Association of Rare Variants in Kidney Developmental Genes With Chronic Kidney Disease and Blood Pressure: A UK Biobank Study

Benjamin L. Spector, MD; Byunggil Yoo, MS; Neil Miller, PhD; Monica Gaddis, PhD; Isabelle Thiffault, PhD; Laurel Willig, MD

ABSTRACT

Introduction: Chronic kidney disease (CKD) and hypertension are heritable traits. The source of this heritability remains largely unknown, and exploration has been limited principally to common genetic variants, with few studies having examined rare variants.

Methods: In this cross-sectional observational study, we evaluate whole exome sequencing data using the UK Biobank to identify the ability of rare variants in 58 kidney developmental genes to predict CKD or elevated blood pressure using logistic regression models with subgroup analysis performed by ancestry.

Results: Significant predictors of CKD included rare variants in *CLCN5* (OR 1.59; 99% CI, 1.02–2.47; $P = 0.007$). Predictors of blood pressure included rare variants in *SIX1* (OR 0.57; 99% CI, 0.35–0.94; $P = 0.004$) and *NPHS1* (OR 0.84; 99% CI, 0.72–0.99; $P = 0.005$), which were protective against blood pressure elevation, and *WT1* (OR 1.58; 99% CI, 1.02–2.45; $P = 0.007$), which was associated with elevated blood pressure. In individuals of White British ancestry, rare variants in *SIX1* protected against elevated blood pressure (OR 0.58; 99% CI, 0.34–0.99; $P = 0.009$). Among individuals of non-White British ancestry, predictors of CKD included rare variants in *SLC12A3* (OR 2.02; 99% CI, 1.08–3.78; $P = 0.004$) and *CALB1* (OR 3.12; 99% CI, 1.15–8.47; $P = 0.003$). Presence of rare variants in *WT1* significantly predicted elevated blood pressure (OR 2.49; 99% CI, 1.08–5.78; $P = 0.005$).

Conclusions: From this study, we conclude that rare variants in kidney developmental genes contribute to the risk of developing CKD and elevated blood pressure. These associations vary by ancestry.

• • •

Author Affiliations: Department of Pediatrics, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin (Spector); Center for Genomic Medicine, Children's Mercy Hospital, Kansas City, Missouri (Yoo, Thiffault, Willig); Bionano Genomics, Inc, San Diego, California (Miller); Department of Emergency Medicine, University of Missouri-Kansas City School of Medicine, Kansas City, Missouri (Gaddis); Department of Pathology and Laboratory Medicine, Children's Mercy Hospital, Kansas City, Missouri (Thiffault); University of Missouri-Kansas City School of Medicine, Kansas City, Missouri (Thiffault); Division of Nephrology, Children's Mercy Hospital, Kansas City, Missouri (Willig).

Corresponding Author: Benjamin L. Spector, MD, Department of Pediatrics, University of Wisconsin School of Medicine and Public Health, 600 Highland Ave, Madison, WI 53792; email bspector2@wisc.edu; ORCID ID 0000-0003-1800-4715

INTRODUCTION

Nearly 15% of the United States population suffers from chronic kidney disease (CKD), with approximately 10% having CKD stage 3 or above, generating Medicare costs exceeding $75 billion annually.[1-3] Despite this high prevalence and economic burden, development of novel CKD treatments has largely stagnated.

CKD and elevated blood pressure are complexly interrelated. Hypertension is among the sequelae of kidney dysfunction due to several pathophysiologic mechanisms, including hyperactivity of the renin-angiotensin-aldosterone access, sodium retention causing fluid overload, and increased sympathetic nervous system activity. Conversely, hypertension is an independent predictor of reduced kidney function. Evidence indicates that an individual's total nephron number, termed "nephron endowment," is inversely correlated to the risk of hypertension and CKD.[4,5]

Genetic studies of congenital renal hypoplasia and hypodysplasia, an extreme form of low nephron endowment, suggest that nephron endowment is genetically predetermined.[6] Similarly, studies have found CKD and estimated glomerular filtration rate (eGFR) have 30% to 70% heritability,[7-10] and blood pressure is 20% to 50% heritable.[11] Genome-wide association studies—the primary method employed to identify the heritability of these traits—have identified a few common variants in kidney developmental genes associated with CKD, supporting the theory that lower nephron endowment may contribute to the general population's CKD risk.[8,12-14] However, much of the heritability of CKD

Table 1. Demographic Features of Individuals With and Without Elevated Blood Pressure

	All (n = 49 989)			White British (n = 41 275)			Non-White British (n = 8714)		
	Present	Absent	*P* value	Present	Absent	*P* value	Present	Absent	*P* value
Elevated blood pressure	42 167 (84.4%)	7822 (15.6%)		35 111 (85.1%)	6164 (14.9%)		7056 (81.0%)	1658 (19.0%)	
Vascular heart disease	5165 (12.2%)	267 (3.4%)	< 0.001	4293 (12.2%)	222 (3.6%)	< 0.001	872 (12.4%)	45 (2.7%)	< 0.001
Diabetes	3732 (8.9%)	191 (2.4%)	< 0.001	2905 (8.3%)	142 (2.3%)	< 0.001	827 (11.7%)	49 (3.0%)	< 0.001
Hyperlipidemia	11 625 (27.6%)	587 (7.5%)	< 0.001	9730 (27.7%)	466 (7.6%)	< 0.001	1895 (26.9%)	121 (7.3%)	< 0.001
Overweight	30 997 (73.5%)	3560 (45.5%)	< 0.001	25 797 (73.5%)	2825 (45.8%)	< 0.001	5200 (73.7%)	735 (44.3%)	< 0.001
Smoker	4223 (10.0%)	903 (11.5%)	< 0.001	3385 (9.6%)	667 (10.8%)	0.004	838 (11.9%)	236 (14.2%)	0.009

Values are displayed as n (%) unless otherwise indicated.

