



Published in final edited form as:

*Hum Genet.* 2009 December ; 126(6): 805–817. doi:10.1007/s00439-009-0732-8.

## Genetic Analysis of Diabetic Nephropathy on Chromosome 18 in African Americans: Linkage Analysis and Dense SNP Mapping

Caitrin W. McDonough<sup>1,2</sup>, Meredith A. Bostrom<sup>1,3,4</sup>, Lingyi Lu<sup>5</sup>, Pamela J. Hicks<sup>1,3,4</sup>, Carl D. Langefeld<sup>5</sup>, Jasmin Divers<sup>5</sup>, Josyf C. Mychaleckyj<sup>6</sup>, Barry I. Freedman<sup>7</sup>, and Donald W. Bowden<sup>1,3,4,7</sup>

<sup>1</sup>Center for Human Genomics, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157

<sup>2</sup>Program in Molecular Medicine and Translational Science, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157

<sup>3</sup>Center for Diabetes Research, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157

<sup>4</sup>Department of Biochemistry, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157

<sup>5</sup>Department of Public Health Sciences, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157

<sup>6</sup>Department of Public Health Sciences, University of Virginia, Charlottesville, VA

<sup>7</sup>Department of Internal Medicine, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157

### Abstract

Genetic studies in Turkish, Native American, European American, and African American (AA) families have linked chromosome 18q21.1-23 to susceptibility for diabetes associated nephropathy. In this study we have carried out fine linkage mapping in the 18q region previously linked to diabetic nephropathy in AAs by genotyping both microsatellite and single nucleotide polymorphisms (SNPs) for linkage analysis in an expanded set of 223 AA families multiplexed for type 2 diabetes associated ESRD (T2DM-ESRD). Several approaches were used to evaluate evidence of linkage with the strongest evidence for linkage in ordered subset analysis with an earlier age of T2DM diagnosis compared to the remaining pedigrees (LOD 3.9 at 90.1cM,  $\Delta P=0.0161$ , NPL  $P$  value = 0.00002). Overall, the maximum LODs and LOD-1 intervals vary in magnitude and location depending upon analysis. The linkage mapping was followed up by performing a dense SNP map, genotyping 2,814 SNPs in the refined LOD-1 region in 1,029 AA T2DM-ESRD cases and 1,027 AA controls. Of the top 25 most associated SNPs, 10 resided within genic regions. Two candidate genes stood out: *NEDD4L* and *SERPINB7*. SNP rs512099, located in intron 1 of *NEDD4L*, was associated under a dominant model of inheritance ( $P$  value = 0.0006; Odds ratio (95% Confidence Interval) (OR (95% CI)) = 0.70 (0.57-0.86)). SNP rs1720843, located in intron 2 of *SERPINB7*, was associated under a recessive model of inheritance ( $P$  value = 0.0017; OR (95% CI) = 0.65 (0.50-0.85)). Collectively, these results suggest that multiple genes in this region may influence diabetic nephropathy susceptibility in AAs.

## Keywords

African American; diabetes type 2; nephropathy; linkage analysis; SNP; association analysis

---

## Introduction

African Americans have a 3.6-fold increased risk of developing end-stage renal disease (ESRD) compared to Caucasian Americans and at least a 1.8-fold increased risk of developing ESRD compared to other racial/ethnic minority groups in the United States (U.S. Renal Data System 2008). Diabetes-associated nephropathy is the most common source of ESRD, accounting for approximately 45% of cases in the U.S., and the incidence rate continues to rise, 2.5% in 2006 (U.S. Renal Data System 2008). Many different studies, using a variety of different methods have shown that there is a genetic component to ESRD in the general population and in African Americans as summarized (Bowden 2003). When looking at familial aggregation, African Americans who have a close relative with ESRD have a 9-fold increased risk of developing ESRD (Freedman et al. 1995), while Caucasian Americans only have an increased risk of 2.7-fold (Spray et al. 1995). Taking these results together, the African American population has a disproportionately higher incidence of ESRD along with a stronger familial component. Diabetes is the single most common contributor to ESRD in the United States, yet the origins of diabetes-associated nephropathy and ESRD are poorly understood. It remains unclear why some diabetes affected individuals will progress to nephropathy and ESRD, and why others will not. This suggests that environmental factors, as well as genetic susceptibility, contribute to nephropathy risk.

We previously reported evidence for linkage of diabetic nephropathy to chromosome 18q21.1-23 in 166 African American families (Bowden et al. 2004). Linkage and association analysis of other populations have provided consistent evidence for a diabetic nephropathy (DN) susceptibility locus/loci on 18q: 18q22.3-23 in Turkish and Pima Indian families (Vardarli et al. 2002), 18q22.3 in European American and American Indian families (Iyengar et al. 2007), and European American type 1 diabetes patients (Ewens et al. 2005). Here we report results from an expanded genetic linkage analysis and association analysis of a high density SNP map across the 18q21.1-23 interval.

## Materials and Methods

### Subjects

**Family-based linkage analysis**—Recruitment and sample collection procedures were approved by the Institutional Review Board at Wake Forest University. DNA samples were collected from self-described African American families with multiple type 2 diabetes mellitus (T2DM) affected subjects with either end-stage renal disease (ESRD) or chronic renal failure (CRF). For the purposes of this report these affected individuals will be referred to collectively as diabetic nephropathy (DN) cases and are treated the same with the exception of analyses that incorporate age at diagnosis of ESRD or duration of diabetes to ESRD. In these analyses only ESRD cases (and not CRF cases) were incorporated into the models since clear definition of age at onset of CRF were not possible.

