Published in final edited form as:

Hum Genet. 2010 August ; 128(2): 195–204. doi:10.1007/s00439-010-0842-3.

Candidate genes for non-diabetic ESRD in African Americans: a genome-wide association study using pooled DNA

Meredith A. Bostrom,

Department of Biochemistry, Wake Forest University, School of Medicine, Winston-Salem, NC 27157-1053, USA. Center for Human Genomics, Wake Forest University, School of Medicine, Winston-Salem, NC 27157-1053, USA. Center for Diabetes Research, Wake Forest University, School of Medicine, Winston-Salem, NC 27157-1053, USA

Lingyi Lu,

Department of Biostatistical Sciences, Wake Forest University, School of Medicine, Winston-Salem, NC 27157-1053, USA

Jeff Chou,

Department of Biostatistical Sciences, Wake Forest University, School of Medicine, Winston-Salem, NC 27157-1053, USA

Pamela J. Hicks,

Department of Biochemistry, Wake Forest University, School of Medicine, Winston-Salem, NC 27157-1053, USA

Jianzhao Xu,

Center for Diabetes Research, Wake Forest University, School of Medicine, Winston-Salem, NC 27157-1053, USA

Carl D. Langefeld,

Department of Biostatistical Sciences, Wake Forest University, School of Medicine, Winston-Salem, NC 27157-1053, USA

Donald W. Bowden, and

Department of Biochemistry, Wake Forest University, School of Medicine, Winston-Salem, NC 27157-1053, USA. Center for Human Genomics, Wake Forest University, School of Medicine, Winston-Salem, NC 27157-1053, USA. Center for Diabetes Research, Wake Forest University, School of Medicine, Winston-Salem, NC 27157-1053, USA

Barry I. Freedman

Section on Nephrology, Department of Internal Medicine/Nephrology, Wake Forest University, School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157-1053, USA

Barry I. Freedman: bfreedma@wfubmc.edu

Abstract

African Americans have increased susceptibility to non-diabetic (non-DM) forms of end-stage renal disease (ESRD) and extensive evidence supports a genetic contribution. A genome-wide association study (GWAS) using pooled DNA was performed in 1,000 African Americans to detect associated genes. DNA from 500 non-DM ESRD cases and 500 non-nephropathy controls

was quantified using gel electrophoresis and spectrophotometric analysis and pools of 50 case and 50 control DNA samples were created. DNA pools were genotyped in duplicate on the Illumina HumanHap550-Duo BeadChip. Normalization methods were developed and applied to array intensity values to reduce inter-array variance. Allele frequencies were calculated from normalized channel intensities and compared between case and control pools. Three SNPs had p values of $<1.0E-6$: rs4462445 (ch 13), rs4821469 (ch 22) and rs8077346 (ch 17). After normalization, top scoring SNPs ($n = 65$) were genotyped individually in 464 of the original cases and 478 of the controls, with replication in 336 non-DM ESRD cases and 363 non-nephropathy controls. Sixteen SNPs were associated with non-DM ESRD ($p < 7.7E-4$, Bonferroni corrected). Twelve of these SNPs are in or near the *MYH9* gene. The four non-*MYH9* SNPs that were associated with non-DM ESRD in the pooled samples were not associated in the replication set. Five SNPs that were modestly associated in the pooled samples were more strongly associated in the replication and/or combined samples. This is the first GWAS for non-DM ESRD in African Americans using pooled DNA. We demonstrate strong association between non-DM ESRD in African Americans with *MYH9*, and have identified additional candidate loci.

Introduction

African Americans have a fourfold higher incidence rate of end-stage renal disease (ESRD) relative to European Americans. This racial disparity is not fully accounted for by differences in socioeconomic status, access to medical care, or control of hypertension. Members of selected African American families are at markedly increased risk for developing ESRD (Freedman et al. 2004). While diabetes is the primary cause of ESRD in nearly 50% of incident cases in African Americans, non-diabetic (non-DM) forms of ESRD, commonly labeled as hypertensive ESRD and chronic glomerulonephritis-associated ESRD, accounted for 40% of ESRD cases in 2008 (U.S. Renal Data System 2008). Intensive hypertension control does not prevent hypertensive ESRD in African Americans (Appel et al. 2008). These observations suggest that an inherited susceptibility to nephropathy is present.

Despite this high incidence rate, only recently have studies begun to shed light on the genetic basis of susceptibility to non-DM ESRD in African Americans. We recruited a large collection of African American pedigrees containing members with non-DM ESRD (hypertension-and chronic glomerulonephritis-associated) and a corresponding large collection of unrelated African American controls lacking nephropathy. Initial linkage studies in African American sib pairs identified several regions linked with non-DM forms of ESRD (Freedman et al. 2004, 2005). Follow-up candidate gene studies evaluated gene variants which may contribute to susceptibility to non-DM forms of ESRD (podocin (Dusel et al. 2005), ephrin-B2 (Hicks et al. 2008), *klotho* (in press)). Most of these variants were relatively rare and did not contribute significantly to risk in the overall population. More recently, MALD analysis (mapping by admixture linkage disequilibrium) identified variants in the *MYH9* gene which are common in the African American population and contribute significantly to risk of non-DM ESRD (Kao et al. 2008; Kopp et al. 2008; Freedman et al. 2009b). Despite the significant impact of *MYH9* on non-DM ESRD, not all subjects with *MYH9* risk alleles progress to ESRD, indeed, approximately 4% of risk allele carriers progress to ESRD, whereas 57% of non-nephropathy controls possess at least one copy of the risk haplotype. This suggests that factors in addition to *MYH9*, environmental and/or genetic, contribute to the increased risk for ESRD in this population.

