The MYH9/APOL1 region and chronic kidney disease in European-Americans

Conall M. O’Seaghdha1,2, Rulan S. Parekh3,4, Shih-Jen Hwang1, Man Li4, Anna Köttgen4,9, Josef Coresh4,5,6, Qiong Yang7, Caroline S. Fox1,8,* and W.H. Linda Kao4,5,6,*,†

1National Heart, Lung and Blood Institute’s Framingham Heart Study and the Center for Population Studies, Framingham, MA, USA, 2Renal Division, Brigham and Women’s Hospital, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA, 3Division of Nephrology, Department of Pediatrics and Medicine, Hospital for Sick Children and University Health Network, University of Toronto, Toronto, Canada, 4Department of Epidemiology, Bloomberg School of Public Health, 5Department of Medicine, School of Medicine and 6Welch Center for Prevention, Epidemiology, and Clinical Research, School of Medicine and Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA, 7Department of Biostatistics, Boston University, Boston, MA, USA, 8Division of Endocrinology, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA, USA and 9Division of Nephrology, Department of Medicine, Freiburg University Hospital, Germany

Received January 14, 2011; Revised March 2, 2011; Accepted March 16, 2011

Polymorphisms in the MYH9 and adjacent APOL1 gene region demonstrate a strong association with non-diabetic kidney disease in African-Americans. However, it is not known to what extent these polymorphisms are present in other ethnic groups. To examine the association of genetic polymorphisms in this region with chronic kidney disease (CKD; estimated glomerular filtration rate < 60 ml/min/1.73 m²) in individuals of European ancestry, we examined rs4821480, an MYH9 single-nucleotide polymorphism (SNP) recently identified as associated with kidney disease in African-Americans, in 13,133 participants from the Framingham Heart Study (FHS) and Atherosclerosis Risk in Communities (ARIC) Study. In addition, we further interrogated the MYH9/APOL1 gene region using 282 SNPs for association with CKD using age-, sex- and center-adjusted models and performed a meta-analysis of the results from both studies. Because of prior data linking rs4821480 and kidney disease, we used a P-value of < 0.05 to test the association with CKD. In the meta-analysis, rs4821480 (minor allele frequency 4.45 and 3.96% in FHS and ARIC, respectively) was associated with higher CKD prevalence in participants free of diabetes (odds ratio 1.44; 95% confidence interval 1.15–1.80; P = 0.001). No other SNPs achieved significance after adjusting for multiple testing. Results utilizing directly genotyped data confirmed the results of the primary analysis. Recently identified APOL1 risk variants were also directly genotyped, but did not account for the observed MYH9 signal. These data suggest that the MYH9 polymorphism rs4821480 is associated with an increased risk of non-diabetic CKD in individuals of European ancestry.

INTRODUCTION

Chronic kidney disease (CKD) is an important worldwide health problem, affecting up to 13.1% of the adult population of the USA (1). Familial clustering of common forms of CKD, such as diabetic nephropathy, suggests that genetic factors play a key role in the pathogenesis (2). Recently, significant discoveries have advanced our understanding of the genetic determinants of several renal phenotypes, including estimated glomerular filtration rate (eGFR) (3), CKD (3), idiopathic focal segmental glomerulosclerosis (FSGS) (4) and end-stage renal disease (ESRD) (5). Notable among these has been the

*To whom correspondence should be addressed at: NHLBI’s Framingham Heart Study, 73 Mt Wayte Avenue, Suite #2, Framingham, MA 01702, USA. Tel: +1 5089353447; Fax: +1 5088722678; Email: foxca@nhlbi.nih.gov (C.S.F.); 615 N. Wolfe St., Room W6513, Baltimore, MD 21205, USA. Tel: +1 4106140945; Fax: +1 4109550863; Email: wkao@jhsph.edu (W.H.L.K.)
†Joint contribution

Published by Oxford University Press 2011.
identification of polymorphisms in the $MYH9/APOL1$ gene region, which has emerged as a kidney disease susceptibility locus of a strong and broad effect (4,5). In individuals of African descent, variants in this genomic region demonstrate a robust association with a variety of kidney disease phenotypes, including non-diabetic kidney disease (4), idiopathic focal segmental glomerulosclerosis, HIV-associated nephropathy, albuminuria (6) and ESRD (4). A recent work, examining large chromosomal regions adjacent to $MYH9$, has identified variants in the adjacent apolipoprotein L1 gene ($APOL1$) that are in strong linkage disequilibrium with previously identified $MYH9$ polymorphisms and which demonstrate even stronger statistical association with kidney disease phenotypes in African-Americans (7).

