Published in final edited form as:

Hum Genet. 2016 November ; 135(11): 1251–1262. doi:10.1007/s00439-016-1714-2.

Association of kidney structure-related gene variants with type 2 diabetes-attributed end-stage kidney disease in African Americans

Meijian Guan¹, Jun Ma^{2,3}, Jacob M. Keaton^{1,4}, Latchezar Dimitrov¹, Poorva Mudgal¹, Mary Stromberg¹, Jason A. Bonomo^{1,4}, Pamela J. Hicks¹, Barry I. Freedman^{2,4}, Donald W. Bowden^{1,4,5}, and Maggie C. Y. Ng^{1,4}

Maggie C. Y. Ng: mng@wakehealth.edu

¹Center for Genomics and Personalized Medicine Research, Medical Center Boulevard, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA

²Section on Nephrology, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA

³Department of Nephrology, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

⁴Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA

⁵Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA

Abstract

African Americans (AAs) are at higher risk for developing end-stage kidney disease (ESKD) compared to European Americans. Genome-wide association studies have identified variants associated with diabetic and non-diabetic kidney diseases. Nephropathy loci, including *SLC7A9*, *UMOD*, and *SHROOM3*, have been implicated in the maintenance of normal glomerular and renal tubular structure and function. Herein, 47 genes important in podocyte, glomerular basement membrane, mesangial cell, mesangial matrix, renal tubular cell, and renal interstitium structure were examined for association with type 2 diabetes (T2D)-attributed ESKD in AAs. Single-variant association analysis was performed in the discovery stage, including 2041 T2D-ESKD cases and 1140 controls (non-diabetic, non-nephropathy). Discrimination analyses in 667 T2D cases-lacking nephropathy excluded T2D-associated SNPs. Nominal associations were tested in an additional 483 T2D-ESKD cases and 554 controls in the replication stage. Meta-analysis of 4218 discovery and replication samples revealed three significant associations with T2D-ESKD at *CD2AP* and *MMP2* ($P_{\text{corr}} < 0.05$ corrected for effective number of SNPs in each locus). Removal of *APOL1* renal-risk genotype carriers revealed additional association at five loci, *TTC21B*, *COL4A3*,

Correspondence to: Maggie C. Y. Ng, mng@wakehealth.edu.

M. Guan and J. Ma were equal contributors.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-016-1714-2) contains supplementary material, which is available to authorized users.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

NPHP3-ACAD11, *CLDN8*, and *ARHGAP24* ($P_{\text{corr}} < 0.05$). Genetic variants at *COL4A3*, *CLDN8*, and *ARHGAP24* were potentially pathogenic. Gene-based associations revealed suggestive significant aggregate effects of coding variants at four genes. Our findings suggest that genetic variation in kidney structure-related genes may contribute to T2D-attributed ESKD in the AA population.

Introduction

Relative to European Americans (EAs), African Americans (AAs) are at higher risk for developing end-stage kidney disease (ESKD). In 2012, the incidence rate of ESKD was 908 cases per million in AAs, nearly 3.3-fold higher than in EAs (USRDS 2014). The excess rate of ESKD in AAs is partly explained by socioeconomic factors (Byrne et al. 1994; Freedman 2002). Diabetes mellitus contributes to more than 44 % of all cases of ESKD in the US and approximately 90 % relate to type 2 diabetes (T2D) (USRDS 2014). Familial aggregation of diabetic kidney disease (DKD), including ESKD attributed to T2D (T2D-ESKD), has been reported (Sequist et al. 1989; Spray et al. 1995; Freedman et al. 1995; Quinn et al. 1996; Iyengar et al. 2015). AAs with T2D who have T2D-ESKD affected relatives have a nearly eightfold higher risk for developing T2D-ESKD compared with those who lack relatives with T2D-ESKD. The association remains after controlling for income, education, serum cholesterol, smoking, and hypertension (Freedman et al. 1995). In addition, prior genomewide association studies (GWAS) have identified variants associated with DKD in multiple populations (Maeda 2004; Pezzolesi et al. 2009; McDonough et al. 2011; Sandholm et al. 2012). These findings suggest that genetic factors may play a role in the development of nephropathy and ESKD in T2D patients. While the apolipoprotein L1 gene (*APOL1*) explains a substantial proportion of the disparity in non-diabetic ESKD in AAs (Tzur et al. 2010; Genovese et al. 2010), less is known about the genetic contributors to the high T2D-ESKD risk in AAs (Bonomo et al. 2014a; Ma et al. 2016; Skorecki and Wasser 2016).

Structural abnormalities in the kidney, including glomerular basement membrane (GBM) thickening, podocyte effacement, mesangial matrix expansion, glomerulosclerosis, and tubulointerstitial fibrosis, are observed during the development of kidney disease in patients with diabetes (Badal and Danesh 2014). A number of gene products play critical roles in maintaining normal kidney substructure and variants in several of these genes that have been associated with DKD. For example, three single nucleotide polymorphisms (SNPs) in the CD2-associated protein gene (*CD2AP*), essential for podocyte function, were associated with ESKD in Europeans with type 1 diabetes (T1D) (Hyvönen et al. 2013). Coding variants in the podocyte slit diaphragm gene nephrin (*NPHS1*) were associated with T2D-ESKD in AAs (Bonomo et al. 2014b). In addition, genetic variants at shroom family member 3 (*SHROOM3*), which plays a role in maintaining the integrity of the glomerular filtration barrier (Yeo et al. 2015), and uromodulin (*UMOD*), a kidney-specific gene expressed in the thick ascending limb (TAL) of the loop of Henle, were associated with kidney filtration in Europeans with T2D (Deshmukh et al. 2013).

These findings suggest that additional variants in glomerular and tubulointerstitial structure-related genes may be associated with ESKD in AAs with T2D. Herein, 47 candidate genes

involved in kidney structure were examined using a high coverage genotyping array, with a customized content of the respective targeted regions, to evaluate the impacts of common and low-frequency SNPs for association with T2D-ESKD in 4885 AA subjects.

Materials and methods

Study samples and clinical characteristics

The discovery stage included 2041 AA participants with T2D-ESKD and 1140 AA controls without diabetes or nephropathy. The replication stage included 483 AA participants with T2D-ESKD and 554 AA controls without diabetes or nephropathy. Trait discrimination analyses included T2D-ESKD cases and controls from the discovery stage, as well as 667 AAs with T2D who lacked evidence of nephropathy (Table 1). All subjects resided in North Carolina, South Carolina, Georgia, Tennessee, and Virginia. Patients were considered to have T2D-ESKD when diabetes was diagnosed after age 25 years and was present for 5 years prior to the onset of ESKD (or in the presence of diabetic retinopathy ensuring adequate T2D durations), and one or more of the following: renal replacement therapy ($N=1928$), estimated glomerular filtration rate (eGFR) <30 ml/min/1.73 m² ($N=43$), or urine albumin:creatinine ratio (UACR) >300 mg/g ($N=70$), when eGFR was missing. Patients with ESKD attributed to non-diabetic causes were excluded. Non-diabetic non-nephropathy controls included participants without diabetes or kidney disease (eGFR >60 ml/min/1.73 m² and UACR <30 mg/g). Individuals with T2D-lacking nephropathy had eGFR >60 ml/min/1.73 m² and UACR <30 mg/g. Detailed descriptions of inclusion criteria for T2D-ESKD patients, non-diabetic non-nephropathy controls, and T2D-lacking nephropathy subjects are included in the Supplementary Methods. This study was approved by the Wake Forest School of Medicine Institutional Review Board. All participants provided written informed consent.