To identify candidate genes associated with non-DM ESRD, we performed a genome-wide association study (GWAS) using 500 African American non-DM ESRD cases and 500 non-nephropathy controls. While a great deal of information can be derived from GWAS, the

cost and labor involved in large scale genotyping is substantial. To reduce the cost and time investment, DNA samples can be combined into pools and loaded on arrays, thereby reducing the total number of arrays necessary. Allele frequencies are generated for each pool and compared between cases and controls. DNA pooling for genome-wide association has previously been successful in the identification of disease-associated loci in type 2 diabetic ESRD (Hanson et al. 2007), rheumatoid arthritis (Steer et al. 2007), and schizophrenia (Kirov et al. 2009). In a comparison between DNA pools genotyped on the Illumina HumanHap300 array and individual genotyping, Macgregor et al. (2008) demonstrated that >80% of the information derived from individual genotyping could be obtained using pooled DNA. The studies of rheumatoid arthritis and schizophrenia also identified known candidate genes in the top hits from the GWAS using pooled DNA, demonstrating the value of this technique in the identification of candidate genes associated with complex disease.

We used pooled DNA samples in a GWAS to identify candidate genes associated with non-DM ESRD in African Americans. Allele frequencies estimated from pooled control samples were compared to frequencies from individual genotyping in order to verify the accuracy of pool construction. We demonstrate that DNA pooling can be a cost effective way to identify candidate genes associated with non-DM ESRD.

Methods

Subjects

Cases were unrelated, self-described African Americans with ESRD receiving renal replacement therapy. Cases were recruited from dialysis and transplant facilities in North Carolina, South Carolina, Virginia and Tennessee. These individuals were categorized as having non-DM ESRD if they had no indication of diabetes in their medical history, either through self-reporting or through review of medical records. Controls were self-reported African Americans >18 years of age who were recruited from community sources and medical clinics in North Carolina. Controls denied a history of kidney disease. Each participant provided 40 ml of blood for DNA extraction. All protocols were approved by the Wake Forest University Institutional Review Board. Percentage of African ancestry was calculated for each individual from genotyping 70 ancestry informative markers (AIMs), as previously published (Keene et al. 2008). Age, BMI, and percentage of African ancestry were compared between cases and controls using a Mann–Whitney Rank Sum Test (SigmaStat 3.5, Systat Software, San Jose, CA). Proportion of females between cases and controls were compared using a z test (SigmaStat 3.5).

DNA extraction and quantification

DNA was extracted from whole blood using the AutoPure LS automated DNA extraction robot (Gentra Systems, Minneapolis, MN). DNA was quantified using a Nanodrop Spectrophotometer (Thermo Scientific, Wilmington, DE). Aliquots were also run individually on agarose gels and compared to a calf thymus standard to confirm quantitation and sample quality. Samples with degraded DNA were excluded.

DNA pooling

DNA from 500 non-DM ESRD case samples and 500 controls were selected for the DNA pools. DNA samples were normalized to 50 ng/μl each. Ten case and control pools were constructed containing 50 samples each, with a final concentration of 50 ng/μl. Each pool was genotyped in duplicate on the Illumina HumanHap550-Duo BeadChip (Illumina, San Diego, CA) using the Illumina Infinium II Assay (<http://www.illumina.com>). BeadChips were imaged on the Illumina BeadStation 500 GX at the Center for Human Genomics at Wake Forest University. Individual red and green channel intensities for each BeadChip

were extracted from the BeadStation. With the BeadChip technology, individual markers are hybridized to several beads, this redundancy leading to increased genotyping success and efficiency. As a result, there are anywhere from 1 to 54 reads per SNP on each chip. The average number of reads per SNP was 15. Individual SNPs with fewer than five reads per chip were removed from further analysis. For each BeadChip, red and green intensity values (reads) for each SNP were averaged, resulting in a single pair of intensity values for each SNP.

Normalization

The fluorescent intensity data were normalized using two methods. The first method, denoted “channel mean” method in Fig. 1, follows the approach by Macgregor et al. (2008) to account for the bias in the green channel intensity. For each array, a multiplier is computed that consists of the ratio of the means of the red channel and the green channel. Fluorescent intensities are then adjusted by the multiplier resulting in the mean green channel intensity being equal to the mean red channel intensity for each array. However, this approach does not adjust for differences in the variance of the two channels and assumes that the relationship between red and green channels can be approximated by a linear function. In the second method, denoted as the “log array mean” method adapted from systematic variation normalization (Chou et al. 2005), the minimum intensity value, considered as background, for each channel (i.e., red, green) and array was subtracted from each SNPs intensity value; here, the minimum ignores those outliers in the raw intensity distribution. The resulting distribution was base 2 logarithm transformed and each channel and array was scaled (multiplied) to the global average in the combined case/control sample. Next, two reference data sets for both the red and green channels were generated by averaging the processed red and green control data across all arrays, respectively. A third degree polynomial regression was applied to the data from each channel against its corresponding (red or green) reference data. The regression line was used only to adjust the channel data. The reference data were fixed and the regression residual was preserved. After the adjustment, the new regression line was a straight line and had an intercept at zero and a slope of one.

Statistical analysis of pooled GWAS

Relative allele frequencies (RAF) were calculated from normalized fluorescent intensities using the following formula: $RAF = (\text{red intensity}) / (\text{red intensity} + \text{green intensity})$. Allele frequencies in cases and controls were analyzed by unpaired two tailed Student's *t* tests under both equal variance and unequal variance assumptions. To control for multiple testing in GWAS, $p < 0.00001$ was taken as the lowest limit of significance to eliminate false positive significant *p* values. The two normalization methods were compared based on the inflation of the test statistics and the corresponding Q-Q plots (Fig. 2). The inflation factor was calculated as the median of χ^2 test statistic with 1 degree of freedom which was transformed from Student's *t* test *p* value divided by the expected value of this median under the null hypothesis of no association of any SNP.

Comparison to individual genotyping

Thirty-six SNPs from the Illumina HuHap550 array were genotyped individually using the Sequenom MassArray Genotyping platform (Sequenom Inc, San Diego, CA) on all case and control samples used in the pooling GWAS. Allele frequencies were compared using a *t* test and resulting *p* values were compared to those calculated from the DNA pooling GWAS.