Almost all studies, to date, have focused on disease-based cohorts of African-Americans or populations known to have a substantial level of African ancestry, such as Hispanics (8–10). Consequently, it is not known whether polymorphisms in the $MYH9/APOL1$ gene region are associated with population-based CKD or whether they are associated with an increased disease risk in other ethnic groups, such as among individuals of European descent.

We examined whether the previously identified $MYH9$ disease-associated single-nucleotide polymorphism (SNP) and part of the E-1 haplotype (4,5), rs4821480, were associated with CKD in participants of European ancestry from the Framingham Heart Study (FHS) and the Atherosclerosis Risk in Communities (ARIC) Study. Given differences in linkage disequilibrium patterns and potential allelic heterogeneity across populations, we also interrogated the $MYH9$ gene region to determine whether other variants are independently associated with CKD in populations of European ancestry. Finally, in view of recent evidence of a statistically stronger association between variants in the $APOL1$ gene and kidney disease in African-Americans (7), we directly genotyped the three SNPs identified as renal risk variants in that analysis in the FHS and extended our interrogation to also include the adjacent $APOL1$ gene region.

RESULTS

Characteristics of the two study samples are shown in Table 1; 4140 participants with 445 CKD cases in FHS and 8993 participants with 791 CKD cases in ARIC contributed information to the present study. The CKD prevalence was 10.7% in the FHS and 8.8% in ARIC.

rs4821480 had a minor allele frequency of 4.45 and 3.96% in the FHS and ARIC, respectively (HapMap CEU frequency is 6.1%). The imputation score for rs4821480 was 0.99 in ARIC and 0.90 in the FHS (Supplementary Material, Table S1). Because of prior data linking rs4821480 and kidney disease (4,5), we used a $P$-value of $<0.05$ to test the association between this SNP and CKD. In a meta-analysis of FHS and ARIC results, rs4821480 in $MYH9$ was associated with CKD in the overall sample ($P = 0.003$; Table 2). Similar to previous reports, the association between rs4821480 and CKD was stronger in individuals without diabetes: the minor G allele of rs4821480 was associated with a 44% increased risk of CKD [odds ratio (OR) 1.44 per copy of the G allele; 95% confidence interval (CI) 1.15–1.80; $P = 0.001$]. We found no statistical evidence for heterogeneity between the diabetes and non-diabetes strata for rs4821480 ($P$-value for heterogeneity $= 0.74$ for rs4821480). Study-specific results can be found in Table 2.

We also directly genotyped rs4821480 in 3146 participants from the FHS and tested for association with CKD; results were similar to those using imputed data (OR per copy of the G allele 1.60; 95% CI 1.10–2.32; $P = 0.014$).

To address the question of whether additional independent alleles in $MYH9$ or $APOL1$ were also associated with CKD in these two populations, a total of 282 additional SNPs spanning the region from $MYH9$ to $APOL4$ were queried. Imputation scores for all SNPs in ARIC and FHS are presented in Supplemental Material, Table S1. We determined 30 of these SNPs to be independent using a pair-wise $r^2$ threshold of 0.2. After Bonferroni correction, the $P$-value for significance was 0.00167 (0.05/30); none of the SNPs in the meta-analysis reached this threshold. Results of the meta-analysis of the 282 SNPs in the $MYH9$ and $APOL1$ gene regions in participants without diabetes are presented in Figure 1.