Briefly, families were originally identified through a proband with T2DM associated ESRD. T2DM was diagnosed in probands developing diabetes and treated with diet and exercise or oral hypoglycemic agents during at least part of their disease history. Medical records were reviewed to verify the etiology of the nephropathy. Renal failure was attributed to diabetes when serum creatinine  $\geq 2.0$  mg/dl with either diabetes duration for  $>10$  years, or proliferative diabetic retinopathy in the absence of other known causes of renal failure.

When proteinuria data was available, all subjects had CRF defined as proteinuria  $\geq 500$  mg/24 hours, a urine protein:creatinine ratio  $\geq 0.5$  mg/g or  $\geq 100$  mg/dl proteinuria on urine dipstick. Diabetic nephropathy (DN) affected siblings and, when possible, other available family members were recruited also. Selection criteria and recruitment strategies have been previously described in detail (Freedman et al. 1997; Yu et al. 1996; Yu et al. 1998; Yu et al. 2000; Freedman et al. 2000; Freedman et al. 2002). The family set for the linkage fine mapping comprised 223 African American families with 270 DN affected sibling pairs, made up primarily of 233 full-sibling pairs and 37 half-sibling pairs, from a total of 476 DN affected individuals. One hundred seventy-one of the families contained 2 affected siblings, 17 families had 3 affected siblings, and 2 families had 4 affected siblings with a total of 796 individual subjects. In general, the family data consisted primarily of individuals from a single generation, with both parents available in none of the families and one parent for 6 families. Of the DN affected individuals, 406 had T2DM with ESRD and 70 had diabetes with chronic renal failure (CRF). Seventy-one individuals in the families had T2DM without a diagnosis of ESRD or CRF, of which 33 were unaffected and 38 had unknown renal status. DNA extraction was performed using the PureGene system (Gentra Systems, Minneapolis, MN).

**Case-control samples for dense SNP mapping and gene analysis**—DNAs from a total of 1,029 DN cases and 1,027 non-diabetic controls without nephropathy were genotyped for association analysis. DN cases were ascertained, recruited, and diagnosed using the same manner as outlined above for the family collection. Controls were healthy, self-reported African Americans born in North Carolina, age  $\geq 18$  years, and denying a personal or family history of kidney disease in 1<sup>st</sup> degree relatives. Controls were recruited from community resources including health fairs, churches, and shopping malls. Recruitment and sample collection procedures were approved by the Institutional Review Board at Wake Forest University. DNA extraction was performed using the PureGene system (Gentra Systems, Minneapolis, MN).

## Genotyping

**Family-based linkage analysis**—There were 64 additional DN-affected sibling pairs added to the African American family population for linkage mapping. These subjects were part of a genome-wide scan completed by the Center for Inherited Disease Research (CIDR), through the National Institute of Diabetes and Digestive and Kidney Diseases–funded Family Investigation of Nephropathy and Diabetes. The marker set was based on Marshfield Panel 8, with  $\sim 10\%$  of the markers changed from the previous Marshfield panel. It was composed of mainly tetra-, tri-, and di- nucleotide repeats, including 385 primer pairs with an average spacing of 9.0cM and no intermarker gaps greater than 20cM.

SNPs for fine mapping were chosen using the marker list from Illumina's HumanLinkage-12 Marker Panel downloaded from the Center for Inherited Disease Research (CIDR) website ([www.cidr.jhmi.edu/snp\\_marker.html](http://www.cidr.jhmi.edu/snp_marker.html)) using genome build version 34. SNPs were picked from the LOD-1 region (bld34: 49,229,785-75,887,528) of the previously published linkage peak (Bowden et al. 2004). SNPs were selected with an average spacing of 0.97cM, and no interSNP gaps greater than 2.6cM. When there was more than one SNP in the region, the SNP with the highest minor allele frequency in the African American population was chosen. Map distances were based on the Rutgers genetic map (Matise et al. 2007).

Forty-nine SNPs were genotyped using the MassARRAY system from Sequenom, Inc. (Sequenom, San Diego, CA) (Buetow et al. 2001) in 796 AA subjects. Primer sequences are available on request. SNP rs872994 failed to genotype in the population. Genotyping

success rates among the other 48 SNPs were >90.5%. For quality controls purposes (QC) 19 samples were run in duplicate. Concordance among blind duplicates was 100%.

One microsatellite marker, D18S880, was also added to the study. D18S880 is in the *CNDP1* gene and has been shown to be associated with DN in European-derived populations (Janssen et al. 2005; Freedman et al. 2007). This marker was genotyped by fragment length analysis on an ABI Prism DNA Analyzer 3700 (Applied Biosystems, Inc., Foster City, CA) using a method similar to that previously described (Janssen et al. 2005). Fragment length was analyzed using ABI Prism GeneMapper software v3.0 (Applied Biosystems, Inc.). There were 19 duplicate samples run in order to ensure quality control. Of the replicate pairs, 100% were concordant.