Control DNA samples from the pooled GWAS were also genotyped individually on the Affymetrix 6.0 (Affymetrix, Santa Clara, CA) chip as part of another GWAS performed by the Center for Inherited Disease Research (CIDR). RAFs calculated from the control pools

were compared with allele frequencies calculated from the individual genotyping on 166,033 SNPs which overlapped between the Affymetrix 6.0 chip and the Illumina HuHap550-Duo BeadChip. Spearman correlation coefficients were calculated between the allele frequencies derived from the pooled and individual genotyping data.

Individual genotyping of top SNPs

Sixty-five SNPs were chosen from the most significant results from the DNA pooling GWAS using both normalization methods, with preference toward the log array mean normalized data. These SNPs were individually genotyped on the 500 African American non-DM ESRD cases and 500 controls from the pooling GWAS using the Sequenom MassArray Genotyping platform (Sequenom Inc, San Diego, CA). Primers were designed using the Sequenom Assay Design 3.1 tool and will be made available upon request. After initial genotyping of the individuals from the pooled GWAS, the 65 SNPs were again individually genotyped in a replicate sample of 336 cases and 363 controls. Genotyping efficiency was greater than 95% for all SNPs. Individual genotyping data were analyzed using the analysis program SNPGE (Steigert et al. 2009; Harley et al. 2008). Data were tested for conformity to Hardy–Weinberg proportions, then for association with disease status using the two degree of freedom test with the accompanying genotypic models (dominant, additive and recessive). Association analyses were adjusted for age, gender, BMI, and percentage of African ancestry.

Due to the highly significant association of the *MYH9* gene with non-diabetic ESRD in African Americans, we tested the individually genotyped SNPs for interaction with *MYH9* risk status. We performed tests of association for gene–gene interaction on cases only between the top 65 SNPs and *MYH9*-E1 risk markers. This test models *MYH9*-E1 risk as the response for all individuals with recessive haplotypes at SNPs rs4821480 and rs3752462 (as described in Kopp et al. 2008) and adjusting for age, gender and percentage of African ancestry. To ensure the assumption of independence between a gene and *MYH9*-E1 risk status, markers on chromosome 22 in LD with *MYH9*-E1 were not considered.

Results

Demographic data

African American non-DM ESRD cases were older, had lower BMI, fewer females, and increased percentage of African ancestry compared to non-nephropathy controls (Table 1). Within case and control groups, age, gender, BMI, and African ancestry were similar between the pooled and replication samples.

Pooled GWAS data normalization

The data generated from the Illumina BeadStation for the pooled DNA GWAS were obtained as raw intensity values for the red and green channels, each representing one allele of the SNP. Previous studies using pooled DNA have demonstrated that the array normalization methods designed for the Illumina BeadStation are not appropriate for pooled DNA data (Macgregor et al. 2008). We normalized the array intensity data using the two methods depicted in Fig. 1: “channel mean” and “log array mean”. After normalization, comparison of the mean fluorescent intensities demonstrated that the “log array mean” normalization method reduced the interarray variance in the fluorescent intensity from 0.031 (channel mean normalized data) to 5.36E–06 (log array mean normalized data).

Allele frequencies were calculated from both sets of normalized data, and allele frequencies were compared between cases and controls using a *t* test. Q–Q plots were used to compare the *p* value data quartiles from each normalization scheme to that of a normal distribution to

estimate the data inflation for each method (Fig. 2). The “log array mean normalization” method reduced the data inflation from 1.28 for the channel mean normalization method to 1.13.

To determine the accuracy of the genotyping on pooled DNA samples, we compared the results of the pooled GWAS to individual genotyping allele frequencies and p values. First, we compared t test p values calculated from the pooled GWAS to that calculated from individual genotyping data at 36 SNPs for all the 500 cases and controls (Supplementary Figure 1). The Spearman correlation coefficient was 0.79 indicating significant concordance between pooling and individual genotyping. Next we compared allele frequencies in the 500 control samples from the pooling GWAS with individual genotyping on the same samples on the Affymetrix 6.0 array. This comparison consisted of 166,033 SNPs which overlapped between the Affymetrix 6.0 and the Illumina HumanHap550-duo arrays. A plot of the RAFs from the pooled and individual genotyping on a single pool of 50 control samples is shown in Fig. 3 (the remaining 9 control pools are shown in Supplementary Figure 2). Although over and under estimation of allele frequency was present in the extremes, the pooling and individual genotyping allele frequencies were well correlated. The Spearman correlation coefficients for each of the ten control pools ranged from 0.96 to 0.97 and were consistent between pools (Supplementary Table 1).

Pooled GWAS

The $\log_{10}(1/p)$ of the top 5,000 SNPs of the log array mean normalized data is shown in Fig. 4. While none of the SNPs reached acceptable levels of genome-wide significance [$\log_{10}(1/p) = 8$], several SNPs were highly associated under the t test (with equal and unequal variance). Three SNPs had p values less than $1.0E-6$, rs4462445 (ch 13), rs4821469 (ch 22) and rs8077346 (ch 17). One of these SNPs, rs4821469, is located between the *ApoL2* and *ApoL4* genes, and about 100 kb from the *MYH9* gene which is strongly associated with non-DM ESRD in African Americans (Kopp et al. 2008; Freedman et al. 2009b). A number of other SNPs in this region of chromosome 22 were also associated with non-DM ESRD and had less significant p values ($<10^{-6}$).