Selection of kidney structure-related candidate genes

Although the precise mechanisms underlying kidney disease progression in patients with T2D are unclear, dysfunction of the glomerular filtration barrier, mesangial cell proliferation, mesangial matrix expansion, renal tubular atrophy, and renal interstitial fibrosis are observed. Therefore, genes related to the structure or function of podocytes, the slit diaphragm, GBM, mesangial cells, mesangial matrix, renal tubules, and renal interstitium were selected. Kidney structure-related genes were selected based on literature search and online database queries of GeneCards and Online Mendelian Inheritance in Man (OMIM) using keywords, including “glomerular”, “podocytes”, “mesangial matrix”, “glomerular basement membrane”, and “renal tubular”. A total of 47 genes reported to impact kidney structures were selected. These included *NPHS1*, *NPHS2*, *CD2AP*, and *ACTN4* in podocytes/slit diaphragms (Fan et al. 2006), type IV collagen, and *LAMB2* in GBM (Miner 2011), *ITGA3* and *MMP2* in mesangial matrix (Prete et al. 1997; Shukrun et al. 2014), and *SLC12A3*, *ATP6V1BA*, and the claudin gene-family in renal tubules (Abu Seman et al. 2014; Brown et al. 2009; Yu 2015) (full list in Supplementary Table 1).

Genotyping, imputation, and quality control

Discovery and trait discrimination samples—Samples from the discovery and trait discrimination stages (to discriminate association with nephropathy from association with T2D per se) were genotyped on a custom Affymetrix Axiom Biobank Genotyping Array (Affymetrix, Santa Clara, CA, USA). Detailed SNP information, custom content design, including fine mapping of candidate regions, genotyping methods, and quality control are described in the Supplementary Methods. The final data set consisted of 3848 subjects with relevant phenotypes for analysis. Relevant to this study, 4588 highly dense SNPs flanking 25-kb upstream and downstream of the 47 selected candidate genes which passed standard quality control (QC) were examined in the discovery stage.

Replication samples—483 AA T2D-ESKD cases and 554 AA controls (non-diabetic, non-nephropathy) were genotyped in the replication stage with the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA, USA). Standard quality controls (QCs) were applied to remove SNPs with call rate <95 %, minor allele frequency (MAF) <0.01, or showing departure from HWE ($P < 1 \times 10^{-4}$). Sample QC was performed to remove subjects with call rates <95 %, contamination, duplicates, or population outliers. SNPs that passed QC were used for pre-phasing and imputation to the cosmopolitan reference haplotype panel from the 1000 Genomes Project phase 1 version 3 (March 2012) (1000 Genomes Project Consortium et al. 2010) using SHAPEIT2 (Delaneau et al. 2012) and IMPUTE2 (Marchini et al. 2007). Imputed variants with imputation info score <0.4 were removed.

Statistical analysis

Single-variant association—Single-variant association in case-control samples was performed using linear mixed models (LMM) implemented in the program MMM (Pirinen et al. 2013) under an additive genetic model with adjustment for age and gender (Baseline model). This method incorporated a genetic relationship matrix (GRM) estimated from a set of high-quality autosomal SNPs as a random effect to control for population structure and cryptic relatedness. Whole genome analysis of discovery samples with 516,226 SNPs that passed QC yielded an inflation factor of 1.02, suggesting that population structure was sufficiently controlled by including GRM as a random effect. In the discovery stage, 4588 genetic variants from 47 candidate gene regions were tested in 2041 AA T2D-ESKD cases and 1140 AA non-diabetic, non-nephropathy controls. Nominally significant SNPs ($P < 0.05$) were further analyzed in the discrimination stages.

To distinguish whether discovery stage associations were driven by T2D alone or T2D-ESKD, 667 AAs with T2D-lacking nephropathy were included in two trait discrimination analyses. First, association was assessed in 2041 AAs with T2D-ESKD versus 667 AAs with T2D-lacking nephropathy. Differences in the effect size between T2D-ESKD versus non-diabetic controls (discovery stage) and T2D-ESKD versus T2D alone (discrimination stage) analyses were tested using Cochran's Q -test implemented in METAL (Willer et al. 2010). SNPs showing heterogeneity (Cochran's Q -test $P < 0.05$) and small effects ($P > 0.05$) in the T2D-ESKD versus T2D non-nephropathy analysis, suggesting lack of association with ESKD, were excluded. Second, the remaining SNPs were tested for association between the

667 AA T2D non-nephropathy participants versus the 1140 AA non-diabetic non-nephropathy controls. SNPs showing nominal association ($P < 0.05$) with T2D were excluded. These analyses removed SNPs that were solely associated with T2D. In secondary discrimination analyses, T2D samples without nephropathy that had T2D durations less than 5 years ($N = 204$) were removed to minimize misclassification.

SNPs remaining after the discovery and discrimination stages were further tested for replication in 483 AA T2D-ESKD cases and 554 AA controls-lacking diabetes and nephropathy. A meta-analysis was performed that included the discovery and replication cases and controls ($N = 4885$) using a fixed-effect inverse variance weighting method implemented in METAL (Willer et al. 2010).

Effective number of SNPs calculation—In each candidate region, the effective number of SNPs was calculated from the eigenvalues of the pair-wise SNP correlation matrix created with composite linkage disequilibrium (LD) for SNPs included in the discovery stage using simple M (Gao et al. 2008). Meta-analysis results were corrected for the effective number of SNPs in each region given their candidacy for associations. A corrected p value (P_{corr}) < 0.05 was considered significant.

Controlling for the effect of the apolipoprotein L1 gene (APOL1)—Two sets of variants in the *APOL1* gene are strong predictors of non-diabetic kidney disease in AAs. To account for the confounding effect of *APOL1*, the same single-variant analyses were repeated (*APOL1*-negative model) by removing 1019 *APOL1* renal-risk-variant carriers and those missing *APOL1* genotypes from T2D-ESKD, T2D non-nephropathy, and non-diabetic non-nephropathy groups, in both discovery and replication cohorts. Specifically, we removed 425 of 2041 T2D-ESKD cases, 293 of 1140 non-diabetic non-nephropathy controls, and 131 of 667 T2D non-nephropathy individuals from the discovery cohort. In addition, 100 of 483 T2D-ESKD cases and 70 of 554 non-diabetic non-nephropathy controls were removed from the replication cohort. Individuals were considered *APOL1* renal-risk-variant carriers if they possessed two G1 alleles (rs60910145 G allele, rs73885319 G allele), two G2 alleles (6 base pair in-frame deletion), or were compound heterozygotes (one G1 and one G2 allele) (Genovese et al. 2010). The power to detect odds ratio (OR) > 1.50 for SNPs with an effect allele frequency 0.05 at α level $= 1 \times 10^{-4}$ under an additive model fell from 68 to 50 % after removing *APOL1* risk allele carriers (<http://csg.sph.umich.edu/abecasis/cats/>).

