Burkhardt, A. S. Butterworth, E. Campana, A. Campbell, H. Campbell, M. Canouil, R. J. Carroll, E. Catamo, J. C. Chambers, M. L. Chee, M. L. Chee, X. Chen, C. Y. Cheng, Y. Cheng, K. Christensen, R. Cifkova, M. Ciullo, M. P. Concas, J. P. Cook, J.

Coresh, T. Corre, C. F. Sala, D. Cusi, J. Danesh, E. W. Daw, M. H. de Borst, A. De Grandi, R. de Mutsert, A. P. J. de Vries, F. Degenhardt, G. Delgado, A. Demirkan, E. Di Angelantonio, K. Dittrich, J. Divers, R. Dorajoo, K. U. Eckardt, G. Ehret, P. Elliott, K. Endlich, M. K. Evans, J. F. Felix, V. H. X. Foo, O. H. Franco, A. Franke, B. I. Freedman, S. Freitag-Wolf, Y. Friedlander, P. Froguel, R. T. Gansevoort, H. Gao, P. Gasparini, J. M. Gaziano, V. Giedraitis, C. Gieger, G. Girotto, F. Giulianini, M. Gögele, S. D. Gordon, D. F. Gudbjartsson, V. Gudnason, T. Haller, P. Hamet, T. B. Harris, C. A. Hartman, C. Hayward, J. N. Hellwege, C. K. Heng, A. A. Hicks, E. Hofer, W. Huang, N. Hutri-Kähönen, S. J. Hwang, M. A. Ikram, O. S. Indridason, E. Ingelsson, M. Ising, V. W. V. Jaddoe, J. Jakobsdottir, J. B. Jonas, P. K. Joshi, N. S. Josyula, B. Jung, M. Kähönen, Y. Kamatani, C. M. Kammerer, M. Kanai, M. Kastarinen, S. M. Kerr, C. C. Khor, W. Kiess, M. E. Kleber, W. Koenig, J. S. Kooner, A. Körner, P. Kovacs, A. T. Kraja, A. Krajcoviechova, H. Kramer, B. K. Krämer, F. Kronenberg, M. Kubo, B. Kühnel, M. Kuokkanen, J. Kuusisto, M. La Bianca, M. Laakso, L. A. Lange, C. D. Langefeld, J. J. Lee, B. Lehne, T. Lehtimäki, W. Lieb, S. C. Lim, L. Lind, C. M. Lindgren, J. Liu, J. Liu, M. Loeffler, R. J. F. Loos, S. Lucae, M. A. Lukas, L. P. Lyytikäinen, R. Mägi, P. K. E. Magnusson, A. Mahajan, N. G. Martin, J. Martins, W. März, D. Mascalzoni, K. Matsuda, C. Meisinger, T. Meitinger, O. Melander, A. Metspalu, E. K. Mikaelsdottir, Y. Milaneschi, K. Miliku, P. P. Mishra, K. L. Mohlke, N. Mononen, G. W. Montgomery, D. O. Mook-Kanamori, J. C. Mychaleckyj, G. N. Nadkarni, M. A. Nalls, M. Nauck, K. Nikus, B. Ning, I. M. Nolte, R. Noordam, J. O'Connell, M. L. O'Donoghue, I. Olafsson, A. J. Oldehinkel, M. Orho-Melander, W. H. Ouwehand, S. Padmanabhan, N. D. Palmer, R. Palsson, Bwjh Penninx, T. Perls, M. Perola, M. Pirastu, N. Pirastu, G. Pistis, A. I.