Table 2. Demographic Features of Individuals With and Without Chronic Kidney Disease

	All (n = 49 989)			White British (n = 41 275)			Non-White British (n = 8714)		
	Present	Absent	*P* value	Present	Absent	*P* value	Present	Absent	*P* value
Chronic kidney diease	1060 (2.1%)	48 929 (97.9%)		870 (2.1%)	40 405 (97.9%)		190 (2.2%)	8524 (97.8%)	
Vascular heart disease	363 (34.2%)	5069 (10.4%)	< 0.001	295 (33.9%)	4220 (10.4%)	< 0.001	68 (35.8%)	849 (10.0%)	< 0.001
Diabetes	242 (22.8%)	3681 (7.5%)	< 0.001	187 (21.5%)	2860 (7.1%)	< 0.001	55 (28.9%)	821 (9.6%)	< 0.001
Hyperlipidemia	607 (57.3%)	11 605 (23.7%)	< 0.001	511 (58.7%)	9685 (24.0%)	< 0.001	96 (50.5%)	1920 (22.5%)	< 0.001
Overweight	869 (82.0%)	33 688 (68.9%)	< 0.001	710 (81.6%)	27 912 (69.1%)	< 0.001	159 (83.7%)	5776 (67.8%)	< 0.001
Smoker	115 (10.8%)	5011 (10.2%)	0.52	93 (10.7%)	3959 (9.8%)	0.38	22 (11.6%)	1052 (12.3%)	0.75
Elevated blood pressure	998 (94.2%)	41 169 (84.1%)	< 0.001	822 (94.5%)	34 289 (84.9%)	< 0.001	176 (92.6%)	6880 (80.7%)	< 0.001

Values are displayed as n (%) unless otherwise indicated.

and elevated blood pressure remains elusive. Few studies have examined the role of rare variants in explaining this missing heritability.

We aimed to identify genes of kidney development in which rare variants are predictive of blood pressure outcomes or CKD. We accomplished this by using the UK Biobank, a biorepository containing genetic information linked to the electronic health records of approximately 500 000 volunteer participants, to examine the relationship between rare variants in kidney developmental genes and kidney dysfunction, including blood pressure elevation and CKD. We hypothesized that very rare variants in genes implicated in nephrogenesis result in abnormal nephron development and decreased nephron endowment, thereby leading to increased risk of elevated blood pressure and CKD.

METHODS

Study Population

This research was conducted using the UK Biobank Resource, a biorepository of volunteer participants aged 40 to 69 years that links genomic data with deidentified electronic medical record information under application 65332. Analysis was limited to those 49 989 individuals for whom whole exome sequencing data were available at the time of access to the biorepository. Subgroup analyses by ancestry were carried out for those identified as having White British ancestry and those with non-White British ancestry. Individuals were stratified as White British based on principal component analysis carried out by the UK Biobank indicating similar genetic ancestry in addition to self-identifying as "White British." This study was determined to be nonhuman subject research by the Children's Mercy Hospital Institutional Review Board under application STUDY00001390.

Study Variables

Phenotypes of Interest

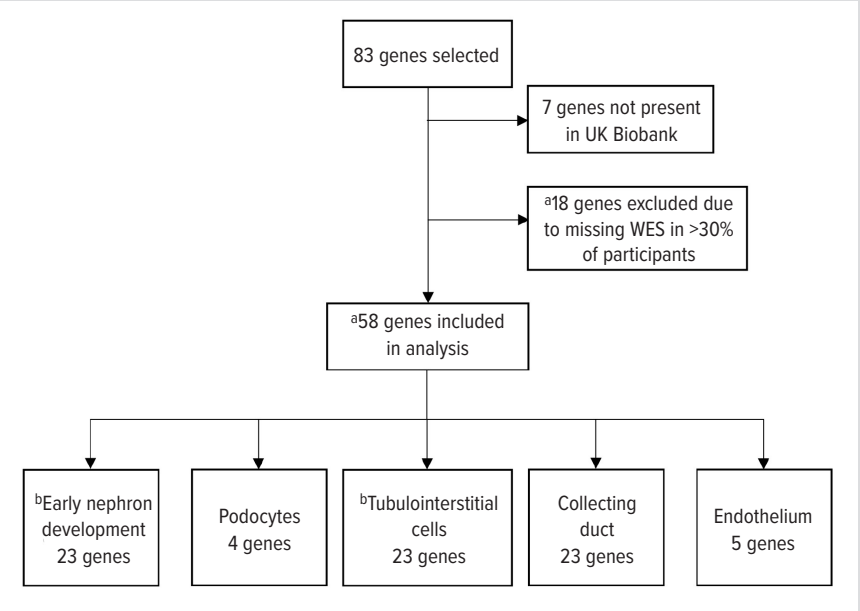
The primary outcomes of interest included the categorical variables of elevated blood pressure and CKD. Presence of elevated blood pressure was determined through use of *International Classification of Diseases 9th Revision* and *10th Revision* (ICD-9 and ICD-10, respectively) codes, reported high blood pressure, reported use of antihypertensive agents, presence of a numeric value in the field "age high blood pressure diagnosed," systolic blood pressure ≥130 mmHg, or diastolic blood pressure ≥80 mmHg. Individuals were considered to have CKD through use of ICD-9 and ICD-10 codes indicating CKD stages 3-5 or end stage kidney disease, calculated eGFR < 60 mL/min/1.73 m^2 using the CKD-EPI 2021 equation,[15] or presence of an end stage kidney disease report (Appendix 1).