Gene-based association—Sequence kernel association test (SKAT) (Wu et al. 2011) and Madsen–Browning test (MB) (Madsen and Browning 2009), implemented in RAREMETAL (Liu et al. 2014), were used to perform gene-based analyses with GRM included as a random effect. Only loss of function (LOF) and missense variants for each region were aggregated which was based on annotations from variant effect predictor (VEP) (McLaren et al. 2010). Candidate regions with only one variant were not tested. Age and gender were included as fixed-effect covariates. A p value < 0.001 , corrected for the number of genes tested, was considered statistically significant.

Bioinformatic characterization

Expression quantitative trait loci analysis—We identified proxies ($r^2 > 0.8$) for top associated variants using SNAP (<https://www.broadinstitute.org/mpg/snap/ldsearch.php>) based on African population and 1000 Genome Project data. Top variants and their LD proxies were then searched in a publically available eQTL database GTEx (<http://www.gtexportal.org/home/>).

Comprehensive functional annotation—Comprehensive variant annotation was performed using whole-genome sequencing annotator (WGSA) through its Amazon Machine Image. WGSA incorporates functional annotation from ANNOVAR (Wang et al. 2010), SnpEff (Cingolani et al. 2012), and VEP (McLaren et al. 2010) using contigs from both RefSeq (Pruitt et al. 2014) and Ensembl (Yates et al. 2016). In addition, it integrates five functional prediction scores, eight conservation scores, four disease-related annotations, and regulatory region-centric resources from multiple epigenomics projects.

Results

The impact of 4588 variants in 47 genes involved in kidney structure was investigated in 5 AA cohorts for associations with T2D-ESKD. Overall, 9 variants, 2 missense and 7 intronic, upstream, or downstream of candidate regions, located in 7 distinct regions reached locus-wide significance in either baseline model or *APOL1*-negative model.

Clinical characteristics of study participants

Characteristics of the samples in discovery, discrimination, and replication stages are shown in Table 1. Clinical characteristics of the individuals in the discovery stage were similar to those in the replication stage. The average age at recruitment was greater for individuals with T2D-ESKD and T2D-lacking nephropathy, compared with non-diabetic non-nephropathy controls ($P < 1 \times 10^{-27}$ in all stages). However, the average age of non-diabetic controls was greater than the average age at T2D diagnosis in T2D-ESKD cases ($P < 1 \times 10^{-9}$).

Association analysis in discovery stage

The impact of 4588 variants in 47 genes involved in kidney structure (Supplementary Table 1) was tested in 2041 T2D-ESKD cases and 1140 non-diabetic non-nephropathy controls for association with T2D-ESKD using a linear mixed model-based method (stage 1, Fig. 1) (Sawcer et al. 2011; Pirinen et al. 2013). Age and gender were included as fixed effects (Baseline model). These analyses did not adjust for PC1 given that the effect of ancestry is accounted for by GRM. A secondary analysis was performed which removed *APOL1* two-renal-risk-variant carriers, given their increased risk for non-diabetic kidney disease (*APOL1*-negative model). For single SNP analysis, only SNPs with MAF ≥ 0.01 were included. The number of SNPs in each gene region ranged from 11 to 205. The coverage for each of the 47 genes was calculated using SNPs with MAF ≥ 0.01 in 1000 Genome phase 1 version 3 AFR samples. The average coverage at $r^2 > 0.5$ (pair-wise LD) was 54.4 %, and varied from 27.6 to 92.8 %. Of 4588 SNPs, 212 showed nominal association ($P < 0.05$) in

baseline model from 39 candidate regions; these were further examined in the discrimination stages.

Discrimination analysis

Two analyses were performed to differentiate T2D-ESKD loci identified in the discovery stage as putative ESKD-associated or T2D-associated loci. An independent cohort of 667 AA individuals with T2D-lacking nephropathy was compared with T2D-ESKD cases and with non-diabetic non-nephropathy controls from the discovery stage in separate analyses (Table 1). First, we tested for association of the 212 associated SNPs between 2041 AA T2D-ESKD cases and 667 AAs with T2D-lacking nephropathy (stage 2a). Cochran's Q-test (Willer et al. 2010) was used to test for differences in allelic effects of association results between stage 1 and stage 2a, assuming that SNPs associated with T2D-ESKD have similar effects in both analyses. The smaller sample size in stage 2a reduced power to detect significant associations. Of the 212 SNPs, 169 did not demonstrate significant heterogeneity in effect sizes in stage 1 and stage 2a ($P_{\text{het}} > 0.05$). They were further tested for association with T2D by comparing 667 AA individuals with T2D-lacking nephropathy with 1140 AA non-diabetic non-nephropathy controls (stage 2b). Evidence of association with T2D was observed for three SNPs ($P < 0.05$); these were excluded from further analysis. The remaining 166 candidate SNPs from 35 distinct genomic regions were carried to the replication stage. Secondary discrimination analyses excluding T2D patients with <5 years disease duration yielded similar results (data not shown).

Replication and meta-analysis

An independent cohort of 483 AA T2D-ESKD cases and 554 AA non-diabetic non-nephropathy controls with imputed dosage data was used to replicate the 166 SNPs selected from stage 2b (stage 3). All SNPs met QC criteria in the replication cohort. The average imputation quality score (info) was 0.90, with the lowest info score of 0.45. Ninety SNPs from 30 candidate regions showed consistent directions of association with T2D-ESKD. Meta-analysis was performed to combine association results from stage 1 and stage 3 (labeled 'stage 4'). Three SNPs reached locus-wide significance after correcting for the effective number of SNPs in the respective regions ($P_{\text{corr}} < 0.05$) (Table 2; Supplementary Table 2). The two most significant locus-wide associations were intronic variants, rs116139597 (OR [95 % confidence interval] (OR [95 % CI]) = 1.28 [1.13–1.46], $P = 1.2 \times 10^{-4}$, $P_{\text{corr}} = 0.0056$) and rs115912771 (OR [95 % CI] = 0.63 [0.49–0.80], $P = 1.6 \times 10^{-4}$, $P_{\text{corr}} = 0.0074$), located in the *CD2AP* region, which overlap with a small portion of the adjacent gene *ADGRF2*. Rs115912771 is located in intron 11 of *CD2AP*, while rs116139597 is in intron 2 of *ADGRF2*, downstream from *CD2AP*. The third locus-wide association (rs7185763, OR [95 % CI] = 0.87 [0.81–0.95], $P = 9.5 \times 10^{-4}$, $P_{\text{corr}} = 0.036$) was a variant located 12-kb upstream from the *MMP2* gene. Of note, two additional associations approached locus-wide significance, rs6804354 (OR [95 % CI] = 1.19 [1.07–1.32], $P = 1.2 \times 10^{-3}$, $P_{\text{corr}} = 0.055$) located 18-kb upstream of the *NPHP3-ACAD11* read-through region and rs9927174 (OR [95 % CI] = 0.88 [0.81–0.94], $P = 5.5 \times 10^{-4}$, $P_{\text{corr}} = 0.060$) located 13-kb downstream from *SLC12A3* (Supplementary Table 3).