Podgornaia, O. Polasek, B. Ponte, D. J. Porteous, T. Poulain, P. P. Pramstaller, M. H. Preuss, B. P. Prins, M. A. Province, T. J. Rabelink, L. M. Raffield, O. T. Raitakari, D. F. Reilly, R. Rettig, M. Rheinberger, K. M. Rice, P. M. Ridker, F. Rivadeneira, F. Rizzi, D. J. Roberts, A. Robino, P. Rossing, I. Rudan, R. Rueedi, D. Ruggiero, K. A. Ryan, Y. Saba, C. Sabanayagam, V. Salomaa, E. Salvi, K. U. Saum, H. Schmidt, R. Schmidt, B. Schöttker, C. A. Schulz, N. Schupf, C. M. Shaffer, Y. Shi, A. V. Smith, B. H. Smith, N. Soranzo, C. N. Spracklen, K. Strauch, H. M. Stringham, M. Stumvoll, P. O. Svensson, S. Szymczak, E. S. Tai, S. M. Tajuddin, N. Y. Q. Tan, K. D. Taylor, A. Teren, Y. C. Tham, J. Thiery, C. H. L. Thio, H. Thomsen, G. Thorleifsson, D. Toniolo, A. Tönjes, J. Tremblay, I. Tzoulaki, A. G. Uitterlinden, S. Vaccargiu, R. M. van Dam, P. van der Harst, C. M. van Duijn, D. R. Velez Edward, N. Verweij, S. Vogelezang, U. Völker, P. Vollenweider, G. Waeber, M. Waldenberger, L. Wallentin, Y. X. Wang, C. Wang, D. M. Waterworth, W. Bin Wei, H. White, J. B. Whitfield, S. H. Wild, J. F. Wilson, M. K. Wojczynski, C. Wong, T. Y. Wong, L. Xu, Q. Yang, M. Yasuda, L. M. Yerges-Armstrong, W. Zhang, A. B. Zonderman, J. I. Rotter, M. Bochud, B. M. Psaty, V. Vitart, J. G. Wilson, A. Dehghan, A. Parsa, D. I. Chasman, K. Ho, A. P. Morris, O. Devuyst, S. Akilesh, S. A. Pendergrass, X. Sim, C. A. Böger, Y. Okada, T. L. Edwards, H. Snieder, K. Stefansson, A. M. Hung, I. M. Heid, M. Scholz, A. Teumer, A. Köttgen, and C. Pattaro. 2019. "A catalog of genetic loci associated with kidney function from analyses of a million individuals." *Nat Genet* 51 (6):957-972. doi: 10.1038/s41588-019-0407-x.

7. Yu, Z., S. C. Kim, K. Vanni, J. Huang, R. Desai, S. N. Murphy, D. H. Solomon, and K. P. Liao. 2019. "Correction to: Association between inflammation and systolic blood

pressure in RA compared to patients without RA." *Arthritis Res Ther* 21 (1):170. doi: 10.1186/s13075-019-1940-9.

8. Yu, Z., C. M. Rebholz, E. Wong, Y. Chen, K. Matsushita, J. Coresh, and M. E. Grams. 2019. "Association Between Hypertension and Kidney Function Decline: The Atherosclerosis Risk in Communities (ARIC) Study." *Am J Kidney Dis* 74 (3):310-319. doi: 10.1053/j.ajkd.2019.02.015.
9. Han, Z., Z. Xiao, K. Kalantar-Zadeh, H. Moradi, T. Shafi, S. S. Waikar, L. D. Quarles, Z. Yu, A. Tin, J. Coresh, and C. P. Kovesdy. 2018. "Validation of a Novel Modified Aptamer-Based Array Proteomic Platform in Patients with End-Stage Renal Disease." *Diagnostics (Basel)* 8 (4). doi: 10.3390/diagnostics8040071.
10. Kim, S. C., M. F. Di Carli, R. K. Garg, K. Vanni, P. Wang, A. Wohlfahrt, Z. Yu, F. Lu, A. Campos, C. F. Bibbo, S. Smith, and D. H. Solomon. 2018. "Asymptomatic hyperuricemia and coronary flow reserve in patients with metabolic syndrome." *BMC Rheumatol* 2:17. doi: 10.1186/s41927-018-0027-6.
11. Prado, M. G., M. D. Iversen, Z. Yu, R. Miller Kroouze, N. A. Triedman, S. S. Kalia, B. Lu, R. C. Green, E. W. Karlson, and J. A. Sparks. 2018. "Effectiveness of a Web-Based Personalized Rheumatoid Arthritis Risk Tool With or Without a Health Educator for Knowledge of Rheumatoid Arthritis Risk Factors." *Arthritis Care Res (Hoboken)* 70 (10):1421-1430. doi: 10.1002/acr.23510.
12. Solomon, D. H., B. Lu, Z. Yu, C. Corrigan, L. R. Harrold, J. S. Smolen, L. Fraenkel, J. N. Katz, and E. Losina. 2018. "Benefits and Sustainability of a Learning Collaborative for Implementation of Treat-to-Target in Rheumatoid Arthritis: Results of a Cluster-