Other clinical covariates included the categorical variables of current smokers, diabetes, vascular heart disease (stroke, angina, myocardial infarction), hyperlipidemia, and overweight defined as body mass index (BMI) ≥ 25 kg/m^2 or by applicable ICD-9 and ICD-10 codes (Appendix 1). Criteria used to identify the presence of these covariates, including ICD-9 and ICD-10 codes and other available parameters, are summarized in Appendix 1.

Selection of Genes

Eighty-three candidate genes were selected for analysis based on prior studies demonstrating their associations with renal development and kidney function.[12] Genes were categorized according to their contributions to 5 structural compartments in kidney development: (1) early nephron development, (2) podocytes, (3) tubulointerstitial cells, (4) collecting duct, or (5) endothelium (Appendix 2). As there are an estimated 11% of variants missing from the reported UK Biobank whole exome sequencing data,[16] only those genes with reported variants in ≥70% of the study population or subgroup of interest were included in final analysis to ensure adequate representation of the cohort.

Figure 1. Flowchart of Gene Selection for Analysis



Abbreviation: WES, whole exome sequencing.
[a]In non-White British individuals, 19 genes were excluded due to missing WES in >30% of participants, resulting in 57 unique genes in the non-White British subgroup analysis.
[b]*LEF1* and *GDNF* were included in the analysis of the early nephron development and tubulointerstitial cells developmental compartments.

Definition of Qualifying Variants

Qualifying rare variants were defined as those with minor allele frequency <0.1% and classified as nonbenign. Variants were annotated using the multistage variant characterization pipeline, Rapid Understanding of Nucleotide variant Effect Software (RUNES),[17] which incorporates Variant Effect Predictor,[18] comparisons to the National Center for Biotechnology Information dbSNP,[19] known disease-causing variants from the Human Gene Mutation Database (HGMD),[20] and in silico prediction of variant consequences with RefSeq[21] and Ensembl[22] annotations. Using this RUNES pipeline, variants were categorized as nonbenign if they were reported previously in HGMD and/or ClinVar (category 1), previously unreported but expected to be pathogenic (category 2), or were a variant of uncertain significance (category 3). Variants were designated as nonqualifying if minor allele frequency was ≥0.1% or if the variant was predicted not to cause disease (category 4) or was known to be neutral and/or benign (category 5). Further detail regarding RUNES categorization is summarized in Appendix 3.

Statistical Analysis

Clinical Covariate Distribution

Chi-square analysis was used to compare the proportions of individuals with clinical covariates outside of the primary outcomes of interest in presence versus absence of disease (Tables 1 and 2).

Association of Qualifying Variants with Outcomes of Interest

To determine association between presence of qualifying variants in different kidney developmental compartments and the primary phenotypes of interest – elevated blood pressure and CKD – binary logistic regression models were created with separate models for each compartment: early nephron development, podocytes, tubulointerstitial cells, collecting duct, and endothelium. Genes included in regression models were those with $P<0.25$ in chi-square univariate analysis assessing association of qualifying variants and the primary outcome of interest. This process was performed separately for each ancestral subgroup analysis (White British and non-White British subgroups). Logistic regression was not carried out for compartments if no genes met the predetermined threshold for inclusion during univariate analysis.

For each outcome of interest, logistic regression analyses included known modifiable risk factors for the disease process. In the case of elevated blood pressure, modifiable risk factors accounted for in the logistic regression models included presence of vascular heart disease, diabetes, overweight, hyperlipidemia, and current smoking status. Given the known associations of vascular heart disease, diabetes, and elevated blood pressure with CKD, these modifiable risk factors were accounted for in the logistic regression models examining CKD. Principal component analysis was performed by the UK Biobank in assignment of ancestry, so accordingly did not require inclusion in our regression models. The degrees of variance of these traits explained by our logistic regression models were calculated by Cox and Snell's R^2 and Nagelkerke's R^2.

Statistical significance was set at a Bonferroni-corrected critical α-level of 0.01 to account for multiple comparisons.

Qualifying Variant Distribution

We hypothesized that individuals with outcomes of interest pos-

Table 3. Logistic Regression Models Assessing the Association of Qualifying Variants With Presence of Elevated Blood Pressure by Developmental Compartment