Association analyses excluding *APOL1* renal-risk carriers

The same series of single-variant analyses were performed under *APOL1*-negative model after removing *APOL1* renal-risk genotype carriers. A total of 70 SNPs showed evidence of association with T2D-ESKD ($P < 0.05$) and no evidence of association with T2D (stage 1 through stage 3). Meta-analysis of T2D-ESKD cases and non-nephropathy controls from the discovery and replication stages (stage 4) revealed six additional SNPs demonstrating locus-wide significance (Table 3; Supplementary Table 4). These included two missense variants, located at *CLDN8* (rs55884670, M97T, OR [95 % CI] = 1.35 [1.12–1.64], $P = 1.89 \times 10^{-3}$, $P_{\text{corr}} = 0.021$) and *COL4A3* (rs34505188, R408H, OR [95 % CI] = 1.55 [1.22–1.97], $P = 4.75 \times 10^{-4}$, $P_{\text{corr}} = 0.029$), respectively. Two variants in perfect LD ($r^2 = 1.0$), rs74504809 (OR [95 % CI] = 0.80 [0.70–0.91], $P = 8.71 \times 10^{-4}$, $P_{\text{corr}} = 0.040$), and rs78174962 (OR [95 % CI] = 0.79 [0.69–0.91], $P = 1.0 \times 10^{-3}$, $P_{\text{corr}} = 0.048$) from the *NPHP3-ACAD11* read-through region also demonstrated locus-wide significance, but are located at introns of *ACAD11*. After conditioning on rs78174962, rs74504809 did not demonstrate association ($P > 0.05$) suggesting that they comprise a single signal. In addition, a locus-wide significant signal (rs6742727, OR [95 % CI] = 0.81 [0.72–0.92], $P = 7.99 \times 10^{-4}$, $P_{\text{corr}} = 0.023$) was observed at an RNA gene *AC009495.3*, located upstream of *TTC21B*. Finally, an intronic variant at *ARHGAP24* (rs10433935, OR [95 % CI] = 0.62 [0.48–0.82], $P = 5.43 \times 10^{-4}$, $P_{\text{corr}} = 0.041$) demonstrated locus-wide significant association with T2D-ESKD. There were two additional associations that approach locus-wide significance, including rs72654165 (OR [95 % CI] = 2.06 [1.37–3.10], $P = 5.3 \times 10^{-4}$, $P_{\text{corr}} = 0.081$) located in the second intron of *COL4A1* and a missense variant rs57737815 (P555R, OR [95 % CI] = 0.70 [0.55–0.89], $P = 3.2 \times 10^{-3}$, $P_{\text{corr}} = 0.060$) located exon 7 of *WNK4* (Supplementary Table 3). All the three locus-wide significant SNPs from baseline model were no longer significant in *APOL1*-negative model ($P_{\text{corr}} > 0.05$); despite similar effect sizes in baseline model and *APOL1*-negative model. This suggests that the attenuation of association may be due to decreased sample size with loss of power (Supplementary Table 5).

Gene-based analysis

To improve statistical power, gene-based analysis was performed to aggregate the effects of functional variants within each candidate region using samples from the discovery and discrimination stages. Only missense and loss of function (LOF) variants were considered, the latter are predicted to cause splice site changes, frameshifts, or stop codon gains (McLaren et al. 2010). Initially, 2041 AA T2D-ESKD cases and 1140 AA non-diabetic non-nephropathy controls were examined in the discovery stage, followed by testing 2041 AA T2D-ESKD cases and 667 AAs with T2D-lacking nephropathy in the discrimination stage under baseline and *APOL1*-negative models. Analyses were performed for 39 genes with variant counts ≥ 2 . After excluding genes showing association with T2D alone, no gene reached defined significance level ($P < 0.001$, correction for 39 comparisons). However, nominal associations were observed at *EMP2* (variant count = 2, $P = 0.0065$), *COL4A3* (variant count = 12, $P = 0.0044$), *WNK4* (variant count = 8, $P = 0.0081$), and *CLDN14* (variant count = 2, $P = 0.0026$) under *APOL1*-negative model (Table 4).

Bioinformatic characterization

We applied expression quantitative trait loci (eQTL) analysis and functional annotation tool to uncover functional relevance underlying the identified associations. In the eQTL analysis, locus-wide significant SNPs from stage 4 and their proxies were queried in the publically available eQTL database GTEx. The intronic SNP rs10433935 in *ARH-GAP24* was strongly associated with expression level of *ARHGAP24* in artery-tibial tissue ($P = 0.000011$), indicating its regulatory role in this gene. In addition, rs6742727 and rs7185763 were associated with transcript abundance of nearby genes *GALNT3* and *AC00949.2* at the *TTC21B* region, and *IRX6* at the *MMP2* region, respectively (Supplementary Table 6).

We used the Amazon Compute Cloud version of WGAS (Liu et al. 2016) to annotate T2D-ESKD-associated SNPs. Three SNPs (rs55884670, M97T at *CLDN8*; rs34505188, R408H at *COL4A3*; and rs57737815, P555R at *WNK4*), were annotated as missense variants by VEP (McLaren et al. 2010), ANNOVAR (Wang et al. 2010) and SnpEff (Cingolani et al. 2012). In addition, rs34505188 and rs57737815 were predicted to be probably damaging by PolyPhen2 (Adzhubei et al. 2010). All three missense SNPs, along with rs10433935 from *ARHGAP24*, were located within highly conserved regions across mammals according to their high GERP++ rejected substitutions scores (Davydov et al. 2010). Two most recently developed functional prediction scores, CADD (Kircher et al. 2014) and DANN (Quang et al. 2015), were evaluated for the significantly associated SNPs. Both methods predict potential pathogenicity of genetic variants by integrating a wide range of annotations into a single metric using machine learning methods (Kircher et al. 2014; Quang et al. 2015). Again, rs34505188, rs57737815, and rs10433935 at *COL4A3*, *WNK4*, and *ARHGAP24* loci, respectively, were predicted within the top pathogenic variants (0.4–8 %) in the human genome. In addition, rs55884670 (*CLDN8*) was ranked within the top 16 and 28 % pathogenic variants in the human genome by CADD and DANN, respectively. For regulatory region-centric annotations, rs57737815 at *WNK4* overlapped with transcription factors-binding site (TFBS) of *POLR2A* in HepG2 cells, as well as DNase I hypersensitive site detected in two cell types. In addition, rs10433935 from *ARHGAP24* overlaps with a general developmental enhancer with <5 % false discovery rate (FDR) (Erwin et al. 2014). Full results are shown in Supplementary Table 7.

Discussion

The contribution of genetic variants in 47 kidney structure-related genes to T2D-ESKD was evaluated in 4885 AAs with high-density genotype data which included coding and non-coding variants. Evidence of association with T2D-ESKD was observed for seven loci, including *CD2AP*, *MMP2*, *TTC21B*, *COL4A3*, *NPHP3-ACAD11*, *CLDN8*, and *ARHGAP24* in single-variant analysis of either baseline or *APOL1*-negative models. The associated variants at *COL4A3* and *CLDN8* were missense mutations. All associated SNPs achieved locus-wide significance in the meta-analysis of T2D-ESKD cases and non-diabetic non-nephropathy controls, and showed no evidence of association with T2D. Gene-based and multiple bioinformatic analyses supported potential cumulative effect and functional relevance at associated loci.