Although no SNPs in the $APOL1$ gene region were associated with CKD, local genomic coverage in this region was poor. This was of particular concern in light of the recent identification of SNPs in the adjacent $APOL1$ gene that have a statistically stronger association with renal risk than $MYH9$ (7). To address this issue, we directly genotyped all three SNPs recently identified by Genovese et al. (rs73885319, rs60910145 and rs71785313) in the FHS ($n = 3107$ participants). We found that these SNPs occurred at extremely low frequency in the FHS (minor allele frequency 0.028–0.057%). Logistic regression revealed no significant association between these variants and CKD ($P$-values 0.14, 0.96 and 0.97 for SNPs rs73885319, rs60910145 and rs71785313, respectively). One carrier of the $MYH9$ G allele of rs4821480 was also a

<table>
<thead>
<tr>
<th>Table 1. Characteristics of the study samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort information</td>
</tr>
<tr>
<td>Study design</td>
</tr>
<tr>
<td>Analyzed sample size</td>
</tr>
<tr>
<td>Sample characteristics</td>
</tr>
<tr>
<td>Age (years)</td>
</tr>
<tr>
<td>Male (%)</td>
</tr>
<tr>
<td>Hypertension (%)</td>
</tr>
<tr>
<td>Diabetes mellitus (%)</td>
</tr>
<tr>
<td>CKD%, eGFR &lt;60 ml/min/1.73 m²</td>
</tr>
</tbody>
</table>

ARIC, Atherosclerosis Risk in Communities Study; FHS, Framingham Heart Study; CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate.

*Serum creatinine measurement and ascertainment of information on demographics, hypertension and diabetes occurred at the same visit.
carrier of one of each \textit{APOL1} risk variants at rs73885319 and rs60910145. Following exclusion of this participant, there was only minimal attenuation of the OR of CKD per copy of the G allele of rs4821480 (OR 1.55; 95% CI 1.06–2.28; \(P = 0.02\)). No additional participants carrying rare variants at each of the three \textit{APOL1} SNPs (\(n = 5\) in total) were also carriers of the \textit{MYH9} G allele at rs4821480.

\section*{DISCUSSION}

We have demonstrated that rs4821480 in the \textit{MYH9/APOL1-4} region is associated with an increased risk of CKD in individuals of European ancestry without diabetes. To date, polymorphisms in this region have been shown to increase the risk of ESRD almost exclusively in African-Americans (4,5). The present study extends knowledge in this area in several key ways. First, it demonstrates that the \textit{MYH9} risk allele, originally identified through admixture mapping in African-Americans, also confers an increased risk for non-diabetic kidney disease when present in individuals of European ancestry. It is important to note, however, that the minor allele frequency among individuals of European ancestry is much lower than that reported among African-Americans (4% when compared with 60%). Second, these findings are notable, in that they were made in a population-based setting as opposed to patient populations with clinically apparent disease. Third, this study describes a novel association between polymorphisms in the \textit{MYH9/APOL1} region and early CKD, whereas most studies have focused on advanced renal disease or ESRD as an outcome. Finally, we have excluded the possibility that the observed signal derives from recently identified \textit{APOL1} renal risk variants (7). This highlights the polymorphic nature of this portion of Chromosome 22, and we cannot rule out the existence of as yet undiscovered additional variants within this region that may account for the observed signal.

\textit{MYH9} is highly conserved among mammalian species, and knockout models die \textit{in utero} (11). It encodes for non-muscle myosin IIA, a mechanoenzyme present in the podocyte, a potential site of action (12), where it is believed to regulate actin dynamics and maintain normal podocyte structure (13).

\begin{table}
\centering
\begin{tabular}{llllll}
\hline
SNP & ARIC OR (95% CI) & \(P\)-value & FHS OR (95% CI) & \(P\)-value & Meta-analysis OR (95% CI) & \(P\)-value \\
\hline
rs4821480\textsuperscript{b} & & & & & & \\
Overall & 1.13 (0.88–1.46) & 0.34 & 1.89 (1.35–2.65) & 0.002 & 1.36 (1.11–1.67) & 0.003 \\
Diabetes & 0.60 (0.28–1.29) & 0.16 & 2.41 (1.03–5.61) & 0.04 & 1.11 (0.62–1.97) & 0.7 \\
No diabetes & 1.26 (0.96–1.66) & 0.10 & 1.84 (1.26–2.67) & 0.001 & 1.44 (1.15–1.80) & 0.001 \\
\hline
\end{tabular}
\caption{Association of rs4821480 in the \textit{MYH9} gene and CKD in participants of the FHS and the ARIC Study\textsuperscript{a}}
\end{table}