	All			White British			Non-White British		
Compartment	**Gene**	**OR (99% CI)**	***P* value**	**Gene**	**OR (99% CI)**	***P* value**	**Gene**	**OR (99% CI)**	***P* value**
Early nephron development	*ANPEP*	0.87 (0.70–1.08)	0.10	*CITED1*	0.66 (0.35–1.23)	0.085	*CRABP2*	0.73 (0.32–1.68)	0.33
	CITED1	0.74 (0.42–1.30)	0.16	*COL2A1*	0.92 (0.74–1.15)	0.34	*EYA1*	1.29 (0.78–2.12)	0.19
	COL2A1	0.86 (0.71–1.04)	0.043	*CRABP2*	1.67 (0.77–3.62)	0.086	*PAX2*	1.97 (0.90–4.34)	0.026
	ETV4	1.55 (0.73–3.29)	0.13	*ETV4*	2.21 (0.84–5.82)	0.036	***WT1***	**2.49 (1.08–5.78)**	**0.005**
	JAG1	0.98 (0.77–1.25)	0.82	*LEF1*	0.84 (0.61–1.15)	0.14			
	LEF1	0.89 (0.68–1.18)	0.29	***SIX1***	**0.58 (0.34–0.99)**	**0.009**			
	SIX1	**0.57 (0.35–0.94)**	**0.004**						
	WT1	**1.58 (1.02–2.45)**	**0.007**						
Podocytes	***NPHS1***	**0.84 (0.72–0.99)**	**0.005**	*NPHS1*	0.823 (0.675–1.004)	0.011			
	PODXL	0.86 (0.66–1.11)	0.12						
Tubulointerstitial cells	*CD248*	1.22 (0.87–1.71)	0.13	*CLCN5*	0.846 (0.670–1.070)	0.066	*CLCN5*	0.67 (0.44–1.02)	0.013
	CLCN5	0.87 (0.70–1.07)	0.079	*COL1A1*	0.861 (0.693–1.069)	0.075	*COL3A1*	1.10 (0.75–1.61)	0.53
	COL1A1	0.95 (0.79–1.15)	0.49	*CSPG4*	0.933 (0.778–1.119)	0.32	*CUBN*	1.21 (0.89–1.66)	0.11
	CSPG4	0.91 (0.78–1.07)	0.15	*DES*	1.233 (0.742–2.050)	0.29	*DES*	1.39 (0.81–2.39)	0.11
	DES	1.27 (0.84–1.90)	0.14	*FOXD1*	0.896 (0.676–1.188)	0.32	*FOXD1*	1.20 (0.76–1.88)	0.30
	LEF1	0.80 (0.61–1.05)	0.035	*LEF1*	0.823 (0.598–1.133)	0.12	*LRP2*	0.94 (0.76–1.16)	0.43
	PAPPA2	1.07 (0.90–1.30)	0.33	*PAPPA2*	1.106 (0.892–1.372)	0.23	*MCAM*	1.13 (0.69–1.83)	0.53
	SERPINE2	0.82 (0.58–1.15)	0.13	*SLC12A1*	1.180 (0.937–1.487)	0.064	*PDGFRB*	0.96 (0.69–1.35)	0.77
	SLC12A1	1.16 (0.94–1.42)	0.067	*TPM2*	1.730 (0.991–3.018)	0.011	*TPM2*	1.43 (0.77–2.65)	0.13
	SLC12A3	0.93 (0.79–1.10)	0.25						
	TPM2	1.38 (0.89–2.15)	0.061						
Collecting duct	*KRT8*	1.11 (0.88–1.40)	0.27	*CALB1*	1.133 (0.787–1.631)	0.38	*KRT8*	1.18 (0.78–1.78)	0.31
				KRT8	1.104 (0.819–1.486)	0.39			
Endothelium	*FLT1*	0.91 (0.78–1.07)	0.14				*KDR*	1.13 (0.83–1.56)	0.31
							TEK	1.10 (0.72–1.68)	0.58

Table 4. Logistic Regression Models Assessing the Association of Qualifying Variants with Presence of Chronic Kidney Disease by Developmental Compartment

	All			White British			Non-White British		
Compartment	**Gene**	**OR (99% CI)**	***P* value**	**Gene**	**OR (99% CI)**	***P* value**	**Gene**	**OR (99% CI)**	***P* value**
Early nephron development	*CITED1*	1.46 (0.45–4.78)	0.41	*ALDH1A1*	1.143 (0.574–2.277)	0.019	*CCND1*	2.50 (0.64–9.82)	0.085
	HNF1A	0.34 (0.11–1.08)	0.016	*CITED1*	1.429 (0.380–5.378)	0.23	*COL2A1*	0.37 (0.10–1.39)	0.054
	ITGA8	0.79 (0.53–1.20)	0.15	*HNF1A*	0.182 (0.029–1.137)	0.19	*CRABP2*	3.23 (0.63–16.63)	0.065
	PAX8	0.48 (0.13–1.78)	0.15	*SIM2*	0.581 (0.242–1.394)	0.037	*EYA1*	1.47 (0.49–4.47)	0.37
	SOX9	0.59 (0.22–1.59)	0.17	*SOX9*	0.414 (0.112–1.525)	0.058	*LEF1*	1.29 (0.42–3.89)	0.56
	WT1	0.70 (0.26–1.89)	0.35				*PAX2*	0.00	1.00
							PDGFRB	1.42 (0.66–3.06)	0.25
							SALL1	1.95 (0.80–4.73)	0.054
							SIM2	1.14 (0.46–2.83)	0.72
Podocytes	*PODXL*	1.47 (0.86–2.51)	0.065	*NPHS2*	0.49 (0.11–2.22)	0.23	*PODXL*	1.83 (0.83–4.06)	0.050
Tubulointerstitial cells	*CDH1*	1.39 (0.78–2.46)	0.14	*CLCN5*	1.49 (0.96–2.31)	0.019	*CDH1*	2.15 (0.80–5.76)	0.045
	CLCN5	**1.59 (1.02–2.47)**	**0.007**	*COL3A1*	1.25 (0.78–2.01)	0.23	*CLCN5*	2.11 (0.87 –5.09)	0.03
	CLDN1	0.57 (0.090–3.57)	0.43	*CSPG4*	0.80 (0.52–1.24)	0.19	*CLDN1*	0.00	1.00
	COL3A1	1.15 (0.71–1.86)	0.46	*LRP2*	0.77 (0.55–1.06)	0.037	*COL1A1*	1.46 (0.75–2.82)	0.15
	CSPG4	0.82 (0.54–1.24)	0.21	*SLC12A3*	0.71 (0.45–1.13)	0.058	*DES*	1.20 (0.42–3.40)	0.66
	FOXD1	1.27 (0.74–2.17)	0.26	*TPM2*	0.30 (0.048–1.88)	0.091	*FOXD1*	2.03 (0.90–4.57)	0.025
	UMOD	1.18 (0.67–2.06)	0.45				*LEF1*	1.80 (0.67–4.82)	0.13
							LRP2	1.21 (0.73–2.00)	0.33
							PAPPA2	1.49 (0.75–2.98)	0.14
							SERPINE2	0.50 (0.04–6.74)	0.49
							SLC12A3	**2.02 (1.08–3.78)**	**0.004**
Collecting duct	*CALB1*	1.48 (0.79–2.78)	0.11				***CALB1***	**3.12 (1.15–8.47)**	**0.003**
	GATA2	0.53 (0.17–1.71)	0.16				*GATA2*	0.00	1.00
	RET	1.47 (0.93–2.35)	0.031				*RET*	1.99 (0.98–4.06)	0.013
Endothelium	*FLT1*	1.31 (0.92–1.87)	0.047	*KDR*	1.27 (0.86 –1.86)	0.11			
	KDR	1.19 (0.84–1.67)	0.20						