The most significant association was observed in two SNPs located at *CD2AP* in baseline model analysis. CD2-associated protein (*CD2AP*) is an adapter protein between membrane proteins and the actin cytoskeleton (Kirsch et al. 1999). *CD2AP* is highly expressed in glomeruli, and has been associated with podocyte injury and focal segmental glomerulosclerosis (FSGS) (Shih et al. 1999; Löwik et al. 2007). Other variants in *CD2AP* demonstrated suggestive associations with ESKD in Europeans with T1D (Hyvönen et al. 2013). Together, this and other studies suggest that *CD2AP* variants are associated with both non-diabetic and diabetic (T1D and T2D) kidney diseases.

The SNP rs7185763 located 12-kb upstream to matrix metalloproteinase 2 (*MMP2*) also achieved locus-wide significance in baseline model analysis. Metalloproteinase 2 is a member of the MMP family and is involved in the degradation and turnover of extracellular matrix (ECM). Decreased expression of matrix metalloproteinase 2 in glomeruli was observed in patients with DKD, suggesting that lower expression level of *MMP2* is associated with mesangial matrix expansion and T2D-ESKD (Prete et al. 1997; Mason and Wahab 2003).

Analyses considering the impact of *APOL1* renal-risk genotypes revealed further insights into genetic susceptibility to T2D-ESKD. In *APOL1*-negative model analysis that removed *APOL1* risk genotype carriers, a low-frequency missense variant R408H (rs34505188) in type IV collagen (*COL4A3*) was associated with T2D-ESKD with the frequency of risk allele A being 0.033. R408H has a lower frequency in African populations (0.028), compared to European (0.069) and East Asian (0.16) based on the 1000 Genomes Project data. The alpha 3 chain in type IV collagen forms a heterotrimer with alpha 4 and alpha 5 chains, forming the major component of the GBM; it may be synthesized by podocytes (Miner 2011). Mutations in *COL4A3* are known to cause autosomal forms of Alport Syndrome (AS) and are associated with FSGS and chronic kidney disease in patients with thin basement membrane nephropathy (Kashtan 1995; Voskarides et al. 2007). Similar to the other five alpha chains in type IV collagen, alpha 3 contains a short amino-terminal domain, a long collagenous central domain consisting of Gly-X-Y repeats, and a non-collagenous domain. Replacement of a glycine in the Gly-X-Y repeat region is the most common mutant type of AS. These missense variants may induce kinks in folding with aberrant formation of the collagen triple helix (Prockop 1992). R408H is located in the Gly-X-Y repeat region and a prior study suggested that this coding variant was not pathogenic for AS (Heidet et al. 2001). However, R408H is located in a highly conserved region, and was predicted within the top 1 % of potentially pathogenic variants in the human genome by both CADD and DANN analyses. Furthermore, *COL4A3* trended toward association with T2D-ESKD in the gene-based analysis with the most significant association with a rare missense mutation rs55816283 [P1109S, $P = 0.00096$, effect allele frequency (EAF) = 0.0008]. Taken together, multiple lines of evidence support that *COL4A3* variants may be involved in ESKD susceptibility in AAs with T2D.

Another missense variant, M97T (rs55884670), in the claudin 8 gene (*CLDN8*) demonstrated locus-wide association with T2D-ESKD in AAs (risk G allele frequency = 0.052). This variant was not observed in non-African populations based on the 1000 Genomes Project data. Claudins are integral membrane components of tight junctions that

support pore and barrier functions of paracellular pathways in epithelial cells. They confer selective paracellular permeability in different segments of the renal tubule (Yu 2015). Claudin 8 is expressed in the thin descending limb of the loop of Henle, the distal tubule, and the collecting duct. It is required for the formation of the paracellular renal chloride channel by interacting with claudin 4 (Kiuchi-Saishin et al. 2002; Li et al. 2004; Hou et al. 2010). Other claudins have been associated with DKD (Gaut et al. 2014; Molina-Jijón et al. 2014); this is the first study to report association of a *CLDN8* gene variant with T2D-ESKD. Interestingly, another claudin family member *CLDN14* also demonstrated a trend toward association with T2D-ESKD in the gene-based analysis. Given the vital role of claudins in the renal tubule, variants in claudin genes may impact progression of kidney disease under the stress of hyperglycemia.

Two locus-wide significant SNPs (rs74504809, rs78174962) in complete LD located in the *NPHP3-ACAD11* region were associated with T2D-ESKD in *APOL1*-negative model. *NPHP3* is involved in ciliogenesis and possibly renal tubular development (Zhou et al. 2010). Mutations in *NPHP3* cause type 3 nephronophthisis, a relatively common cause of kidney failure in children and young adults (Olbrich et al. 2003). Both variants were located outside *NPHP3*, their potential regulatory role in *NPHP3* gene function remains to be determined.

The SNP rs10433935 located at *ARHGAP24* which encodes Rho GTPase-activating protein 24 was also associated with T2D-ESKD. Multiple lines of evidence, including eQTL analysis and WGSa annotation, suggest that rs10433935 overlaps with a highly functional region in *ARHGAP24*. Rho GTPase-activating protein 24 is involved in cell polarity, cell morphology, and cytoskeletal organization (Katoh and Katoh 2004). It plays a role in podocyte differentiation and contributes to the balance of podocyte Rho A and Rac1 signaling (Akilesh et al. 2011). Furthermore, a mutation in this gene has been associated with familial FSGS (Akilesh et al. 2011).

Collectively, these results suggest that variation at glomerular and tubulointerstitial-related structural genes may contribute to T2D-ESKD susceptibility in AAs. DKD is a complex disease with multiple environmental risk factors and polygenic inheritance. Microvascular changes, for example, abnormal capillary permeability and retinopathy, may precede early structural glomerular changes in patients with diabetes. Thus, variations in structural genes may be manifest specifically in the kidneys, despite possible altered expression in other organs (Alpert et al. 1972; Girach and Vignati 2006). It is noteworthy that mutations in *CD2AP*, *COL4A3*, and *ARHGAP24* are known to cause the podocyte disorder FSGS (Pollak 2014). The present findings suggest that structural kidney genetic variants may contribute to both DKD and FSGS. Alternatively, some T2D-ESKD cases might have been misclassified and had proteinuric FSGS or GBM diseases with coincident T2D. Most cases lacked a kidney biopsy; this limitation is present in virtually all large-scale genetic studies of T2D-ESKD. We carefully attempted to minimize effects of misclassification by removing patients with T2D who had other causes of kidney disease and by removing *APOL1* risk genotype carriers who likely had non-diabetic kidney diseases (*APOL1*-negative model). The latter should lead to enrichment for true DKD cases, and additional loci may be discovered as these samples were less heterogeneous despite the smaller sample size. A similar approach

led to the identification of *FRMD3* association with DKD in non-*APOL1* carriers (Freedman et al. 2011). Renal tubulointerstitial fibrosis and ECM accumulation have great impact on progression of DKD, as well as on other etiologies of nephropathy. Our results are consistent with variation in tubulointerstitium-related structural genes, such as *MMP2*, *NPHP3*, *TTC21B*, *CLDN8*, *CLDN14*, and *WNK4* contributing to interstitial injury and fibrosis in patients susceptible to DKD.