Randomized Controlled Phase II Clinical Trial." *Arthritis Care Res (Hoboken)* 70 (10):1551-1556. doi: 10.1002/acr.23508.

13. Sparks, J. A., M. D. Iversen, Z. Yu, N. A. Triedman, M. G. Prado, R. Miller Kroouze, S. S. Kalia, M. L. Atkinson, E. A. Mody, S. M. Helfgott, D. J. Todd, P. F. Dellaripa, B. L. Bermas, K. H. Costenbader, K. D. Deane, B. Lu, R. C. Green, and E. W. Karlson. 2018. "Disclosure of Personalized Rheumatoid Arthritis Risk Using Genetics, Biomarkers, and Lifestyle Factors to Motivate Health Behavior Improvements: A Randomized Controlled Trial." *Arthritis Care Res (Hoboken)* 70 (6):823-833. doi: 10.1002/acr.23411.
14. Wang, P., S. E. Smith, R. Garg, F. Lu, A. Wohlfahrt, A. Campos, K. Vanni, Z. Yu, D. H. Solomon, and S. C. Kim. 2018. "Identification of monosodium urate crystal deposits in patients with asymptomatic hyperuricemia using dual-energy CT." *RMD Open* 4 (1):e000593. doi: 10.1136/rmdopen-2017-000593.
15. Wei, J., R. Hou, Y. Xi, A. Kowalski, T. Wang, Z. Yu, Y. Hu, E. K. Chandrasekar, H. Sun, and M. K. Ali. 2018. "The association and dose-response relationship between dietary intake of α-linolenic acid and risk of CHD: a systematic review and meta-analysis of cohort studies." *Br J Nutr* 119 (1):83-89. doi: 10.1017/s0007114517003294.
16. Yoshida, K., Z. Yu, G. A. Greendale, K. Ruppert, Y. Lian, S. K. Tedeschi, T. C. Lin, S. Haneuse, R. J. Glynn, S. Hernández-Díaz, and D. H. Solomon. 2018. "Effects of analgesics on bone mineral density: A longitudinal analysis of the prospective SWAN cohort with three-group matching weights." *Pharmacoepidemiol Drug Saf* 27 (2):182-190. doi: 10.1002/pds.4362.
17. Yu, Z., S. C. Kim, K. Vanni, J. Huang, R. Desai, S. N. Murphy, D. H. Solomon, and K. P. Liao. 2018. "Association between inflammation and systolic blood pressure in RA

compared to patients without RA." *Arthritis Res Ther* 20 (1):107. doi: 10.1186/s13075-018-1597-9.