Each pedigree was examined for consistency of familial relationships using PREST (Pedigree Relationship Statistical Test) (McPeck et al. 2000). When the self-reported familial relationships were inconsistent with that determined from the observed genotypic data for that pedigree, then 1) the pedigree was modified when the identity by descent (IBD) statistics suggested a very clear alternative, or 2) a minimal set of genotypic data was converted to missing. Each genetic marker was also examined for Mendelian inconsistencies using PedCheck (O'Connell et al. 1998), and sporadic problem genotypes were converted to missing.

**Case-control dense SNP mapping and gene analysis**—A dense SNP map was completed by the Center for Inherited Disease Research (CIDR), and was funded by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). SNPs were identified through SNP databases, HapMap resources, and Illumina design scores. Initially, all Yoruban (YRI) SNP genotype, allele frequency, and LD data from HapMap (public data release #21; includes phase I and II data) for SNPs in the fine mapped LOD-1 region (bld35: 48,263,676-65,774,799) were selected. SNPs were then filtered based on minor allele frequency (MAF), retaining only SNPs with  $MAF \geq 0.05$ . Next, the resulting SNP set was scored by Illumina using their proprietary assay designability scoring algorithm. In order to maximize probability of successful genotyping, the minimum threshold Illumina design score was set to 0.8. Tag SNPs were identified from these SNPs using an algorithm that selects an Illumina-designable SNP to tag each bin of SNPs (designable+non-designable) using a threshold LD score ( $r^2 \geq 0.8$ ). These SNPs were added to the final SNP list.

Next, haplotype tagging SNPs were identified by evaluating the HapMap genotype data using the Haploview program (Barrett et al. 2005) to generate the set of Gabriel-algorithm (Gabriel et al. 2002) defined haplotype blocks and compute haplotype frequencies for haplotypes of frequency 1% or greater. The *tagsnps* program (D.O. Stram, USC) was used to identify haplotype tagging SNPs (htSNPs) for these Gabriel-algorithm derived haplotype blocks in the region. *tagsnps* (beta version 2) generates htSNP tag sets using a greedy stepwise algorithm to reduce the block-tagging SNP sets until the minimum haplotype variance (Rh2) drops below a threshold (in this case 0.8). These SNPs were added to the final SNP list which was filtered for duplicate SNPs.

Residual SNP map slots were backfilled using designable bin tag SNPs identified by Perlegen in an independent study and using physical map coverage from HapMap data. A search was conducted for any adjacent inter-SNP region >20 kb in length in the SNP map and these regions were checked for published Perlegen tag SNPs with  $MAF \geq 0.05$  in AA populations (Perlegen ethnicity samples are distinct from HapMap samples). After merging the Perlegen tag SNPs with the map and uniqueness filtering, the map was checked for SNPs that were  $\leq 60$  bases apart since assayed SNPs closer than 60 bases cannot be included in the same Illumina multiplex oligo pool. Where necessary, SNPs were deleted from the map to

meet the 60 bp separation requirement, with deletion preference in the order: *physical map coverage SNP* > *Perlegen tag SNP* > *htSNP* > *HapMap tag SNP*. This step was followed by a final manual check and editing to arrive at 2×1536 SNPs (actually 2,962 genotyping SNPs with 40 SNPs reserved for QC purposes and 70 SNPs reserved for admixture adjustment purposes).

3,072 (2×1536) SNPs were genotyped in 1,029 AA DN cases and 1,027 AA healthy controls using Illumina's Custom Genotyping Service. Genotyping success rates were >98.3%. For quality controls purposes (QC) 48 samples were run in duplicate. Concordance among blind duplicates was >99.9%.

**Admixture Adjustment for case-control**—Seventy biallelic ancestral informative markers (AIMs) were genotyped in 1,029 AA case subjects, 1,027 AA control subjects, 44 Yoruba Nigerians, and 39 European Americans using Illumina's Custom Genotyping Service or the MassARRAY system from Sequenom, Inc as the protocol specified.

### Linkage analyses

Multipoint linkage analyses were carried out using NPL regression analyses using the  $NPL_{\text{pairs}}$  statistics outputted from a modified version of Genehunter (Langefeld et al. 1999; Langefeld et al. 2001; Davis et al. 2001). The NPL regression approach is a conditional logistic regression analysis in which the family-specific NPL statistic (*e.g.*,  $NPL_{\text{pairs}}$ ) at one or more loci is the predictor variable. This methodology was used in the initial genome scan and is described in detail (Bowden et al. 2004).