Individual genotyping and replication

Top scoring SNPs from both normalization methods, with prioritization on the “log array mean” dataset were chosen for follow-up through individual genotyping of the original group of 500 cases and 500 controls and replication in an additional group of 336 cases and 363 controls. Of the 1,000 cases and controls, 36 cases and 22 controls were not included in analysis due to failed genotyping or missing phenotypic data. For individual genotyping of the pooled samples, three SNPs, rs4751759, rs17608620, and rs139998, deviated from Hardy–Weinberg equilibrium in the control samples ($7.7E-4$; Bonferroni corrected). Association analysis revealed that 16 SNPs were associated with non-DM ESRD ($p < 7.7E-4$; Table 2). Twelve of these SNPs are located on chromosome 22, in or near the *MYH9* gene. Of the remaining four SNPs, two are located on chromosome 8, rs4733947 and rs10112543, although on different arms [8p11.22 (*TACCI*) and 8q23.3 (*CSMD3*), respectively] and rs2034906 and rs4462445 are located on chromosome 3 (in *HLTF*) and chromosome 13 (3' of *FLT1*), respectively.

Nearly all significantly associated *MYH9* SNPs remained associated with non-DM ESRD in the replication samples (Table 2). When the two sets of samples were combined, the chromosome 22 SNPs in *MYH9* were the most strongly associated, with p values ranging from $2.26E-8$ to $1.78E-19$. The four non-*MYH9* SNPs that were associated with non-DM ESRD in the original set of pooled samples were modestly associated in the combined set, most likely driven by the data in the pooled samples.

SNPs from individual genotyping were prioritized as follows: (1) significant association with non-DM ESRD in the original samples, replication samples and combined analysis, (2) association in the combined sample was stronger than association in either of the two groups individually, (3) significant association in either the original sample or the replication set and potential for functional relevance in ESRD. Using these criteria, we identified six genes or regions for follow-up studies (Table 3). While many of these SNPs are not significantly associated under Bonferroni corrected p values, the increased significance in the combined analysis, and potential functional relevance of some of these genes indicate that they should be targets for follow-up candidate gene analysis.

MYH9 interaction analysis

We tested the 55 individually genotyped SNPs (excluding SNPs on chromosome 22) for interaction with *MYH9* risk haplotype status (Supplementary Table 2). While none of the SNPs were associated under a Bonferroni corrected p value ($7.0E-04$), three SNPs were modestly associated with either the pooled, replication or combined set of samples, rs11579379 (ch 1), rs10112543 (ch 8), and rs4751759 (ch 10).

Discussion

We have demonstrated the utility of pooled DNA samples in the first genome-wide association analysis for non-DM ESRD in African Americans. There are several advantages of using pooled DNA for large scale genotyping, in particular the reduction in time and cost of genotyping. For example, we ran 40 Illumina Beadchips in 1 week for this GWAS (10 case and 10 controls pools run in duplicate), whereas genotyping individual samples would have required 1,000 chips and several months of laboratory time. The use of pooled DNA for GWAS does have some limitations. Any pooling-based study is subject to error imposed by the creation of the pool. Great care was taken to accurately determine concentration and quality of the DNA used for the pooling GWAS. Based on the high correlation between allele frequencies from pooling and individual genotyping, we can conclude that the error due to pooling is minimal. Using pooled samples we also have limited power to detect certain effects. That we were able to detect SNPs in the most strongly associated gene to date (*MYH9*) demonstrates that we had enough power to detect a strong association and lends confidence to some of the other high ranking SNPs we have identified. Finally, the HuHap550 bead chip that was used for this GWAS was not comprehensive for an African American sample and some signals may have been missed due to inadequate coverage for this population.

Success of a pooled GWAS is based largely on the accuracy of pool construction. Using the individual genotyping data on the controls from a GWAS that was performed on the Affymetrix 6.0 array, we were able to show a high level of correlation (0.96–0.97) between allele frequencies derived from pooled DNA and individual genotyping. This indicated highly accurate pool construction and genotyping. The level of correlation was also consistent among the individual pools, further supporting the quality of the pooled genotyping data.

Despite the high quality of the pooling data, we were unable to demonstrate association of any SNP at genome-wide levels of significance [$\log_{10}(1/p) = 8$]. Three SNPs had p values approaching genome-wide significance. Notably, there were a number of SNPs with values above five in the *MYH9* region of chromosome 22. Variants in the *MYH9* have a substantial effect on risk for non-DM ESRD in African Americans (Kao et al. 2008; Kopp et al. 2008). That we replicated this association with our highest scoring SNPs in the pooled GWAS lends veracity to our data and supports the potential for true association at other high scoring loci.

The pooling GWAS was followed-up with individual genotyping of the top scoring SNPs from both normalization methods; with emphasis on the log array mean normalized data because it had the lowest inflation (1.13). Significant p values in chromosome 22 SNPs in the region of *MYH9* ranged from 0.027 to 2.6×10^{-14} . The majority of these SNPs were also associated with non-DM ESRD in the smaller replicate sample and in the combined analysis. Four other SNPs (rs4658749, rs350792, rs1928722, and rs155705) were well associated ($p < 3.0 \times 10^{-4}$) in the pooled GWAS, but were not as well associated in the individual genotyping of the pooled and replicate samples. These SNPs, located in *KIF26B*, *LOC254312*, *TMEM132D* and an intergenic region, were more (or equally) significantly associated after the samples from individual genotyping were combined. It is unclear as to why these SNPs were not associated in the replication samples. The replication sample set was smaller and therefore had less power than the analysis of the original set of cases and controls or all samples combined. It is possible that there was not enough power in the replication sample set to detect association and that the association was equally or more significant in the combined sample set leads us to believe that these associations may be genuine.

MYH9 has already been shown to be strongly associated with ESRD and has been studied in a number of renal disease phenotypes in African Americans (Freedman et al. 2009a, b, c, d). We hypothesize that there are secondary factors which, combined with *MYH9* risk alleles, lead to ESRD. For example, in HIV-associated nephropathy (HIVAN), having *MYH9* risk alleles and HIV infection significantly increases the risk for ESRD (vs. HIV alone) (Kopp et al. 2008). Other “secondary hits” could include other (non-HIV) viral infections, environmental factors, or additional genes, particularly those whose functions interact with *MYH9*.