\textsuperscript{a}Results are presented as odds ratio for CKD, defined as eGFR < 60 ml/min/1.73 m\(^2\), 95\% CI in parentheses.

\textsuperscript{b}Presented per copy of the G allele; the test for homogeneity between the diabetes and non-diabetes strata in the meta-analysis was \(P = 0.74\).
Despite the identification of multiple disease-associated SNPs in intron 23 of MYH9 (4), and recent regional dense mapping localizing a kidney disease susceptibility hotspot to between introns 13 and 15 (14), the causal variant remains unknown. Recently, SNPs in the last exon of the adjacent APOL1 gene were identified as having a statistically stronger association with non-diabetic CKD than MYH9 in African-Americans (7). Although biologic evidence is thus far lacking, it is believed that these APOL1 variants are likely to be causal. No residual MYH9 signal remained after adjusting for the effects of both APOL1 SNPs in regression analyses, whereas adjusting for multiple MYH9 SNPs failed to attenuate the APOL1 signal. Furthermore, the APOL1 SNPs demonstrated a strong linkage disequilibrium with many MYH9 variants, and the E-1 haplotype in particular, leading the authors to conclude that APOL1 variations are the likely explanation for the MYH9 association. A subsequent analysis independently identified these variants and demonstrated them to be more strongly associated with ESRD than previously reported MYH9 variants from the same group (15).

We have shown that these APOL1 SNPs do not explain renal disease susceptibility in this chromosomal region in European-Americans from the FHS. In a rapidly advancing field, it remains plausible that further independent variants in APOL1, MYH9, or other genes may be identified. For example, separate analyses indicate additional, as-yet-undefined, MYH9 variants associated with FSGS susceptibility (16). Strong linkage disequilibrium patterns are evident in this region of Chromosome 22 (17–19), which appears to have arisen as a result of intense selective pressure within the past 10 000 years (20). This phenomenon appears to be due to a selective advantage conferred to APOL1 mutation carriers, mediated by trypanolytic activity of their serum not present in the wild-type, which confers resistance to African sleeping sickness (7).

Studies in other ethnic groups, such as Europeans with hypertension, and most notably American-Indians in the Strong Heart Study (21), have generally failed to replicate the association with MYH9 and kidney disease. However, an increased risk of non-diabetic ESRD and the E-1 haplotype was recently demonstrated among Hispanic-Americans (9). Furthermore, the authors identified the novel S1 MYH9 haplotype as being associated with an even greater risk of kidney disease. This identification of the S1 haplotype suggests that a possible reason for the lack of replication seen in other studies may be that the SNPs selected for analysis did not include ethnicity-appropriate polymorphisms. This highlights the importance of extending these findings to the full spectrum of ethnicities, and additional studies are needed to characterize the entire population at risk. The rs4821480 risk allele occurs more frequently among African-Americans, whereas protective alleles that form part of the E-2 and E-3 haplotypes are more frequent among individuals of European ancestry (4). This disparity has been proposed as a genetic explanation for the variation in risk for FSGS and hypertensive ESRD between these populations (22).

It is unlikely that unsuspected African ancestry explains our findings for several reasons. First, participants in the FHS are composed almost exclusively of whites of European descent, and this analysis was limited to only ARIC participants who are of European descent. Second, in both studies, individuals who did not cluster with the HapMap Caucasians from Utah based on genome-wide markers were excluded; furthermore, adjustment for principal components was performed in both studies and the results remained unchanged (23). Independent principal component analyses of comparable large European-American data sets consistently demonstrate that their population substructure has two axes of variation, tri-secting such populations into components roughly corresponding to northwest European, southeast European and Ashkenazi Jewish ancestry (24). Third, the minor allele frequency of rs4821480 in both the FHS and ARIC was similar to that in the HapMap CEU, a population believed to be representative of European populations (25). Finally, we have shown that APOL1 risk variants, which have been shown to account for the excess of renal risk among African-Americans, occur at extremely low frequency in our sample.