sessed more qualifying variants relative to the overall study population. To assess this hypothesis, the number of qualifying variants within each individual was categorized into 1 of 5 groups: no qualifying variants, 1 variant, 2 variants, 3 to 5 variants, and 6 or more variants. Chi-square analysis was performed comparing individuals with and without disease for each phenotype of interest to determine if the distribution within each cohort matched the distribution of the larger study population. Subgroup analyses by ancestry were performed using identical methods. Statistical significance was set at a critical α-level of 0.05.

Statistical Software

Statistical analysis was performed using SPSS version 28.0.0.0 (Statistical Package for the Social Sciences; SPSS Inc, Chicago, Illinois).

RESULTS

Genes Included in Analysis

Of the 83 genes initially identified, 58 were included in final analysis (57 in the non-White British subgroup) (Figure 1, Tables 3 and 4). A full listing of identified genes and rationale for exclusion is provided in Appendix 2. A numeric count of unique qualifying rare variants per gene, median number of individuals possessing each variant by gene, RUNES category, and functional impact is summarized in Appendix 4.

Association of Qualifying Variants and Phenotypes of Interest

Elevated Blood Pressure

When examining the cohort in its entirety, we found that qualifying variants in *SIX1* (OR 0.57; 99% CI, 0.35–0.94; $P = 0.004$) and *NPHS1* (OR 0.84; 99% CI, 0.72–0.99; $P = 0.005$) were protective against elevated blood pressure, whereas qualifying variants in *WT1* (OR 1.58; 99% CI, 1.02–2.45; $P = 0.007$) were predictive of elevated blood pressure. Among White British individuals, qualifying variants in *SIX1* were protective against blood pressure elevation (OR 0.58; 99% CI, 0.34–0.99; $P = 0.009$). In subgroup analysis of non-White British individuals, presence of qualifying variants in *WT1* in the early nephron development compartment was the only statistically significant predictor of elevated blood pressure (OR 2.49; 99% CI, 2.49–5.78; $P = 0.005$). Though statistical significance was not reached, qualifying variants in *CLCN5* in tubulointerstitial cell regression model approached significance in protection against elevated blood pressure (OR 0.67; 99% CI, 0.49–0.92; $P = 0.013$). Among genes significantly associated with elevated blood pressure, there was no difference in distribution of functional impact of variants between individuals with versus those without elevated blood pressure (Appendix 5A).

Odds ratios of all genes included in the regression models of the primary and subgroup analyses are summarized in Table 3.

Chronic Kidney Disease

In the analysis of the cohort in its entirety, presence of a qualifying variant in *CLCN5* as part of the logistic regression model for the tubulointerstitial cell compartment was the only significant predictor of CKD (OR 1.59; 99% CI, 1.02–2.47; $P = 0.007$).

There were no genes in which qualifying variants were significant predictors of CKD in White British subgroup analysis. In non-White British individuals, genes in which qualifying variants were significant predictors of CKD included *SLC12A3* (OR 2.02; 99% CI, 1.08–3.78; $P = 0.004$) for tubulointerstitial cells and *CALB1* in the collecting duct compartment (OR 3.12; 99% CI, 1.15–8.47; $P = 0.003$). Among genes significantly associated with CKD, there was no difference in distribution of functional impact of variants between individuals with versus those without CKD (Appendix 5B).

Odds ratios of all genes included in regression models of primary and subgroup analyses are summarized in Table 4.

Qualifying Variant Distribution

Elevated Blood Pressure

There were no significant differences in the proportion of qualifying variant counts among individuals with elevated blood pressure versus the full study population. This was true in the primary analysis, as well as in subgroup analyses (Figure 2A-C).