The remaining loci we have identified may fall into the category of “second hits” or gene–gene interactions (Table 3). *KIF26B* (kinesin family member 26B) is a nucleotide binding protein which is involved in microtubule motor activity in the cell and could be important for maintaining podocyte cytoskeletal structure. *CTXN3*, cortixin 3, is expressed exclusively in the brain and kidney; however, its function is unknown (Wang et al. 2007). *TMEM132D* is a transmembrane protein with no known functional relevance to renal disease. There was also significant association at one SNP (rs1928722) in a hypothetical gene *LOC254312* and one SNP (rs350792) in an intergenic region with no known genes within 500 kb. These genes and regions will be tagged as part of follow-up candidate gene analysis in our non-DM ESRD population. Future studies include genotyping top hits from this GWAS on individual case and control samples using a 1,536 GoldenGate array from Illumina. Analysis of this data will include standard association analyses, as well as interaction with *MYH9*, to identify potential “second hit” genes which might work in concert with *MYH9* to lead to ESRD.

We performed the first GWAS for non-DM ESRD in African Americans using pooled DNA. Using allele frequencies from individual genotyping on control samples, we have demonstrated accuracy in pool construction and genotyping in these samples and developed a novel method for normalization of the intensity data to reduce inter-array variance in the samples. In the pooling GWAS, we confirmed association with the *MYH9* gene, further supporting the accuracy of the pooled DNA genotyping. We also identified several genes and regions that will be followed-up as part of candidate gene analysis with the intent of identifying a “second hit” gene that may interact with *MYH9* and contribute to the risk of ESRD in African Americans.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported in part by NIH Grants R01 DK 070941 (BIF) and R01 DK53591 (DWB), and by the NIDDK and NCI Intramural Research Programs. MAB was supported by F32 DK080617 from the NIDDK. We gratefully acknowledge the contributions of the participants as well as the physicians who were part of the study and the work of our study coordinators Joyce Byers, Carrie Smith, Mitzie Spainhour, Cassandra Bethea, and Sharon Warren.

References

- Appel LJ, Wright JT Jr, Greene T, Kusek JW, Lewis JB, Wang X, Lipkowitz MS, Norris KC, Bakris GL, Rahman M, Contreras G, Rostand SG, Kopple JD, Gabbai FB, Schulman GI, Gassman JJ, Charleston J, Agodoa LY. Long-term effects of renin-angiotensin system-blocking therapy and a low blood pressure goal on progression of hypertensive chronic kidney disease in African Americans. *Arch Intern Med.* 2008; 168:832–839. [PubMed: 18443258]
- Chou JW, Paules RS, Bushel PR. Systematic variation normalization in microarray data to get gene expression comparison unbiased. *J Bioinform Comput Biol.* 2005; 3:225–241. [PubMed: 15852502]
- Dusel JA, Burdon KP, Hicks PJ, Hawkins GA, Bowden DW, Freedman BI. Identification of podocin (NPHS2) gene mutations in African Americans with nondiabetic end-stage renal disease. *Kidney Int.* 2005; 68:256–262. [PubMed: 15954915]
- Freedman BI, Langefeld CD, Rich SS, Valis CJ, Sale MM, Williams AH, Brown WM, Beck SR, Hicks PJ, Bowden DW. A genome scan for ESRD in black families enriched for nondiabetic nephropathy. *J Am Soc Nephrol.* 2004; 15:2719–2727. [PubMed: 15466277]
- Freedman BI, Bowden DW, Rich SS, Valis CJ, Sale MM, Hicks PJ, Langefeld CD. A genome scan for all-cause end-stage renal disease in African Americans. *Nephrol Dial Transplant.* 2005; 20:712–718. [PubMed: 15701670]
- Freedman BI, Hicks PJ, Bostrom MA, Comeau ME, Divers J, Bleyer AJ, Kopp JB, Winkler CA, Nelson GW, Langefeld CD, Bowden DW. Non-muscle myosin heavy chain 9 gene MYH9 associations in African Americans with clinically diagnosed type 2 diabetes mellitus-associated ESRD. *Nephrol Dial Transplant.* 2009a; 24:3366–3371. [PubMed: 19567477]
- Freedman BI, Hicks PJ, Bostrom MA, Cunningham ME, Liu Y, Divers J, Kopp JB, Winkler CA, Nelson GW, Langefeld CD, Bowden DW. Polymorphisms in the non-muscle myosin heavy chain 9 gene (MYH9) are strongly associated with end-stage renal disease historically attributed to hypertension in African Americans. *Kidney Int.* 2009b; 75:736–745. [PubMed: 19177153]
- Freedman BI, Kopp JB, Winkler CA, Nelson GW, Rao DC, Eckfeldt JH, Leppert MF, Hicks PJ, Divers J, Langefeld CD, Hunt SC. Polymorphisms in the nonmuscle myosin heavy chain 9 gene (MYH9) are associated with albuminuria in hypertensive African Americans: the HyperGEN study. *Am J Nephrol.* 2009c; 29:626–632. [PubMed: 19153477]
- Freedman BI, Nagaraj SK, Lin JJ, Gautreaux MD, Bowden DW, Iskandar SS, Stratta RJ, Rogers J, Hartmann EL, Farney AC, Reeves-Daniel AM. Potential donor-recipient MYH9 genotype interactions in posttransplant nephrotic syndrome after pediatric kidney transplantation. *Am J Transplant.* 2009d; 9:2435–2440. [PubMed: 19764949]
- Hanson RL, Craig DW, Millis MP, Yeatts KA, Kobes S, Pearson JV, Lee AM, Knowler WC, Nelson RG, Wolford JK. Identification of PVT1 as a candidate gene for end-stage renal disease in type 2 diabetes using a pooling-based genome-wide single nucleotide polymorphism association study. *Diabetes.* 2007; 56:975–983. [PubMed: 17395743]
- Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci. *Nat Genet.* 2008; 40:204–210. [PubMed: 18204446]