### Ordered subsets analysis

Ordered subset analyses (OSA) (Hauser et al. 2004) were calculated to investigate the influence of a pedigree's mean age of diagnosis of diabetes, mean age of diagnosis of ESRD, and mean duration of diabetes before diagnosis of ESRD (similar to NPL regression analysis above). OSA ranks each family by the family-level value of a covariate of interest and identifies the contiguous subset of families that maximize the evidence for linkage. In the OSA with the mean age at diabetes diagnosis, each pedigree was ranked from lowest to highest for age at diabetes diagnosis. The family with the lowest mean age at diabetes diagnosis entered into the analysis and the corresponding LOD score was computed on the target chromosome (*e.g.*, chromosome 18) for that family. Next, a second linkage analysis on the target chromosome 18 was computed combining the two families with the two lowest mean ages at diabetes diagnosis values. The *i*<sup>th</sup> OSA analysis proceeds by computing a linkage analysis on the target chromosome using the subset of families with the *i*<sup>th</sup> lowest mean ages at diabetes diagnosis. This process is repeated until all families have been added to the linkage analysis. The subset of families that yield the largest LOD score on the target chromosome is taken as the LOD score of interest. The location that maximizes the LOD score on a chromosome will vary as the subset of families analyzed changes. The statistical significance of the change in the LOD score was evaluated by a permutation test under the null hypothesis that the ranking of the covariate is independent of the family's LOD score on the target chromosome. Thus, the families were randomly permuted with respect to the covariate ranking and an analysis proceeded as above for each permutation of these data. The resulting empirical distribution of the change in the LOD scores yielded a chromosome-specific P-value. In this example, the family-level means were ranked in ascending order; however, we repeated the analysis ranking in descending order.

### Association analysis in the Dense SNP map

Tests for genotypic association were performed on each SNP individually using SNPADDMIX, a component of the SNPGBWA program (Harley et al. 2008) which includes the



capability to perform association calculations adjusting for covariates. Genotypic association reported here is for analyses incorporating adjustment for African ancestry proportions. The primary inference is based on the 2 degree of freedom global test of genotypic association. If significant, then the individual genetic models (dominant, additive and recessive) were examined for context. This is consistent with the Fisher's protected least significant difference (LSD) multiple comparisons procedure.

Ancestral allele frequencies were estimated from the results of the AIMs genotyped in the Yoruba Nigerians and the European Americans. Individual ancestral proportions were generated for each subject using FRAPPE (Tang et al. 2005), an EM algorithm, under a two-population model. The influence of other possible covariates: age, BMI and gender, on evidence of association was tested using SNPADDMIX. Two SNP haplotype analysis was completed using the program Dandelion ([www.phs.wfubmc.edu](http://www.phs.wfubmc.edu)). Linkage disequilibrium was calculated as defined by Gabriel (Gabriel et al. 2002) with the program Haploview (Barrett et al. 2005).

## Results

### Clinical and phenotypic data for African American DN individuals

The clinical and phenotypic characteristics for the genotyped samples, both the family samples for linkage analysis and the case-control samples for the dense SNP map, are summarized in Table 1. The genotyped samples have a higher proportion of females, probably reflecting the increased prevalence of T2DM among African American women (Center for Disease Control and Prevention 2008), survival, and participation bias. The cases in the family sample and the cases in the case-control sample have an overlap of 189 subjects, thus are not entirely independent. The two sets of cases are broadly comparable across all characteristics. While the age of enrollment is higher in the cases compared to the controls, note that the age of diagnosis of diabetes in the cases is lower than the age of recruitment for the non-diabetic controls. As we have reported previously (Bowden et al. 2004), the age at diabetes diagnosis and age at ESRD onset are strongly correlated. There were 53 DN subjects with age of onset of diabetes less than 25 years of age. The prevalence of type 1 diabetes in African Americans is low, and the average BMI at enrollment into the study of this early diabetes onset group was 31.4 (range 21.7-44.0) suggesting that many of these individuals do have T2DM. We cannot, however, exclude the possibility that some subjects have type 1 diabetes. The great majority of the ESRD affected subjects were enrolled within 5 years of developing ESRD. Overall the DN affected individuals were overweight or obese on average at the time of their enrollment in the study (median BMI 30.6 for linkage set and 29.6 for association sample). Controls for the association study were also overweight. On average, the diabetes affected individuals in the linkage set have poorer glucose control than cases in the association data set.

### Fine mapping analyses

Linkage mapping was performed on the LOD-1 region on chromosome 18q (18q21.1-23) that was previously identified (Bowden et al. 2004). This dataset encompassed 13 microsatellite markers (including D18S880) and 48 SNPs genotyped in the enlarged family population. The average spacing between microsatellite markers was 8.7cM. The average spacing between SNP markers was 0.97cM, and there were no interSNP gaps greater than 2.6cM. The average spacing across all markers in the LOD-1 region was 0.89cM. The marker list and the Rutgers genetic map locations (Matise et al. 2007) are detailed in Supplementary Table 1.

With the inclusion of additional subjects and additional markers from fine mapping to the original scan, a multipoint linkage analysis was completed on chromosome 18. The single locus, NPL regression analysis results are detailed in Table 2. When including all markers: microsatellites and SNPs (Micro+SNP; Combined), there was a reduction in the LOD score (LOD 0.36, 110.6 cM, near D18S1371; Table 2). However, when the analysis was completed using just the microsatellites alone (Micro Only), the LOD score remained at the same magnitude (LOD 1.08, 121.6 cM, near D18S1390; Table 2) that was shown previously (Bowden et al. 2004). These results are also shown in Figure 1 (Micro+SNP, black line; Micro Only, gray line).

Ordered subsets analysis (OSA), which assesses linkage under the premise that it can be uncovered more readily in subgroups of families within a population, distinguished by specific phenotypic traits, was also performed. This method was used to look for differential evidence of linkage based on age at diagnosis of ESRD and age at diagnosis of diabetes. The pedigree subsets that were identified from the OSA were also analyzed using NPL regression. These results are shown in Table 3 and depicted in Figure 1.