This study has limitations. As stated, it is difficult to fully exclude all cases misclassified as DKD due to the frequent lack of kidney biopsy material. However, removing patients with ESKD attributed to non-diabetic etiologies, as well as excluding *APOL1* renal-risk allele carriers, should minimize misclassification despite considerable reductions in power. An additional limitation is that many other kidney structure-related genes, for instance *SLC5A2*, which plays a role in tubuloglomerular feedback and hyperfiltration injury (Škrtić and Cherney 2015), could not be included due to the gene selection criteria. Higher density genetic data remain necessary to thoroughly examine these genes. Moreover, we lack prospective follow-up data to investigate genetic contributions to the progression of kidney disease in diabetic patients; this could be a powerful analysis to understand the dynamics of DKD. Finally, we did not adjust for several well-established DKD risk factors in our analyses, including blood pressures, degrees of albuminuria, body mass index, HbA_{1c}, or medication usage, because these change frequently over time as disease progresses.

Conclusion

This is the first study to comprehensively survey the genetic effects of kidney structure-related loci in T2D-ESKD susceptibility in the AA population. Using high-density genetic data and a multi-stage study design, strong genetic associations were identified between several kidney structure genes and T2D-ESKD risk. T2D-ESKD-associated variants at *COL4A3*, *CLDN8*, and *ARHGAP24* were potentially pathogenic. Future functional studies are warranted to illustrate the potential mechanisms underlying the effect of these newly identified genetic associations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH Grants R01 DK53591 (DWB), DK070941, and DK084149 (BIF), and by the National Natural Science Foundation of China (No. 81200488). This work has also been made possible through an International Society of Nephrology Fellowship and Shanghai Jiaotong University K.C. Wong Medical Fellowship Fund (Jun Ma). We acknowledge the contributions of the study participants, coordinators, physicians, staff, and laboratory.

References

- Abecasis GR, Altshuler D, et al. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467:1061–1073. [PubMed: 20981092]
- Abu Seman N, He B, Ojala JRM, et al. Genetic and biological effects of sodium-chloride cotransporter (SLC12A3) in diabetic nephropathy. *Am J Nephrol*. 2014; 40:408–416. [PubMed: 25401745]

- Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7:248–249. [PubMed: 20354512]
- Akilesh S, Suleiman H, Yu H, et al. Arhgap24 inactivates Rac1 in mouse podocytes, and a mutant form is associated with familial focal segmental glomerulosclerosis. *J Clin Invest*. 2011; 121:4127–4137. [PubMed: 21911940]
- Alpert JS, Coffman JD, Balodimos MC, et al. Capillary permeability and blood flow in skeletal muscle of patients with diabetes mellitus and genetic prediabetes. *N Engl J Med*. 1972; 286:454–460. [PubMed: 4550615]
- Badal SS, Danesh FR. New insights into molecular mechanisms of diabetic kidney disease. *Am J Kidney Dis*. 2014; 63:S63–S83. [PubMed: 24461730]
- Bonomo JA, Guan M, Ng MCY, et al. The ras responsive transcription factor RREB1 is a novel candidate gene for type 2 diabetes associated end-stage kidney disease. *Hum Mol Genet*. 2014a
- Bonomo JA, Ng MCY, Palmer ND, et al. Coding variants in nephrin (NPHS1) and susceptibility to nephropathy in African Americans. *Clin J Am Soc Nephrol CJASN*. 2014b; 9:1434–1440. [PubMed: 24948143]
- Brown D, Paunescu TG, Breton S, Marshansky V. Regulation of the V-ATPase in kidney epithelial cells: dual role in acid-base homeostasis and vesicle trafficking. *J Exp Biol*. 2009; 212:1762–1772. [PubMed: 19448085]
- Byrne C, Nedelman J, Luke RG. Race, socioeconomic status, and the development of end-stage renal disease. *Am J Kidney Dis Off J Natl Kidney Found*. 1994; 23:16–22.
- Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)*. 2012; 6:80–92. [PubMed: 22728672]
- Davydov EV, Goode DL, Sirota M, et al. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol*. 2010; 6:e1001025. [PubMed: 21152010]
- Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2012; 9:179–181.
- Deshmukh HA, Palmer CNA, Morris AD, Colhoun HM. Investigation of known estimated glomerular filtration rate loci in patients with type 2 diabetes. *Diabet Med J Br Diabet Assoc*. 2013; 30:1230–1235.
- Erwin GD, Oksenberg N, Truty RM, et al. Integrating diverse datasets improves developmental enhancer prediction. *PLoS Comput Biol*. 2014; 10:e1003677. [PubMed: 24967590]
- Fan Q, Xing Y, Ding J, et al. The relationship among nephrin, podocin, CD2AP, and alpha-actinin might not be a true “interaction” in podocyte. *Kidney Int*. 2006; 69:1207–1215. [PubMed: 16501493]
- Freedman BI. End-stage renal failure in African Americans: insights in kidney disease susceptibility. *Nephrol Dial Transpl*. 2002; 17:198–200.
- Freedman BI, Tuttle AB, Spray BJ. Familial predisposition to nephropathy in African-Americans with non-insulin-dependent diabetes mellitus. *Am J Kidney Dis*. 1995; 25:710–713. [PubMed: 7747724]
- Freedman BI, Langefeld CD, Lu L, et al. Differential effects of MYH9 and APOL1 risk variants on FRMD3 association with diabetic ESRD in African Americans. *PLoS Genet*. 2011
- Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet Epidemiol*. 2008; 32:361–369. [PubMed: 18271029]
- Gaut JP, Hoshi M, Jain S, Liapis H. Claudin-1 and Nephrin label cellular crescents in diabetic glomerulosclerosis. *Hum Pathol*. 2014; 45:628–635. [PubMed: 24529330]
- Genovese G, Friedman DJ, Ross MD, et al. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science*. 2010; 329:841–845. [PubMed: 20647424]
- Girach A, Vignati L. Diabetic microvascular complications—can the presence of one predict the development of another? *J Diabetes Complicat*. 2006; 20:228–237. [PubMed: 16798474]
- Heidet L, Arrondel C, Forestier L, et al. Structure of the human type IV collagen gene COL4A3 and mutations in autosomal Alport syndrome. *J Am Soc Nephrol JASN*. 2001; 12:97–106. [PubMed: 11134255]