- Hicks PJ, Staten JL, Palmer ND, Langefeld CD, Ziegler JT, Keene KL, Sale MM, Bowden DW, Freedman BI. Association analysis of the ephrin-B2 gene in African-Americans with end-stage renal disease. *Am J Nephrol.* 2008; 28:914–920. [PubMed: 18580054]
- Kao WH, Klag MJ, Meoni LA, Reich D, Berthier-Schaad Y, Li M, Coresh J, et al. MYH9 is associated with nondiabetic end-stage renal disease in African Americans. *Nat Genet.* 2008; 40:1185–1192. [PubMed: 18794854]
- Keene KL, Mychaleckyj JC, Leak TS, Smith SG, Perlegas PS, Divers J, Langefeld CD, Freedman BI, Bowden DW, Sale MM. Exploration of the utility of ancestry informative markers for genetic association studies of African Americans with type 2 diabetes and end stage renal disease. *Hum Genet.* 2008; 124:147–154. [PubMed: 18654799]
- Kirov G, Zaharieva I, Georgieva L, Moskvina V, Nikolov I, Cichon S, Hillmer A, Toncheva D, Owen MJ, O'Donovan MC. A genome-wide association study in 574 schizophrenia trios using DNA pooling. *Mol Psychiatry.* 2009; 14:796–803. [PubMed: 18332876]
- Kopp JB, Smith MW, Nelson GW, Johnson RC, Freedman BI, Bowden DW, Oleksyk T, McKenzie LM, Kajiyama H, Ahuja TS, Berns JS, Briggs W, Cho ME, Dart RA, Kimmel PL, Korbet SM, Michel DM, Mokrzycki MH, Schelling JR, Simon E, Trachtman H, Vlahov D, Winkler CA. MYH9 is a major-effect risk gene for focal segmental glomerulosclerosis. *Nat Genet.* 2008; 40:1175–1184. [PubMed: 18794856]
- Macgregor S, Zhao ZZ, Henders A, Nicholas MG, Montgomery GW, Visscher PM. Highly cost-efficient genome-wide association studies using DNA pools and dense SNP arrays. *Nucleic Acids Res.* 2008; 36:e35. [PubMed: 18276640]
- Steer S, Abkevich V, Gutin A, Cordell HJ, Gendall KL, Merriman ME, Rodger RA, Rowley KA, Chapman P, Gow P, Harrison AA, Highton J, Jones PB, O'Donnell J, Stamp L, Fitzgerald L, Iliev D, Kouzmine A, Tran T, Skolnick MH, Timms KM, Lanchbury JS, Merriman TR. Genomic DNA pooling for whole-genome association scans in complex disease: empirical demonstration of efficacy in rheumatoid arthritis. *Genes Immun.* 2007; 8:57–68. [PubMed: 17159887]
- Steigert, ML.; Grab, JD.; Guy, RT.; Langefeld, CD. SNPGWA Version 4.0 (computer program). Public Health Sciences, Wake Forest University; 2009.
- U.S. Renal Data System. Annual Data Report: Atlas of chronic kidney disease and end-stage renal disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; 2008.
- Wang HT, Chang JW, Guo Z, Li BG. In silico-initiated cloning and molecular characterization of cortxin 3, a novel human gene specifically expressed in the kidney and brain, and well conserved in vertebrates. *Int J Mol Med.* 2007; 20:501–510. [PubMed: 17786280]

Fig. 1.
Scheme for the normalization of Illumina BeadArray intensity data

Fig. 2.
Q-Q plots of the t test p value data quartiles compared to a normal distribution. Plots and corresponding inflation values are shown for both normalization methods

Fig. 3. Comparison of relative allele frequency (RAF) between individual genotyping and pooling of samples from one control pool of 50 individuals using 166,033 SNPs

Fig. 4.

$\text{Log}_{10}(1/p)$ of the top 5,000 SNPs from the two sample t test with equal and unequal variance using the “log array mean” dataset

Table 1
Demographic data for the entire collection of non-diabetic ESRD cases and non-nephropathy controls

	<i>N</i>	Female (%)	Age	BMI	African ancestry
Pooled					
Non-diabetic ESRD	500	45.1 ^a	54.6 ± 14.8 ^a	26.7 ± 6.7 ^a	79.3 ± 10.9 ^a
Controls	500	53.7	50.6 ± 11.6	30.0 ± 7.0	76.6 ± 11.6
Replication					
Non-diabetic ESRD	336	42.2 ^a	53.5 ± 14.2 ^a	27.0 ± 7.2 ^a	80.4 ± 9.8 ^a
Controls	363	57.7	47.1 ± 12.1	30.0 ± 7.1	77.5 ± 10.5
Combined					
Non-diabetic ESRD	800	43.9 ^a	54.1 ± 14.6 ^a	26.8 ± 6.9 ^a	79.8 ± 10.4 ^a
Controls	841	55.6	49.1 ± 11.9	30.0 ± 7.1	77.0 ± 11.1

Data include the percentage of females, average (±SD) age, BMI and percentage of African ancestry (as determined from ancestry informative markers)

^aSignificantly different between cases and controls *P* < 0.05

Table 2

Minor allele frequency (MAF), Hardy–Weinberg equilibrium (HWE) and genotypic association (GA) under the best model for individual genotyping of the top 65 SNPs in the samples used for pooling (464 cases and 478 controls), the replication (336 cases and 363 controls) and the combined set (800 cases and 841 controls)