When subsetting by age at diagnosis of ESRD, both the microsatellite+SNP marker set and the microsatellite only marker set showed evidence of significant change in  $P$  value ( $\Delta P$ ) when looking at an optimal subset compared to the complete population (Table 3). The microsatellite+SNP marker set displayed an OSA maximum LOD score of 3.36 at 80.6cM using 34 pedigrees (15%) with an earlier age of diagnosis, compared to the entire sample LOD score of 0.04 ( $\Delta P = 0.0148$ , NPL  $P$  value = 0.00008; Table 3) (Yellow line; Figure 1). In the microsatellite only marker set, the OSA maximum LOD score was 3.34 at 84.1cM in the optimum subset of 31 pedigrees (14%), whereas the maximum LOD score in the whole population was 0.52 ( $\Delta P = 0.0222$ ; NPL  $P$  value = 0.00009; Table 3) (Red line; Figure 1).

The OSA was also performed subsetting by age at diagnosis of diabetes. Using the microsatellite only marker set, there were two subsets, early onset and optimal slice, that showed a significant  $P$  value for change. The early onset subset exhibited an OSA maximum LOD score of 3.18 at 90.1cM (Table 3). This was obtained using 136 pedigrees (61%) with the earliest age at diagnosis of diabetes. The entire sample LOD was 1.23 ( $\Delta P = 0.0168$ , NPL  $P$  value = 0.00013; Table 3) (Blue line; Figure 1). The optimal slice subset, which also had an earlier age of diabetes diagnosis compared to the remaining families, showed a max LOD score of 3.90 at 90.1cM using 111 pedigrees (50%) compared to a LOD score of 1.23 in the whole population ( $\Delta P = 0.0161$ , NPL  $P$  value = 0.00002; Table 3) (Green line; Figure 1). There were no statistically significant differences ( $\Delta P$ ) observed between the entire sample LOD score and the OSA maximized LOD score in the microsatellite+SNP marker set. Instead, the same subsets of pedigrees identified from the microsatellite only OSA analysis were used for analysis in the microsatellite+SNP marker set. The early onset subset displayed a trimodal peak, with OSA maximized LOD scores of 1.20, 1.82, and 2.12 at 73.1cM, 96.1cM, and 113.6cM respectively (NPL  $P$  values = 0.01854, 0.00383, and 0.0018 respectively; Table 3) (Pink line; Figure 1). This was done using the same 136 pedigrees that were identified in the microsatellite only early onset subset. The optimal slice subset also showed a trimodal peak in the microsatellite+SNP marker set. The OSA maximized LOD scores were 1.36, 2.47, and 1.95 at 74.6cM, 90.6cM, and 113.6cM respectively (NPL  $P$  values = 0.0123, 0.00075, and 0.00275 respectively; Table 3) (Maroon line; Figure 1). This was performed using the 111 pedigrees that were in the optimal slice subset in the microsatellite only marker set.

### Dense SNP map analysis

A dense SNP map was performed on the LOD-1 region, 18q21.1-18q22.2. There were 2,814 bi-allelic variants (95.0%) successfully genotyped in this region in 1,029 AA T2DM-ESRD

case subjects, and 1,027 AA non-diabetic, non-nephropathy controls. The average marker density was 6 kb with a range of 55 bp to 97 kb. Sixty-three SNPs (2.2%) were nominally inconsistent with Hardy Weinberg Equilibrium (HWE) proportions ( $P$ -value $<0.01$ ) in the combined population (Supplementary Table 2A). These SNPs were noted as inconsistent with HWE, but still included for analysis. Minor allele frequencies (MAF) are also shown in Supplementary Table 2A. Genotype frequencies and counts for each SNP are shown in Supplementary Table 2B. There were 55 SNPs (1.9%) with MAFs of less than 0.05 in both the case and control populations and 3 SNPs that were monomorphic in our AA case-control population (Supplementary Table 2A).

A plot of the admixture-adjusted genotypic results is shown in Figure 2. These results are further detailed in Supplementary Tables 3A (2DF and dominant genetic model) and 3B (additive and recessive genetic models). There were 342 SNPs (12.2%) that showed nominal evidence of association ( $P$ -value $<0.05$ ) with T2DM-ESRD under one or more tests of association. One-hundred and forty-six SNPs showed evidence of association under the 2 df test, 180 were associated under a dominant genetic model, 129 were associated under an additive genetic model, and 150 were associated under a recessive genetic model (Figure 2; Supplementary Tables 3A and 3B).

The 25 SNPs most associated under the 2 df test are shown in Table 4. SNPs were excluded from this list if they were out of HWE in the case, control, or combined population with a  $P$ -value $<0.001$ , or if they were associated under the recessive genetic model and had a minor allele homozygous genotype frequency of less than 0.10. Ten of the top 25 SNPs fell within a gene; 13 SNPs were within ~300 kb or less of a gene or hypothetical protein; and 2 SNPs were in intergenic regions (no genes within 500 kb). The 2 df  $P$ -values ranged from 0.001 – 0.0114 (Table 4). When looking at genetic models, 13 of the SNPs were most associated under a dominant genetic model ( $P=0.0003$ – $0.0210$ ; Table 4), 8 were most associated under an additive genetic model ( $P=0.0017$ – $0.0052$ ; Table 4), and 4 were most associated under a recessive genetic model ( $P=0.0015$ – $0.0050$ ; Table 4).