- Hou J, Renigunta A, Yang J, Waldegger S. Claudin-4 forms paracellular chloride channel in the kidney and requires claudin-8 for tight junction localization. *Proc Natl Acad Sci USA*. 2010; 107:18010–18015. [PubMed: 20921420]
- Hyvönen ME, Ihalmio P, Sandholm N, et al. CD2AP is associated with end-stage renal disease in patients with type 1 diabetes. *Acta Diabetol*. 2013; 50:887–897. [PubMed: 23681557]
- Iyengar SK, Sedor JR, Freedman BI, et al. Genome-wide association and trans-ethnic meta-analysis for advanced diabetic kidney disease: Family Investigation of Nephropathy and Diabetes (FIND). *PLoS Genet*. 2015; 11:e1005352. [PubMed: 26305897]
- Kashtan CE. Clinical and molecular diagnosis of Alport syndrome. *Proc Assoc Am Phys*. 1995; 107:306–313. [PubMed: 8608415]
- Katoh M, Katoh M. Identification and characterization of ARHGAP24 and ARHGAP25 genes in silico. *Int J Mol Med*. 2004; 14:333–338. [PubMed: 15254788]
- Kircher M, Witten DM, Jain P, et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014; 46:310–315. [PubMed: 24487276]
- Kirsch KH, Georgescu MM, Ishimaru S, Hanafusa H. CMS: an adapter molecule involved in cytoskeletal rearrangements. *Proc Natl Acad Sci USA*. 1999; 96:6211–6216. [PubMed: 10339567]
- Kiuchi-Saishin Y, Gotoh S, Furuse M, et al. Differential expression patterns of claudins, tight junction membrane proteins, in mouse nephron segments. *J Am Soc Nephrol*. 2002; 13:875–886. [PubMed: 11912246]
- Li WY, Huey CL, Yu ASL. Expression of claudin-7 and -8 along the mouse nephron. *Am J Physiol Ren Physiol*. 2004; 286:F1063–F1071.
- Liu DJ, Peloso GM, Zhan X, et al. Meta-analysis of gene-level tests for rare variant association. *Nat Genet*. 2014; 46:200–204. [PubMed: 24336170]
- Liu X, White S, Peng B, et al. WGS: an annotation pipeline for human genome sequencing studies. *J Med Genet*. 2016; 53:111–112. [PubMed: 26395054]
- Löwik MM, Groenen PJTA, Pronk I, et al. Focal segmental glomerulosclerosis in a patient homozygous for a CD2AP mutation. *Kidney Int*. 2007; 72:1198–1203. [PubMed: 17713465]
- Ma J, Guan M, Bowden DW, et al. Association analysis of the cubilin (CUBN) and megalin (LRP2) genes with ESRD in African Americans. *Clin J Am Soc Nephrol*. 2016; 11:1034–1043. [PubMed: 27197912]
- Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet*. 2009; 5:e1000384. [PubMed: 19214210]
- Maeda S. Genome-wide search for susceptibility gene to diabetic nephropathy by gene-based SNP. *Diabetes Res Clin Pract*. 2004; 66:S45–S47. [PubMed: 15563979]
- Marchini J, Howie B, Myers S, et al. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007; 39:906–913. [PubMed: 17572673]
- Mason RM, Wahab NA. Extracellular matrix metabolism in diabetic nephropathy. *J Am Soc Nephrol*. 2003; 14:1358–1373. [PubMed: 12707406]
- McDonough CW, Palmer ND, Hicks PJ, et al. A genome wide association study for diabetic nephropathy genes in African Americans. *Kidney Int*. 2011; 79:563–572. [PubMed: 21150874]
- McLaren W, Pritchard B, Rios D, et al. Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. *Bioinformatics*. 2010; 26:2069–2070. [PubMed: 20562413]
- Miner JH. Glomerular basement membrane composition and the filtration barrier. *Pediatr Nephrol Berl Ger*. 2011; 26:1413–1417.
- Molina-Jijón E, Rodríguez-Muñoz R, Namorado M, del C, et al. Oxidative stress induces claudin-2 nitration in experimental type 1 diabetic nephropathy. *Free Radic Biol Med*. 2014; 72:162–175. [PubMed: 24726862]
- Olbrich H, Fliegau M, Hoefele J, et al. Mutations in a novel gene, NPHP3, cause adolescent nephronophthisis, tapeto-retinal degeneration and hepatic fibrosis. *Nat Genet*. 2003; 34:455–459. [PubMed: 12872122]

- Pezzolesi MG, Poznik GD, Mychaleckyj JC, et al. Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes. *Diabetes*. 2009; 58:1403–1410. [PubMed: 19252134]
- Pirinen M, Donnelly P, Spencer CCA. Efficient computation with a linear mixed model on large-scale data sets with applications to genetic studies. *Ann Appl Stat*. 2013; 7:369–390.
- Pollak MR. Familial FSGS. *Adv Chronic Kidney Dis*. 2014; 21:422–425. [PubMed: 25168831]
- Prete DD, Anglani F, Forino M, et al. Down-regulation of glomerular matrix metalloproteinase-2 gene in human NIDDM. *Diabetologia*. 1997; 40:1449–1454. [PubMed: 9447953]
- Prockop DJ. Mutations in collagen genes as a cause of connective-tissue diseases. *N Engl J Med*. 1992; 326:540–546. [PubMed: 1732793]
- Pruitt KD, Brown GR, Hiatt SM, et al. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res*. 2014; 42:D756–D763. [PubMed: 24259432]
- Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*. 2015; 31:761–763. [PubMed: 25338716]
- Quinn M, Angelico MC, Warram JH, Krolewski AS. Familial factors determine the development of diabetic nephropathy in patients with IDDM. *Diabetologia*. 1996; 39:940–945. [PubMed: 8858216]
- Sandholm N, Salem RM, McKnight AJ, et al. New susceptibility loci associated with kidney disease in type 1 diabetes. *PLoS Genet*. 2012
- Sawcer S, Hellenthal G, Pirinen M, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*. 2011; 476:214–219. [PubMed: 21833088]
- Sequist ER, Goetz FC, Rich S, Barbosa J. Familial clustering of diabetic kidney disease. *N Engl J Med*. 1989; 320:1161–1165. [PubMed: 2710189]
- Shih NY, Li J, Karpitskii V, et al. Congenital nephrotic syndrome in mice lacking CD2-associated protein. *Science*. 1999; 286:312–315. [PubMed: 10514378]
- Shukrun R, Vivante A, Pleniceanu O, et al. A human integrin- α 3 mutation confers major renal developmental defects. *PLoS One*. 2014; 9:e90879. [PubMed: 24621570]
- Skorecki K, Wasser WG. Beyond APOL1: genetic inroads into understanding population disparities in diabetic kidney disease. *Clin J Am Soc Nephrol*. 2016; 11:928–931. [PubMed: 27197913]
- Škrtić M, Cherney DZI. Sodium-glucose cotransporter-2 inhibition and the potential for renal protection in diabetic nephropathy. *Curr Opin Nephrol Hypertens*. 2015; 24:96–103. [PubMed: 25470017]
- Spray BJ, Atassi NG, Tuttle AB, Freedman BI. Familial risk, age at onset, and cause of end-stage renal disease in white Americans. *J Am Soc Nephrol JASN*. 1995; 5:1806–1810. [PubMed: 7787148]
- Tzur S, Rosset S, Shemer R, et al. Missense mutations in the APOL1. *Hum Genet*. 2010; 128:345–350. [PubMed: 20635188]
- United States Renal Data System. 2014 USRDS annual data report: epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda. 2014
- Voskarides K, Damianou L, Neocleous V, et al. COL4A3/ COL4A4 mutations producing focal segmental glomerulosclerosis and renal failure in thin basement membrane nephropathy. *J Am Soc Nephrol*. 2007; 18:3004–3016. [PubMed: 17942953]
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010; 38:e164–e164. [PubMed: 20601685]
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010; 26:2190–2191. [PubMed: 20616382]
- Wu MC, Lee S, Cai T, et al. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*. 2011; 89:82–93. [PubMed: 21737059]
- Yates A, Akanni W, Amode MR, et al. Ensembl 2016. *Nucleic Acids Res*. 2016; 44:D710–D716. [PubMed: 26687719]
- Yeo NC, O'Meara CC, Bonomo JA, et al. Shroom3 contributes to the maintenance of the glomerular filtration barrier integrity. *Genome Res*. 2015; 25:57–65. [PubMed: 25273069]
- Yu ASL. Claudins and the kidney. *J Am Soc Nephrol JASN*. 2015; 26:11–19. [PubMed: 24948743]