Marker	CHR	MA	Pooled samples (cases = 464, controls = 478)			Replication samples (cases = 336, controls = 363)			Combined samples (cases = 800, controls = 841)		
			Control MAF	Control HWE <i>p</i>	GA Best <i>p</i>	Model	Control MAF	Control HWE <i>p</i>	GA Best <i>p</i>	Model	Model
rs1200100	1	A	0.262	0.470	0.731	add	0.261	0.100	0.918	add	add
rs2273366	1	G	0.129	1.000	0.039	dom	0.127	0.054	0.575	add	add
rs6684819	1	T	0.177	0.201	0.044	add	0.164	0.700	0.921	add	add
rs2992644	1	C	0.391	0.631	0.012	add	0.427	0.130	0.325	add	add
rs4658749	1	A	0.375	0.326	0.001	add	0.320	0.629	0.719	add	add
rs111579379	1	A	0.515	0.851	0.001	add	0.496	1.000	0.756	add	add
rs4581855	2	A	0.366	0.235	0.005	add	0.429	0.827	0.432	add	add
rs4402815	2	A	0.049	0.090	0.309	add	0.048	0.566	0.948	add	add
rs350792	2	T	0.261	0.024	0.054	add	0.254	0.403	0.063	add	add
rs7565678	2	G	0.454	0.353	0.224	add	0.469	0.522	0.191	add	add
rs1434080	2	C	0.200	0.566	0.004	dom	0.166	0.342	0.570	add	add
rs12986657	2	T	0.132	0.104	0.015	add	0.094	0.754	0.039	add	add
rs888090	2	T	0.370	0.547	0.047	add	0.418	0.667	0.473	add	add
rs1822120	3	C	0.344	0.472	0.218	add	0.348	0.241	0.136	rec	add
rs2034906	3	T	0.129	0.150	1.808E-04	add	0.170	0.263	0.996	add	add
rs1993522	3	C	0.071	0.280	0.205	add	0.055	0.295	0.404	add	add
rs7638569	3	A	0.403	0.446	0.054	add	0.442	0.068	0.746	add	dom
rs6779973	3	A	0.440	0.853	0.549	add	0.417	0.744	0.872	add	add
rs13138705	4	T	0.122	1.000	0.024	dom	0.150	0.533	0.327	add	add
rs10030137	4	T	0.098	0.434	0.008	add	0.120	0.448	0.472	add	add
rs9990730	4	A	0.163	1.000	0.013	dom	0.189	1.000	0.323	add	add
rs6869417	5	T	0.179	0.875	0.002	add	0.178	0.148	0.870	add	add
rs10063554	5	A	0.184	0.642	0.932	add	0.200	0.617	0.186	add	add
rs248694	5	A	0.371	0.429	0.094	add	0.359	0.138	0.001	dom	dom
rs6569884	6	G	0.163	0.865	0.097	add	0.140	0.026	0.067	dom	add
rs958386	6	T	0.151	0.030	0.733	add	0.132	0.486	0.097	add	add

Marker	CHR	MA	Pooled samples (cases = 464, controls = 478)				Replication samples (cases = 336, controls = 363)				Combined samples (cases = 800, controls = 841)			
			Control MAF	Control HWE <i>p</i>	GA Best <i>p</i>	Model	Control MAF	Control HWE <i>p</i>	GA Best <i>p</i>	Model	Control MAF	Control HWE <i>p</i>	GA Best <i>p</i>	Model
rs4286900	7	A	0.134	0.322	0.095	add	0.111	1.000	0.848	add	0.124	0.522	0.169	add
rs17647278	7	A	0.032	1.000	0.114	add	0.021	1.000	0.231	add	0.027	1.000	0.566	add
rs10245560	7	A	0.257	0.811	0.004	add	0.199	0.243	0.066	add	0.232	0.382	0.276	add
rs4733947	8	G	0.324	1.000	7.621E-06	add	0.298	0.706	0.426	add	0.313	0.808	2.134E-04	add
rs2247329	8	T	0.170	1.000	0.168	add	0.170	0.457	0.251	add	0.170	0.619	0.028	dom
rs10112543	8	C	0.200	0.772	3.735E-05	add	0.279	1.000	0.428	add	0.234	1.000	0.016	add
rs2102103	8	T	0.250	0.326	0.002	dom	0.205	0.330	0.293	add	0.230	0.922	0.115	add
rs12156507	9	G	0.409	0.683	0.029	add	0.393	0.368	0.811	add	0.402	0.763	0.064	add
rs888401	9	G	0.466	0.111	0.003	add	0.410	0.913	0.933	add	0.441	0.157	0.022	add
rs1928722	10	G	0.206	0.888	0.008	add	0.215	0.532	0.248	add	0.210	0.530	0.005	add
rs3012499	10	A	0.227	0.239	0.016	add	0.278	0.012	0.092	add	0.249	0.007	0.457	add
rs1484170	10	C	0.171	0.621	0.158	add	0.179	1.000	0.908	add	0.175	0.808	0.360	add
rs4751759	10	A	0.177	2.288E-05	0.004	add	0.139	0.001	0.236	add	0.160	0.000	0.138	add
rs10886596	10	A	0.159	0.081	0.546	add	0.168	1.000	0.957	add	0.163	0.200	0.749	add
rs155705	12	A	0.116	0.819	0.001	dom	0.115	0.797	0.293	add	0.115	1.000	0.001	dom
rs4462445	13	C	0.510	0.454	8.655E-06	add	0.461	0.018	0.435	add	0.489	0.041	0.003	add
rs2170765	13	C	0.291	0.738	0.037	add	0.261	0.490	0.128	add	0.278	0.436	0.512	add
rs7998345	13	G	0.263	0.723	0.002	add	0.228	1.000	0.615	add	0.248	0.853	0.007	add
rs2596179	15	C	0.326	0.142	0.008	add	0.348	1.000	0.938	add	0.336	0.309	0.046	add
rs16946398	16	G	0.129	0.839	0.002	add	0.170	0.345	0.669	add	0.146	0.332	0.030	dom
rs8077346	17	T	0.145	0.059	0.068	add	0.128	0.057	0.636	add	0.137	0.008	0.085	add
rs2186820	18	G	0.291	0.373	0.026	add	0.246	0.061	0.700	add	0.272	0.053	0.056	add
rs1785394	18	T	0.338	0.038	0.144	add	0.347	0.003	0.645	add	0.342	0.000	0.366	add
rs8105198	19	C	0.286	0.575	0.089	add	0.305	0.168	0.896	add	0.294	0.155	0.142	add
rs17608620	19	T	0.068	0.000	0.007	add	0.034	0.000	0.038	add	0.053	0.000	0.430	add
rs134378	22	G	0.209	0.331	0.006	add	0.186	1.000	0.142	add	0.199	0.514	0.002	add
rs7364143	22	A	0.394	0.378	6.445E-08	add	0.399	0.507	6.755E-06	add	0.396	0.271	2.091E-12	add
rs9610448	22	T	0.345	0.101	7.986E-06	add	0.306	0.266	0.001	add	0.328	0.041	3.393E-08	add
rs7289037	22	T	0.336	1.000	2.311E-10	add	0.357	0.419	7.005E-06	add	0.345	0.535	1.453E-14	add
rs4820222	22	G	0.354	0.221	1.813E-07	dom	0.317	0.221	1.779E-04	add	0.338	0.072	3.493E-10	add