In addition to looking at the most associated SNPs, SNPs were also prioritized by genes and/or regions that contained multiple associated SNPs. These results are shown in Table 5. The region contains 97 known genes. After prioritization, there were 23 genes that contained one or more SNPs nominally associated ( $P=0.0006$ – $0.0441$ ; Table 5). There were also 12 genes with nearby SNPs (within 500 kb) associated ( $P=0.0025$ – $0.0390$ ; Table 5).

## Discussion

We have previously shown evidence of linkage to diabetic nephropathy in African American families on chromosome 18q21.1-23 (Bowden et al. 2004). This region has also been identified in genome wide linkage scans for diabetic nephropathy in other populations: Turkish, Pima Indian, European American, and American Indian families (Vardarli et al. 2002; Iyengar et al. 2007). Following studies have excluded the candidate gene *ZNF236*, located at 18q22-23, from playing a role in diabetic nephropathy (Halama et al. 2003), but have discovered that the carnosinase genes, *CNDP1* and *CNDP2*, located at 18q22.3, influence diabetic nephropathy susceptibility (Janssen et al. 2005; Freedman et al. 2007; McDonough et al. 2009). D18S880, a trinucleotide repeat in exon 2 of the *CNDP1* gene, which encodes for a leucine repeat in the signal peptide of the preprotein, was associated with diabetic nephropathy in a small European population (Janssen et al. 2005) and a European-American population (Freedman et al. 2007). The association was seen in individuals who were homozygous for the shortest allelic form of the repeat (5L-5L); this genotype was significantly more common in the absence of diabetic nephropathy (Janssen et al. 2005; Freedman et al. 2007). However, there was no association seen with D18S880 in African Americans (McDonough et al. 2009). There were associations



seen with other variants and haplotypes in the *CNDP2* gene in African Americans. These results were associated with both risk and protection, and some were observed at low frequencies in the population (McDonough et al. 2009). This suggests that there are other variants at the 18q21.1-18q23 locus that are contributing to DN risk in African Americans.

In order to refine the results of our previously published linkage peak (Bowden et al. 2004), we increased our AA family population from 206 DN-affected sibling pairs to 270. We also increased the marker density in the region by adding an additional microsatellite marker, D18S880, and 48 SNPs. We examined the data using multiple methods of analysis. First, we used NPL analysis with two different marker sets (Micro Only and Micro+SNP). We saw a reduction in the LOD score using the combined marker set, however the LOD score remained at the same magnitude we had previously published (Bowden et al. 2004) using just the microsatellite markers. In order to maximize our LOD scores, we performed OSA with age of diagnosis of ESRD and age of diagnosis of diabetes. This provided us with maximized LOD scores in pedigrees with earlier ages of onset of ESRD and earlier ages of onset of diabetes. Our significant maximized LOD scores ranged from 3.18 to 3.90 and from 80.6cM to 90.1cM.

We have looked at the data numerous ways in order to see if one model or one approach is better. Overall, we observe a wide range of results depending on the analysis method and the combination of markers used. This is consistent with the commonly accepted explanation of results from linkage mapping at this time: that there are multiple loci at the same region that are shared among and between families (Altshuler et al. 2008). These results are clearly complex; however, there is continued evidence for linkage in this region after increasing the number of families and the number of markers.

We further investigated this region in African Americans by performing a dense SNP map. The dense SNP map increased our coverage to an average marker density of 6kb. These results were evaluated in two ways. First, we ranked SNPs by order of magnitude of association. Ten of the top 25 SNPs were in a gene. Second, we looked at genes that contained multiple associated SNPs. Using this method, we compiled a list of 23 genes that contained one or more associated SNPs and 12 genes had one or more associated SNPs nearby (within 500kb).

After comparing the results between the two different SNP prioritization methods, two candidate genes stood out: *NEDD4L* and *SERPINB7*. *NEDD4L* is a ubiquitin ligase located at 18q21. *NEDD4L* has been previously shown to influence the regulation of renal sodium reabsorption (Manunta et al. 2008; Sile et al. 2008; Dunn et al. 2002) and has been associated with essential hypertension in Caucasians and African Americans (Russo et al. 2005), and in Han Chinese (Wen et al. 2008). The Nedd4L protein is expressed in the distal nephron (Araki et al. 2008; Umemura et al. 2006). *SERPINB7*, also known as *MESGIN*, is located at 18q21.33 (Scott et al. 1999), is predominantly expressed in human mesangial cells, and is up regulated in IgA nephropathy (Miyata et al. 1998). Two polymorphisms in the 3'UTR of *SERPINB7*, 2093C and 2180T, were associated with IgA nephropathy in a Chinese population (Li et al. 2004). These markers form the 2093C-2180T haplotype that was associated with a more severe form of IgA nephropathy in a Chinese population (Xia et al. 2006), and poor renal survival in Korean IgA nephropathy patients (Lim et al. 2008). Inagi et al. (2006) created a severe diabetic nephropathy mouse model by overexpressing *serpinb7*, *RAGE*, and *iNOS*. There is also evidence that *serpinb7* is unregulated in diabetic nephropathy, in turn inhibiting plasmin and MMP activities, leading to mesangial matrix accumulation (Ohtomo et al. 2008).