Zhou W, Dai J, Attanasio M, Hildebrandt F. Nephrocystin-3 is required for ciliary function in zebrafish embryos. *Am J Physiol Ren Physiol*. 2010; 299:F55–F62.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1.
Workflow of kidney structure-related gene analyses. *T2D* type 2 diabetes, *ESKD* end-stage kidney disease, *MAF* minor allele frequency

Table 1

Clinical characteristics of study cohorts

	Discovery stage		Discrimination stage		Replication stage	
	T2D-ESKD cases	Non-diabetic, non-nephropathy controls	T2D-lacking nephropathy	T2D-ESKD cases	Non-diabetic, non-nephropathy controls	
<i>N</i>	2041	1140	667	483	554	
Female (%)	57.2	51.7	64.5	62	55.4	
Age (years)	61.4 ± 10.8	46.5 ± 12.0	55.7 ± 11.6	64.5 ± 8.6	49.5 ± 11.1	
Age at T2D onset (years)	38.6 ± 12.7	–	46.2 ± 12.2	45.4 ± 9.8	–	
Duration of T2D for T2D-lacking nephropathy cohort	–	–	9.6 ± 9.1	–	–	
Duration of T2D prior to ESKD (years)	19.6 ± 9.9	–	–	16.1 ± 9.0	–	
Duration of ESKD (years)	3.6 ± 3.6	–	–	3.2 ± 3.7	–	
Fasting serum glucose (mg/dl)	–	96.5 ± 20.9	–	–	85.1 ± 10.0	
Serum creatinine (mg/dl)	–	0.97 ± 0.2	0.94 ± 0.2	–	0.95 ± 0.2	
Body mass index (kg/m ²)	30.8 ± 7.1	29.7 ± 7.4	33.1 ± 7.8	29.3 ± 6.5	28.4 ± 4.7	

Categorical data expressed as percentage; continuous data as mean ± SD

T2D type 2 diabetes, ESKD end-stage kidney disease, *N* number

Table 2
T2D-ESKD-associated SNPs in meta-analysis from discovery and replication cohorts (Baseline model)

Candidate region	SNP	Annotation	Chr	Position	EA	Discovery (T2D-ESKD cases vs. non-diabetic, non-nephropathy controls)			Replication (T2D-ESKD cases vs. non-diabetic, non-nephropathy controls)			Meta-analysis		
						N case/control	EAF case/control	OR	P	N case/control	EAF	OR	P	P _{corr}
CD2AP	rs115912771	Intronic	6	47,551,182	A	2040/1138	0.019/0.032	0.64	0.0010	483/554	0.027	0.60	0.060	0.0074
CEBP	rs116139597	Downstream	6	47,633,775	C	2038/1136	0.11/0.080	1.28	0.0013	483/554	0.096	1.30	0.034	0.0056
MIR22	rs7185763	Upstream	16	55,411,450	C	2034/1136	0.28/0.33	0.87	0.0042	483/554	0.30	0.88	0.098	0.036

Baseline model: adjusted for age and sex, *APOL1* risk genotype carriers included
SNPs: single nucleotide polymorphism, *T2D* type 2 diabetes, *ESKD* end-stage kidney disease, *Chr* chromosome, *EA* effect allele, *N* number, *EAF* effect allele frequency, *OR* odds ratio, *P* value, *CI* confidence interval, *P_{corr}* value adjusted for multiple comparison

Table 3

Additional associations in combined analysis after removing *APOL1* renal-risk genotype carriers (*APOL1*-negative model)

Candidate region	SNP	Annotation	Chr	Position	EA	Discovery (T2D-ESKD cases vs. non-diabetic, non-nephropathy controls)			Replication (T2D-ESKD vs. non-diabetic, non-nephropathy controls)			Meta-analysis		
						N case/control	EAF case/control	OR	P	N case/control	EAF	OR	P	P _{corr}
<i>TTC21B</i>	rs6742727	Intronic	2	166,692,686	T	1583/828	0.11/0.15	0.76	0.00025	383/484	0.13	0.94	0.62	0.023
<i>COL4A3</i>	rs34505188	Missense	2	228,128,568	A	1613/846	0.035/0.024	1.57	0.0035	383/484	0.037	1.50	0.054	0.029
<i>NPHP3-ACAD11</i>	rs78174962	Intronic	3	132,306,663	A	1612/847	0.086/0.11	0.80	0.0058	383/484	0.094	0.78	0.075	0.048
<i>NPHP3-ACAD11</i>	rs74504809	Intronic	3	132,366,578	G	1613/846	0.099/0.12	0.81	0.0084	383/484	0.10	0.75	0.035	0.040
<i>ARHGAP24</i>	rs10433935	Intronic	4	86,597,245	C	1608/843	0.020/0.029	0.63	0.003	383/484	0.031	0.62	0.077	0.041
<i>CLDN8</i>	rs55884670	Missense	21	31,587,954	G	1612/845	0.060/0.043	1.29	0.024	383/484	0.048	1.56	0.022	0.021

APOL1-negative model: adjusted for age and sex, *APOL1* risk carriers removed

SNP single nucleotide polymorphism, *T2D* type 2 diabetes, *ESKD* end-stage kidney disease, *Chr* chromosome, *EA* effect allele, *N* number, *EAF* effect allele frequency, *OR* odds ratio, *P_p* value, *CI* confidence interval, *P_{corr}* value adjusted for multiple comparison

Table 4

Top associations from the gene-based analysis in *APOL1*-negative model

Gene	Method	T2D-ESKD vs. non-diabetic, non-nephropathy controls (<i>N</i> = 2463)					T2D-lacking nephropathy vs. non-T2D (<i>N</i> = 1383)	
		Number of SNPs	<i>P</i>	Best SNP	<i>P</i> (best SNP)	EAF (best SNP)	Number of SNPs	<i>P</i>
<i>EMP2</i>	SKAT	2	0.0065	rs73503834	0.0064	0.0333	2	0.93
<i>COL4A3</i>	SKAT	12	0.0044	rs55816283	0.00096	0.000813	12	0.56
<i>WNK4</i>	SKAT	8	0.0081	rs57737815	0.005	0.0279	8	0.64
<i>CLDN14</i>	Burden	2	0.0026	rs140918123	0.032	0.00406	2	0.78

APOL1-negative model: adjusted for age and sex, *APOL1* risk genotype carriers excluded
T2D type 2 diabetes, *ESKD* end-stage kidney disease, *P* *p* value, *EAF* effect allele frequency, *SKAT* sequence kernel association test, *Burden* Madsen–Browning test, *Best SNP* the most significant SNP within each region