The well-characterized participants from both the FHS and ARIC Study with clear and validated measures of kidney disease are the strengths of this study. There are also some important limitations. First, participants in our analysis were middle-aged to elderly, and therefore, these findings cannot be generalized to younger individuals or other ethnicities. Second, although the magnitude of the ORs for the association between rs4821480 and diabetic and non-diabetic CKD appeared different, they were not statistically different, suggesting that a weaker association between rs4821480 and diabetic nephropathy may exist, which we were underpowered to detect. Third, we cannot rule out undiscovered additional rare variants within this region that may account for the observed signal. As such, further work in this area is warranted. Finally, our findings need to be replicated in large population-based and disease-specific studies, and further studies in other ethnic cohorts are also necessary.

In conclusion, the polymorphism rs4821480 in MYH9 is associated with an increased risk of non-diabetic CKD among individuals of European ancestry. This work broadens the spectrum of MYH9/APOL1-associated kidney disease beyond African-Americans and suggests that additional undiscovered risk variants may exist in this chromosomal region. Further studies are required to explore this possibility.

MATERIALS AND METHODS

Framingham Heart Study

The original cohort of the FHS began enrollment in 1948 (26). The Framingham offspring cohort began in 1971, enrolling 5124 participants; the design and methodology have been reported previously (27,28). Original cohort participants attending examination cycle 15 (1977–1979) or cycle 24 (1995–1998) (n = 2338) and offspring cohort participants attending the second examination cycle (1979–1983) or cycle 7 (1998–2001) (n = 4182) were included in the present study. Of these 6520 participants, complete genotype data were available for 4140. Participants in the FHS are composed primarily of whites of European descent (29).

The institutional review board of the Boston University Medical Center approved the study, and all participants provided written informed consent.
Atherosclerosis Risk in Communities Study

Initial recruitment of the 15 792 participants in ARIC occurred by probability sampling between 1987 and 1989. Individuals were mostly of European or African ancestry and aged 45–64 years. They underwent the baseline examination (visit 1) and three subsequent examinations approximately every 3 years (30). Samples from participants were excluded from genotyping if they did not consent to genetic studies (n = 53). For the present study, genotype data were available for 8993 white participants after data cleaning (3). Written informed consent was obtained, and the institutional review boards of each study center approved the study protocols.

Kidney function measurements and definition of CKD

In both studies, kidney function was measured using an eGFR from the simplified modification of diet in renal disease study equation (31). Serum creatinine was measured using a modified Jaffe method. Serum creatinine was calibrated to account for variations in quantification, as described previously (32).

In the FHS, a definition based on the cumulative prevalence of CKD was used to maximize power and reduce misclassification. Cases included participants with GFR ≤ 60 ml/min/1.73 m² at both the initial examination cycle (15th for original and 2nd for offspring cohort) and the later examination cycle (24th for original and 7th for offspring cohort) or diagnosed at the later examination cycles.

A cumulative CKD case definition was also used in ARIC; cases were defined based on eGFR < 60 ml/min/1.73 m² at study visits 1, 2 or 4, when creatinine was measured. Individuals with CKD at an earlier study visit who reverted to being non-cases at a later visit were only counted as cases if they also had an ICD code for kidney disease listed on a hospital discharge record or death certificate which was collected from study outset in 1987 to January 1, 2005 (33).

CKD covariate assessment

In both studies, participants underwent blood testing and were assessed for CKD risk factors. Type 2 diabetes was defined as fasting blood glucose ≥ 126 mg/dl (7.0 mmol/l) taken at the examination, or the use of insulin and/or oral hypoglycemic medication. In the ARIC Study, individuals who self-reported physician’s diagnosis of diabetes or had a non-fasting glucose ≥ 200 mg/dl were also considered to have diabetes. Hypertension was defined as either a systolic blood pressure (BP) ≥ 140 mmHg or a diastolic BP ≥ 90 mmHg, taken from the mean of two measurements, or the use of anti-hypertensive medication.