18. Yu, Z., S. H. Ley, Q. Sun, F. B. Hu, and V. S. Malik. 2018. "Cross-sectional association between sugar-sweetened beverage intake and cardiometabolic biomarkers in US women." *Br J Nutr* 119 (5):570-580. doi: 10.1017/s0007114517003841.
19. Yu, Z., B. Lu, J. Agosti, A. Bitton, C. Corrigan, L. Fraenkel, L. R. Harrold, E. Losina, J. N. Katz, and D. H. Solomon. 2018. "Implementation of Treat-to-Target for Rheumatoid Arthritis in the US: Analysis of Baseline Data From a Randomized Controlled Trial." *Arthritis Care Res (Hoboken)* 70 (5):801-806. doi: 10.1002/acr.23343.
20. Yu, Z., N. Yang, B. M. Everett, M. Frits, C. Iannaccone, J. Coblyn, M. Weinblatt, N. Shadick, D. H. Solomon, and K. P. Liao. 2018. "Impact of Changes in Inflammation on Estimated Ten-Year Cardiovascular Risk in Rheumatoid Arthritis." *Arthritis Rheumatol* 70 (9):1392-1398. doi: 10.1002/art.40532.
21. Zak, A., C. Corrigan, Z. Yu, A. Bitton, L. Fraenkel, L. Harrold, J. S. Smolen, and D. H. Solomon. 2018. "Barriers to treatment adjustment within a treat to target strategy in rheumatoid arthritis: a secondary analysis of the TRACTION trial." *Rheumatology (Oxford)* 57 (11):1933-1937. doi: 10.1093/rheumatology/key179.
22. Lee, Y. C., A. S. Karlamangla, Z. Yu, C. C. Liu, J. S. Finkelstein, G. A. Greendale, S. D. Harlow, and D. H. Solomon. 2017. "Pain Severity in Relation to the Final Menstrual Period in a Prospective Multiethnic Observational Cohort: Results From the Study of Women's Health Across the Nation." *J Pain* 18 (2):178-187. doi: 10.1016/j.jpain.2016.10.012.

23. Solomon, D. H., E. Losina, B. Lu, A. Zak, C. Corrigan, S. B. Lee, J. Agosti, A. Bitton, L. R. Harrold, T. Pincus, H. Radner, Z. Yu, J. S. Smolen, L. Fraenkel, and J. N. Katz. 2017. "Implementation of Treat-to-Target in Rheumatoid Arthritis Through a Learning Collaborative: Results of a Randomized Controlled Trial." *Arthritis Rheumatol* 69 (7):1374-1380. doi: 10.1002/art.40111.
24. Wang, Y., M. Lee, P. Liu, L. Shi, Z. Yu, Y. Abu Awad, A. Zanobetti, and J. D. Schwartz. 2017. "Doubly Robust Additive Hazards Models to Estimate Effects of a Continuous Exposure on Survival." *Epidemiology* 28 (6):771-779. doi: 10.1097/ede.0000000000000742.
25. Yu, Z., T. Huang, Y. Zheng, T. Wang, Y. Heianza, D. Sun, H. Campos, and L. Qi. 2017. "PCSK9 variant, long-chain n-3 PUFAs, and risk of nonfatal myocardial infarction in Costa Rican Hispanics." *Am J Clin Nutr* 105 (5):1198-1203. doi: 10.3945/ajcn.116.148106.
26. Afshin, A., D. Babalola, M. McLean, Z. Yu, W. Ma, C. Y. Chen, M. Arabi, and D. Mozaffarian. 2016. "Information Technology and Lifestyle: A Systematic Evaluation of Internet and Mobile Interventions for Improving Diet, Physical Activity, Obesity, Tobacco, and Alcohol Use." *J Am Heart Assoc* 5 (9). doi: 10.1161/jaha.115.003058.
27. Yu, Z., V. S. Malik, N. Keum, F. B. Hu, E. L. Giovannucci, M. J. Stampfer, W. C. Willett, C. S. Fuchs, and Y. Bao. 2016. "Associations between nut consumption and inflammatory biomarkers." *Am J Clin Nutr* 104 (3):722-8. doi: 10.3945/ajcn.116.134205.