Chronic Kidney Disease

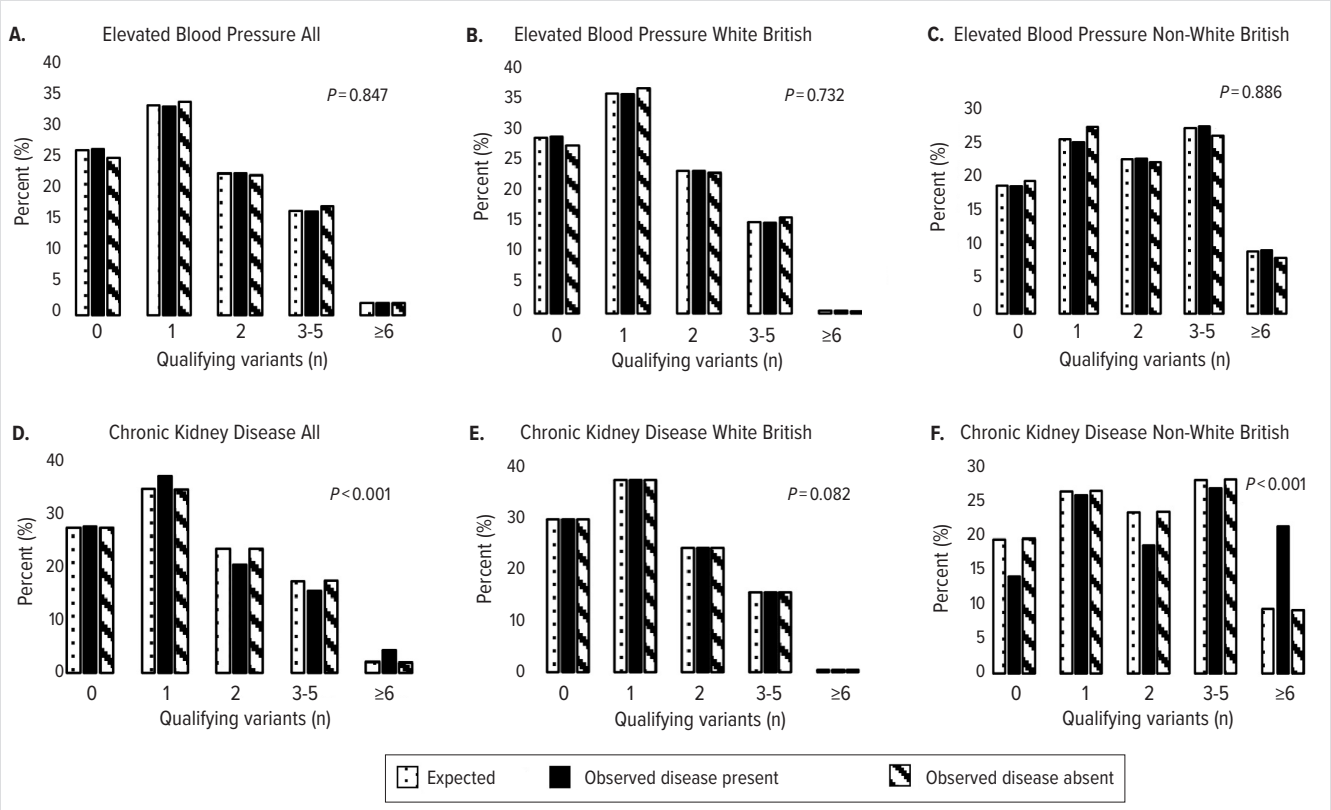
The proportion of variant counts within individuals with CKD differed significantly from proportions found in the overall study population ($P < 0.001$), with overall higher proportions of individuals with no qualifying variants or 6 or more qualifying variants among individuals with CKD. When examining participants of White British ancestry, we found that the distribution of variant counts within individuals with CKD did not significantly differ from proportions found in the overall White British subpopulation ($P = 0.082$). In individuals of non-White British ancestry, we found that a higher proportion of individuals with 6 or more qualifying variants was present in individuals with CKD versus the larger non-White British subpopulation ($P < 0.001$) (Figure 2D-F).

DISCUSSION

Our study highlights several important findings. We demonstrated that rare variants in kidney developmental genes are associated with hypertension and CKD, and that the implicated genes vary by ancestry. Further, while rare variants in some genes predict deleterious consequences, their presence in others confers a protective effect. Finally, individuals with CKD possess higher numbers of qualifying rare variants in kidney developmental genes than is expected relative to the larger population.

We found that rare variants in some genes implicated in structural kidney development are associated with development of elevated blood pressure and CKD later in life. This finding is in keeping with our hypothesis that rare variants in these genes contribute to nephron endowment and supports the idea that rare

Figure 2. Expected Versus Observed Qualifying Variant Counts in the Overall Cohort (A, D), Individuals Genetically Identified as White British (B,E), and Genetically Identified as Non-White British (C,F)



genetic variants—rather than exclusively common variants—contribute to the missing heritability in kidney function and disease. However, large proportions of the variance in blood pressure and CKD are not accounted for by our logistic regression models, which individually explain 6.9% to 15.3% of the variance in the case of blood pressure, and 1.1% to 9.2% of the variance in CKD. This unexplained variance suggests other factors (eg, variants in other genes, gene-gene, or gene-environment interactions) are at play, and evaluation of rare variants should be expanded to a wider selection of genes.