We have looked at both of these genes in reference to diabetic nephropathy in our African American population. As we saw the most significant results in intron 1 of *NEDD4L* in the

dense SNP map, we examined this gene from exon 1 to exon 3. We successfully genotyped 91 SNPs in this region; however, we only saw nominal association (P values = 0.025-0.0459) with five SNPs (data not shown). In addition, we successfully genotyped 26 SNPs across the *SERPINB7* gene. We saw association with 6 SNPs throughout the gene (data not shown); however, these results require further validation. While we did not examine the entire genic region of *NEDD4L*, these results suggest that *NEDD4L* does not play a major role in diabetic nephropathy susceptibility in African Americans. Mutations in *SERPINB7* may be involved in the mesangial matrix accumulation observed in diabetic nephropathy; however, this gene must be further investigated before any conclusion can be drawn on its relationship to diabetic nephropathy in African Americans.

Our study is not without limitations. One advantage of genome-wide linkage scans is the ability to identify regions where multiple, rare variants which lead to disease, are shared among and/or between families. These variants may not have been identified in our population based case-control dense SNP map follow-up study. In addition, due to our recruitment design, we are unable to distinguish if the linkage and associations we observed were with diabetes, nephropathy, or both.

Overall, we have performed a comprehensive evaluation of the 18q21-18q23 genomic region in African American. The results of the fine mapping performed in our African American T2DM-DN family population demonstrated continued evidence for linkage with the inclusion of additional subjects, and additional markers. The dense SNP map produced several candidate genes, including *NEDD4L* and *SERPINB7*, which warrant further investigation. Taken together, this suggests that there may be multiple loci in this region that affect diabetic nephropathy susceptibility in African Americans.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We wish to thank the patients, their relatives, and staff of the Southeastern Kidney Council, Inc./ESRD Network 6 for their participation. This work was supported by NIH grants RO1 DK066358 (DWB), R01 DK053591 (DWB), R01 HL56266 (BIF), R01 DK070941 (BIF) and in part by the General Clinical Research Center of the Wake Forest University School of Medicine grant M01 RR007122. Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University; contract number N01-HG-65403.

## References

- Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science*. 2008; 322:881–8. [PubMed: 18988837]
- Araki N, Umemura M, Miyagi Y, Yabana M, Miki Y, Tamura K, Uchino K, Aoki R, Goshima Y, Umemura S, Ishigami T. Expression, transcription, and possible antagonistic interaction of the human *Nedd4L* gene variant: implications for essential hypertension. *Hypertension*. 2008; 51:773–7. [PubMed: 18268134]
- Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005; 21:263–5. [PubMed: 15297300]
- Bowden DW. Genetics of kidney disease. *Kidney Int Suppl*. 2003:S8–12. [PubMed: 12864867]
- Bowden DW, Colicigno CJ, Langefeld CD, Sale MM, Williams A, Anderson PJ, Rich SS, Freedman BI. A genome scan for diabetic nephropathy in African Americans. *Kidney Int*. 2004; 66:1517–26. [PubMed: 15458446]
- Buetow KH, Edmonson M, MacDonald R, Clifford R, Yip P, Kelley J, Little DP, Strausberg R, Koester H, Cantor CR, Braun A. High-throughput development and characterization of a

genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc Natl Acad Sci U S A*. 2001; 98:581–4. [PubMed: 11136232]

Center for Disease Control and Prevention. Age-Adjusted Percentage of Civilian, Noninstitutionalized Population with Diagnosed Diabetes, by Race and Sex, United States, 1986-2006. 2008 [21 May 2009]. Available via CDC. <http://www.cdc.gov/diabetes/statistics/prev/national/figraceethsex.htm>

Davis CC, Brown WM, Lange EM, Rich SS, Langefeld CD. Nonparametric linkage regression. II: Identification of influential pedigrees in tests for linkage. *Genet Epidemiol*. 2001; 21 1:S123–9. [PubMed: 11793654]

Dunn DM, Ishigami T, Pankow J, von Niederhausern A, Alder J, Hunt SC, Leppert MF, Lalouel JM, Weiss RB. Common variant of human NEDD4L activates a cryptic splice site to form a frameshifted transcript. *J Hum Genet*. 2002; 47:665–76. [PubMed: 12522688]

Ewens KG, George RA, Sharma K, Ziyadeh FN, Spielman RS. Assessment of 115 candidate genes for diabetic nephropathy by transmission/disequilibrium test. *Diabetes*. 2005; 54:3305–18. [PubMed: 16249459]

Freedman BI, Hicks PJ, Sale MM, Pierson ED, Langefeld CD, Rich SS, Xu J, McDonough C, Janssen B, Yard B, van der Woude FJ, Bowden DW. A leucine repeat in the carnosinase gene CNDP1 is associated with diabetic end-stage renal disease in European Americans. *Nephrol Dial Transplant*. 2007; 22:1131–5. [PubMed: 17205963]

Freedman BI, Rich SS, Yu H, Roh BH, Bowden DW. Linkage heterogeneity of end-stage renal disease on human chromosome 10. *Kidney Int*. 2002; 62:770–4. [PubMed: 12164858]

Freedman BI, Tuttle AB, Spray BJ. Familial predisposition to nephropathy in African-Americans with non-insulin-dependent diabetes mellitus. *Am J Kidney Dis*. 1995; 25:710–3. [PubMed: 7747724]