Genotyping and imputation

Briefly, both studies directly genotyped between 550 000 and 900 000 SNPs using Affymetrix whole-genome genotyping arrays as described previously (6.0, ARIC; 500K and 50K gene-centric, FHS) (3). All genotyping was carried out according to the manufacturer’s instructions between 2006 and 2008. Genotypes were imputed to a common set of ~2.5 million high-quality HapMap SNPs, using the Phase II CEU HapMap individuals as a reference panel. Software used for imputation was MACH v1.0.15/16 (http://www.sph.umich.edu/csg/abecasis/MACH/). The FHS accounted for relatedness of participants. Imputed genotypes were presented as an allelic dosage, expressed as a fractional value between 0 and 2.

The imputation score for rs4821480 was 0.99 in ARIC and 0.90 in the FHS. Imputation scores are a measure of data quality between 0 and 1, where scores close to 1 indicate high imputation confidence. Imputation scores for all SNPs in the ARIC and FHS are presented in Supplemental Material, Table S1. We defined the G allele of rs4821480 as the ‘minor allele’.

Direct genotyping of rs4821480 and three recently identified APOL1 risk variants (rs73885319, rs60910145 and rs71785313) was performed using a combination of an Illumina panel and Taqman in participants from the original and offspring cohorts (call rates 99–100%).

Statistical analyses

In the FHS, the first 10 principal components estimated from the genotype data using the Eigenstrat program (34) were not found to be significantly associated with prevalent CKD. In ARIC, the association of CKD with the first 10 principal components was tested and not found to be significant. Consequently, only genomic control adjustment was performed to account for population stratification. λ (35) for overall CKD was 1.014 and was 1.007 and 1.018 for the diabetic and non-diabetic strata, respectively.

In total, 283 SNPs (rs4821480 and 282 others) in the region from APOL4 to MYH9 were available in our imputed genome-wide association study data set, ranging from 34 915 188 to 35 111 626 bp on Chromosome 22 (build 36). Because of prior data linking rs4821480 and kidney disease, we used a P-value of < 0.05 to test the association between this SNP and CKD. Using a pair-wise $r^2 < 0.2$ to determine independence, we found 30 of the remaining 282 MYH9/APOL SNPs to be independent. After the Bonferroni correction, the P-value threshold for significance was 0.00167 (0.05/30) for the additional 282 SNPs in the MYH9/APOL gene region.

In both studies, logistic regression was used to assess whether SNPs were associated with CKD status, using an additive genetic model. In the FHS, logistic regression was implemented using generalized estimating equations (in the R package) to account for familial correlations. We also stratified our analyses by diabetes status. Data from the FHS and ARIC were meta-analyzed using METAL (http://www.sph.umich.edu/csg/abecasis/Metal/index.html). In both studies, all statistical analyses were performed using R.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.
ACKNOWLEDGEMENTS

ARIC: The authors thank the staff and participants of the ARIC study for their important contributions. FHS: This research was conducted in part using data and resources from the FHS of the National Heart, Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine.

Conflict of Interest statement. None declared.

FUNDING

Framingham Heart Study. This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart, Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine.

The analyses reflect the intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARE) project. This work was partially supported by the National Heart, Lung and Blood Institute’s Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc., for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGa-II) funded by the Robert David Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center.

Atherosclerosis Risk in Communities Study. The ARIC Study is carried out as a collaborative study supported by National Institutes of Health and Boston University School of Medicine.

The analyses reflect the intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARE) project. This work was partially supported by the National Heart, Lung and Blood Institute’s Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc., for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGa-II) funded by the Robert David Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center.

Atherosclerosis Risk in Communities Study. The ARIC Study is carried out as a collaborative study supported by National Institutes of Health and Boston University School of Medicine.

REFERENCES