Interestingly, the specific genes in which rare variants predicted elevated blood pressure or CKD differed by ancestry. It is possible we are capturing the rare variants that predispose some populations to hypertension and CKD (eg, individuals of South Asian, sub-Saharan African, Aborigine, and Hispanic descent) when compared to White British individuals. Alternatively, such variation may arise from the heterogeneity of the non-White British subgroup, as this subgroup represents an admixture of many ancestral backgrounds. The characterization of certain variants as "rare" may be inaccurate, as they may be common variants within a more specific ancestral subgroup that is underrepresented in the UK Biobank. In this way, the number of qualifying variants within the non-White British subgroup may be overestimated.

Surprisingly, our analyses demonstrated that in some instances, rare qualifying variants have a protective rather than deleterious effect on blood pressure. This finding potentially stems from our definition of a qualifying variant. Because variants of uncertain significance (RUNES category 3) were included in our definition and can convey deleterious, neutral, or protective effects, some qualifying variants may be protective but have yet to be classified. To examine this effect further, we performed separate binary logistic regression analyses examining the ability of qualifying RUNES category 3 variants to predict elevated blood pressure and CKD, with results demonstrating associations in protective as well as deleterious directions (Appendix 6). Adding to this hypothesis is the large proportion of variants of uncertain significance that comprise qualifying variants in genes predictive of disease presence/absence, with those ranging from 64.1% to 100% of the qualifying variants in each predictive gene depending on disease state and ancestry (Appendix 7). An additional consideration is that qualifying variants in some kidney developmental genes may lead to impaired handling of sodium and water, as is seen in many tubulopathies, which could lead to hypotension. This consideration is supported by a prior study by Ji et al demonstrating that rare variants in the salt handling genes *SLC12A3*, *SLC12A1*, and *KCNJ1(ROMK)* are associated with reduced blood pressure.[23] Of note, rare variants in

SLC12A3 and *SLC12A1* were evaluated in our study but were not significantly associated with blood pressure.

Finally, we found that individuals with CKD possess higher numbers of rare variants in kidney developmental genes than would be expected in our entire cohort and in the non-White British subgroup. The fact that this association did not reach significance in the White British subgroup once again calls into question our characterizations of variants as "rare" among the non-White British subgroup. We attempted to account for this potential discrepancy by calculating expected qualifying variant counts for the whole study cohort and for each of the ancestral subgroups separately. Our finding that, among those with CKD, the proportion of individuals with high carriage of qualifying variants (ie, 6 or more variants) was ≥2 times the proportion found in the overall populations in the primary analysis and in non-White British subgroup analysis suggests there could be an additive effect relating the number of rare variants in kidney developmental genes possessed by an individual to CKD risk (Figure 2).

Limitations

Our study does have limitations. Relative to the general population of the United Kingdom, the UK Biobank tends toward a healthy subject bias with lower rates of all-cause mortality, cancer, smoking, alcohol use, and obesity, in addition to fewer self-reported health conditions.[24] This bias is seen in our analysis, with only 2.1% of the study population having CKD stage 3 and above compared to the estimated 10% of the general population.[1] Despite this bias, a large-scale meta-analysis by Batty et al comparing risk associations in the UK Biobank relative to risk associations in pooled data from 18 English and Scottish studies representative of the general population found that the associations in the UK Biobank were generalizable.[25] This finding supports our assertion that the significant associations in rare genetic variants and elevated blood pressure and CKD in the present study can be applied more generally.

However, caution is still needed when generalizing findings to the non-White British subgroup. As discussed previously, the admixed nature of the non-White British group and low representation of this group within the UK Biobank could lead to mischaracterization of genetic variants as rare due to ethnic disparities within the UK Biobank, thereby increasing risk of type 1 error. Alternatively, because the low number of non-White British individuals decreases the power of this subgroup analysis, the potential for a type 2 error rate is elevated as it pertains to associations in non-White British individuals. These possibilities highlight the need to enhance diversity and decrease disparities that exist within genomic research. The limitation of these relatively low sample sizes may be overcome to some extent with larger replicative studies, as the number of individuals possessing whole exome sequencing data within the UK Biobank has rapidly expanded to more than 450 000 participants since we accessed the biorepository.

A final limitation of our study is the use of whole exome sequencing and the restriction to genes implicated in kidney development. As we are increasingly appreciating, much genetic regulation takes place in intronic regions, which is missed through use of whole exome sequencing. The limited number of genes we examined limits our ability to examine gene-gene and gene-environment interactions. However, by limiting our analysis to those genes involved in nephrogenesis, we were able to support the hypothesis that rare variants in genes that could impact nephron endowment are predictive of elevated blood pressure and CKD, an inference that could not be made as easily in a broader genome-wide association study.

CONCLUSIONS

This study demonstrates that rare variants in kidney developmental genes can help predict the presence or absence of elevated blood pressure and chronic kidney disease; however, the implicated genes vary based on ancestry. These findings indicate that rare variants explain a portion of the risk of developing hypertension and CKD and serve as putative targets for disease risk screening in the future as the field of precision medicine continues to expand.

Financial Disclosures: None declared.

Funding/Support: This research was supported in part by the Children's Mercy Hospital Clinical Research Fellowship Award, The Sam and Helen Kaplan Research Fund in Pediatric Nephrology, and The McLaughlin Family Endowed Chair in Nephrology. None of these funding entities had a role in study design; collection, analysis, or interpretation of the data; writing the report; or the decision to submit the report for publication.

Acknowledgements: This research has been conducted using the UK Biobank Resource under Application Number 65332. This work uses data provided by patients and collected by the National Health Service (NHS) as part of their care and support. We thank the Medical Writing Center at Children's Mercy Kansas City for editing this manuscript.