Marker	CHR	MA	Pooled samples (cases = 464, controls = 478)				Replication samples (cases = 336, controls = 363)				Combined samples (cases = 800, controls = 841)			
			Control MAF	Control HWE <i>p</i>	GA Best <i>p</i>	Model	Control MAF	Control HWE <i>p</i>	GA Best <i>p</i>	Model	Control MAF	Control HWE <i>p</i>	GA Best <i>p</i>	Model
rs4821467	22	T	0.335	0.829	2.483E-10	add	0.357	0.563	2.526E-06	add	0.345	0.528	4.771E-15	add
rs4821469	22	A	0.504	0.706	2.601E-14	dom	0.473	0.750	4.737E-07	dom	0.491	0.944	1.784E-19	dom
rs2010467	22	A	0.423	0.444	2.363E-07	add	0.453	0.915	0.013	add	0.436	0.617	2.267E-08	add
rs713753	22	T	0.217	0.582	1.026E-08	dom	0.216	0.209	8.343E-08	dom	0.216	0.178	4.052E-15	dom
rs2157256	22	C	0.367	0.489	3.285E-10	dom	0.322	0.394	7.591E-06	add	0.348	1.000	3.341E-14	dom
rs2239784	22	G	0.441	1.000	1.477E-06	add	0.452	0.915	3.279E-09	dom	0.446	1.000	7.280E-14	add
rs139933	22	T	0.223	0.686	4.133E-06	add	0.240	0.469	3.704E-04	add	0.230	0.429	4.201E-09	add
rs139998	22	A	0.417	4.757E-04	1.127E-06	dom	0.404	0.445	0.001	add	0.411	0.001	2.824E-09	dom
rs16991689	22	G	0.081	0.534	0.027	add	0.057	0.322	0.556	add	0.071	0.294	0.169	add

Genotypic association was calculated with adjustment for logBMI, age, gender and African ancestry

CHR chromosome, MA minor allele

Table 3

Genotyping association *p* values, odds ratios (OR) and confidence intervals (CI) for SNPs significant in either the pooled or replication sample and the combined analysis

Gene	SNP	Pooled			Replication			Combined		
		Best <i>p</i>	OR	CI	Best <i>p</i>	OR	CI	Best <i>p</i>	OR	CI
<i>KIF26B</i>	rs4658749	0.001	0.69	0.56–0.86	0.719	0.96		0.005	0.79	
None	rs350792	0.054	0.80	0.64–1	0.063	0.78	0.6–1.01	0.008	0.79	0.67–0.94
Near <i>CTXN3</i>	rs248694	0.094	1.19	0.97–1.45	0.001	1.82	1.3–2.55	4.987E-04	1.47	1.18–1.83
<i>LOC254312</i>	rs1928722	0.008	1.36	1.08–1.71	0.248	1.17	0.9–1.53	0.005	1.28	1.07–1.52
<i>TMEM132D</i>	rs155705	0.001	0.53	0.36–0.76	0.293	0.81	0.56–1.19	0.001	0.64	0.49–0.84
<i>MYH9</i>	rs134378	0.006	0.69	0.53–0.9	0.142	0.79	0.58–1.08	0.002	0.73	0.6–0.89
	rs7364143	6.445E-08	1.71	1.41–2.08	6.755E-06	1.68	1.34–2.11	2.091E-12	1.7	1.46–1.96
	rs9610448	7.986E-06	0.62	0.5–0.76	0.001	0.65	0.51–0.85	3.393E-08	0.63	0.54–0.74
	rs7289037	2.311E-10	1.93	1.58–2.37	7.005E-06	1.7	1.35–2.15	1.453E-14	1.82	1.56–2.12
	rs4820222	1.813E-07	0.47	0.36–0.63	1.779E-04	0.61	0.47–0.79	3.493E-10	0.6	0.51–0.7
	rs4821467	2.483E-10	1.97	1.6–2.43	2.526E-06	1.76	1.39–2.23	4.771E-15	1.86	1.59–2.18
	rs4821469	2.601E-14	0.31	0.22–0.41	4.737E-07	0.42	0.3–0.59	1.784E-19	0.35	0.28–0.44
	rs2010467	2.363E-07	1.67	1.38–2.03	0.013	1.34	1.06–1.68	2.267E-08	1.52	1.31–1.77
	rs713753	1.026E-08	0.39	0.28–0.54	8.343E-08	0.36	0.25–0.53	4.052E-15	0.38	0.3–0.48
	rs2157256	3.285E-10	0.41	0.31–0.54	7.591E-06	0.56	0.43–0.72	3.341E-14	0.44	0.36–0.54
	rs2239784	1.477E-06	0.60	0.49–0.74	3.279E-09	0.37	0.26–0.51	7.280E-14	0.56	0.48–0.65
	rs139933	4.133E-06	1.67	1.34–2.09	3.704E-04	1.58	1.23–2.04	4.201E-09	1.64	1.39–1.93
	rs139998	1.127E-06	0.49	0.37–0.66	0.001	0.66	0.52–0.83	2.824E-09	0.52	0.42–0.65
	rs16991689	0.027	0.63	0.42–0.95	0.556	1.15	0.72–1.86	0.169	0.81	0.59–1.1