MANUSCRIPTS

1. Yu Z, Grams ME, Ndumele CE, Boerwinkle E, North K, Rebholz CM, Giovannucci EL, Coresh J. Association between midlife obesity and kidney function trajectories: the Atherosclerosis Risk in Communities (ARIC) Study. *Under review.*
2. Iversen MD, Prado MG, Yu Z, Triedman NA, Miller Kroouze R, Kalia SS, Costenbader KH, Lu B, Green RC, Sparks JA, Karlson EW. Impact of personalized rheumatoid arthritis (RA) risk education on decisional balance for health behaviors associated with the risk of developing RA. *Under review.*
3. Yu Z, Jin J, Tin A, Köttgen A, Yu B, Chen J, Ballantyne C, Hoogeveen R, Arking D, Chatterjee N, Grams M, Coresh J. Polygenic risk scores of kidney function, circulating proteome, and incident kidney diseases: the Atherosclerosis Risk in Communities (ARIC) Study. *In preparation.*
4. Yu Z, Wongvibulsin S, Matsushita K, Arking DE, Coresh J, Zeger SL. Machine learning for sudden cardiac death (SCD) prediction: the Atherosclerosis Risk in Communities (ARIC) Study. *In preparation.*

PRESENTATIONS

American Society of Human Genetics (ASHG) Annual Meeting, Houston, US Oct 2019

Kidney function and blood pressure: A Mendelian randomization study. Poster.

Society for Epidemiologic Research (SER) Annual Meeting, Baltimore, US Jun 2018

Association between midlife obesity and kidney function trajectories: the Atherosclerosis Risk in Communities Study. Poster.

American College of Rheumatology (ACR) Annual Meeting, Washington, US Nov 2016

Association between inflammation and systolic blood pressure at normal and elevated C-reactive protein levels. Poster.

Boston Nutrition Obesity Research Center Annual Symposium, Boston, US Aug 2015

Association between nut consumption and inflammatory biomarkers. Poster.

AD HOC PEER REVIEW ACTIVITIES

Arthritis Research and Therapy	2019 & 2020
Clinical Journal of the American Society of Nephrology (CJASN)	2020
British Medical Journal (BMJ)	2020
British Journal of Nutrition (BJN)	2020
Journal of Clinical Densitometry	2020
Diabetes Care	2018 & 2019
Journal of the American Heart Association (JAHA)	2017 & 2019
Public Health Nutrition	2019
Heliyon	2019
BMJ Open	2018
Journal of Hospital Medicine	2018
Blood Purification	2018
BioMed Central (BMC) Nutrition	2015

REVIEW SERVICES

Atherosclerosis Risk in Communities Study (ARIC) Publications Committee	2019 - 2020
Society for Epidemiologic Research (SER) Abstract Review	2020

SELECTED HONORS AND AWARDS

Delta Omega Scholarship, Johns Hopkins Bloomberg School of Public Health 2020

Travel Award, Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium 2019

Comstock Training Fund, Johns Hopkins Bloomberg School of Public Health 2019

Ellen B. Gold Fund for Epidemiology, Johns Hopkins Bloomberg School of Public Health 2019

President Honor Roll, Hong Kong Baptist University 2008 - 2013

Lee Koon Shin Scholarship 2010 & 2013

First Class Honors, Hong Kong Baptist University 2013

Scholastic Award (highest academic standing), Hong Kong Baptist University 2013

Chiu Chow Chamber of Commerce Limited Scholarship, Hong Kong Baptist University 2013

C.V. Starr Scholarship Fund, Starr Foundation 2012

School of Chinese Medicine Scholarship, Hong Kong Baptist University 2012

Hong Kong Association of University Women (HKAUM) Scholarship, HKAUM 2011

Scholarship for Excellent Mainland Student (full scholarship), Hong Kong Baptist University 2008

TEACHING EXPERIENCES

Teaching Assistant at Johns Hopkins Bloomberg School of Public Health:

Statistical Methods in Public Health III

Epidemiologic Methods II

Epidemiology of Diabetes and Obesity

Principles of Epidemiology

Epidemiology in Evidence-Based Policy

Analysis of Longitudinal Data

Data Analysis Workshop I Data Analysis Workshop II

SKILLS

Computer: R, Python, SAS, Unix, plink, git, Latex

Experiment: Western blotting, cell culture