Data Availability: The data generated from this study can be found within the published article and its appendices. Raw data of the UK Biobank are available from https://www.ukbiobank.ac.uk.

Appendices: Available at wmjonline.org.

REFERENCES

1. Hill NR, Fatoba ST, Oke JL, et al. Global prevalence of chronic kidney disease - a systematic review and meta-analysis. *PLoS One*. 2016;11(7):e0158765. doi:10.1371/journal.pone.0158765

2. Honeycutt AA, Segel JE, Zhuo X, Hoerger TJ, Imai K, Williams D. Medical costs of CKD in the Medicare population. *J Am Soc Nephrol*. 2013;24(9):1478-1483. doi:10.1681/ASN.2012040392

3. Collins AJ, Foley RN, Herzog C, et al. US Renal Data System 2010 annual data report. *Am J Kidney Dis*. 2011;57(1 Suppl 1):A8-e526. doi:10.1053/j.ajkd.2010.10.007

4. Brenner BM, Garcia DL, Anderson S. Glomeruli and blood pressure: less of one, more the other? *Am J Hypertens*. 1988;1(4 Pt 1):335-347. doi:10.1093/ajh/1.4.335

5. Wang X, Johnson AC, Williams JM, et al. Nephron deficiency and predisposition to renal injury in a novel one-kidney genetic model. *J Am Soc Nephrol*. 2015;26(7):1634-1646. doi:10.1681/ASN.2014040328

6. Cain JE, Di Giovanni V, Smeeton J, Rosenblum ND. Genetics of renal hypoplasia:

insights into the mechanisms controlling nephron endowment. *Pediatr Res*. 2010;68(2):91-98. doi:10.1203/PDR.0b013e3181e35a88

7. Fox CS, Yang Q, Cupples LA, et al. Genomewide linkage analysis to serum creatinine, GFR, and creatinine clearance in a community-based population: the Framingham Heart Study. *J Am Soc Nephrol.* 2004;15(9):2457-2461. doi:10.1097/01.ASN.0000135972.13396.6F

8. Gorski M, Tin A, Garnaas M, et al. Genome-wide association study of kidney function decline in individuals of European descent. *Kidney Int.* 2015;87(5):1017-1029. doi:10.1038/ki.2014.361

9. 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(2):157-165. doi:10.1159/000484327

10. Satko SG, Sedor JR, Iyengar SK, Freedman BI. Familial clustering of chronic kidney disease. *Semin Dial.* 2007;20(3):229-236. doi:10.1111/j.1525-139X.2007.00282.x

11. Salfati E, Morrison AC, Boerwinkle E, Chakravarti A. Direct estimates of the genomic contributions to blood pressure heritability within a population-based cohort (ARIC). *PLoS One*. 2015;10(7):e0133031. doi:10.1371/journal.pone.0133031

12. Wuttke M, Li Y, Li M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet.* 2019;51(6):957-972. doi:10.1038/s41588-019-0407-x

13. Gorski M, van der Most PJ, Teumer A, et al. 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function. *Sci Rep.* 2017;7:45040. doi:10.1038/srep45040

14. Chasman DI, Fuchsberger C, Pattaro C, et al. Integration of genome-wide association studies with biological knowledge identifies six novel genes related to kidney function. *Hum Mol Genet.* 2012;21(24):5329-5343. doi:10.1093/hmg/dds369

15. Inker LA, Eneanya ND, Coresh J, et al. New creatinine- and cystatin C-based equations to estimate GFR without race. *N Engl J Med.* 2021;385(19):1737-1749. doi:10.1056/NEJMoa2102953

16. Halldorsson BV, Eggertsson HP, Moore KHS, et al. The sequences of 150,119 genomes in the UK Biobank. *Nature*. 2022;607(7920):732-740. doi:10.1038/s41586-022-04965-x

17. Saunders CJ, Miller NA, Soden SE, et al. Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci Transl Med.* 2012;4(154):154ra135. doi:10.1126/scitranslmed.3004041

18. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17(1):122. doi:10.1186/s13059-016-0974-4

19. Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res*. Aug 1999;9(8):677-9. doi:10.1101/gr.9.8.677

20. Stenson PD, Ball EV, Howells K, Phillips AD, Mort M, Cooper DN. The Human Gene Mutation Database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. *Hum Genomics.* 2009;4(2):69-72. doi:10.1186/1479-7364-4-2-69

21. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016;44(D1):D733-D745. doi:10.1093/nar/gkv1189

22. Yates AD, Achuthan P, Akanni W, et al. Ensembl 2020. *Nucleic Acids Res.* 2020;48(D1):D682-D688. doi:10.1093/nar/gkz966

23. Ji W, Foo JN, O'Roak BJ, et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet.* 2008;40(5):592-599. doi:10.1038/ng.118

24. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol.* 2017;186(9):1026-1034. doi:10.1093/aje/kwx246

25. Batty GD, Gale CR, Kivimäki M, Deary IJ, Bell S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. *BMJ*. 2020;368:m131. doi:10.1136/bmj.m131

advancing the art & science of medicine in the midwest

WMJ

WMJ (ISSN 2379-3961) is published through a collaboration between The Medical College of Wisconsin and The University of Wisconsin School of Medicine and Public Health. The mission of *WMJ* is to provide an opportunity to publish original research, case reports, review articles, and essays about current medical and public health issues.

© 2025 Board of Regents of the University of Wisconsin System and The Medical College of Wisconsin, Inc.

Visit www.wmjonline.org to learn more.