Freedman BI, Yu H, Anderson PJ, Roh BH, Rich SS, Bowden DW. Genetic analysis of nitric oxide and endothelin in end-stage renal disease. *Nephrol Dial Transplant*. 2000; 15:1794–800. [PubMed: 11071967]

Freedman BI, Yu H, Spray BJ, Rich SS, Rothschild CB, Bowden DW. Genetic linkage analysis of growth factor loci and end-stage renal disease in African Americans. *Kidney Int*. 1997; 51:819–25. [PubMed: 9067916]

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. *Science*. 2002; 296:2225–9. [PubMed: 12029063]

Halama N, Yard-Breedijk A, Vardarli I, Akkoyun I, Yard B, Janssen B, van der Woude FJ. The Kruppel-like zinc-finger gene ZNF236 is alternatively spliced and excluded as susceptibility gene for diabetic nephropathy. *Genomics*. 2003; 82:406–11. [PubMed: 12906866]

Harley JB, Alarcón-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, Vyse TJ, Langefeld CD, Nath SK, Guthridge JM, Cobb BL, Mirel DB, Marion MC, Williams AH, Divers J, Wang W, Frank SG, Namjou B, Gabriel SB, Lee AT, Gregersen PK, Behrens TW, Taylor KE, Fernando M, Zidovetzki R, Gaffney PM, Edberg JC, Rioux JD, Ojwang JO, James JA, Merrill JT, Gilkeson GS, Seldin MF, Yin H, Baechler EC, Li QZ, Wakeland EK, Bruner GR, Kaufman KM, Kelly JA. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX1, KIAA1542 and other loci. *Nat Genet*. 2008; 40:204–10. [PubMed: 18204446]

Hauser ER, Watanabe RM, Duren WL, Bass MP, Langefeld CD, Boehnke M. Ordered subset analysis in genetic linkage mapping of complex traits. *Genet Epidemiol*. 2004; 27:53–63. [PubMed: 15185403]

Inagi R, Yamamoto Y, Nangaku M, Usuda N, Okamoto H, Kurokawa K, van Ypersele de Strihou C, Yamamoto H, Miyata T. A severe diabetic nephropathy model with early development of nodule-like lesions induced by megalin overexpression in RAGE/iNOS transgenic mice. *Diabetes*. 2006; 55:356–66. [PubMed: 16443768]

Iyengar SK, Abboud HE, Goddard KA, Saad MF, Adler SG, Arar NH, Bowden DW, Duggirala R, Elston RC, Hanson RL, Ipp E, Kao WH, Kimmel PL, Klag MJ, Knowler WC, Meoni LA, Nelson RG, Nicholas SB, Pahl MV, Parekh RS, Quade SR, Rich SS, Rotter JI, Scavini M, Schelling JR, Sedor JR, Sehgal AR, Shah VO, Smith MW, Taylor KD, Winkler CA, Zager PG, Freedman BI.

Genome-wide scans for diabetic nephropathy and albuminuria in multiethnic populations: the family investigation of nephropathy and diabetes (FIND). *Diabetes*. 2007; 56:1577–85. [PubMed: 17363742]

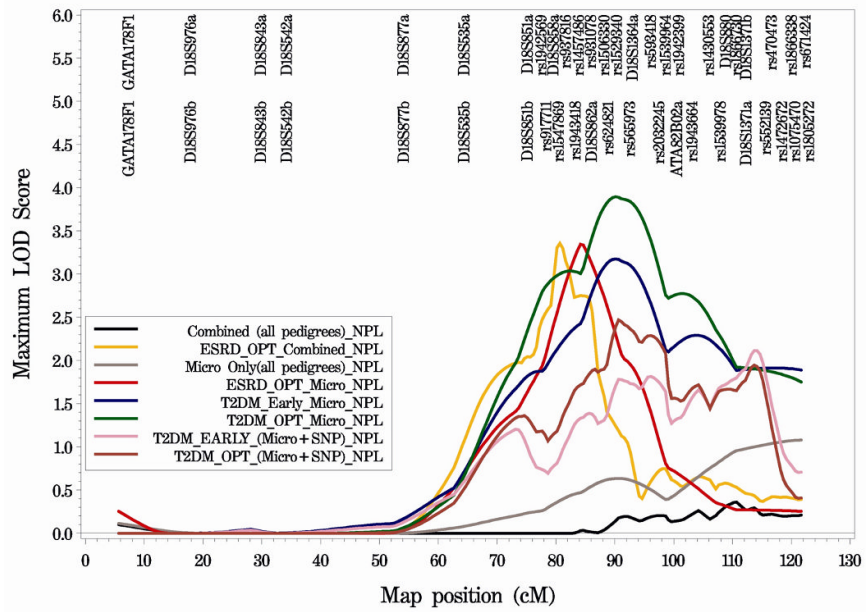
- Janssen B, Hohenadel D, Brinkkoetter P, Peters V, Rind N, Fischer C, Rychlik I, Cerna M, Romzova M, de Heer E, Baelde H, Bakker SJ, Zirie M, Rondeau E, Mathieson P, Saleem MA, Meyer J, Köppel H, Sauerhoefer S, Bartram CR, Nawroth P, Hammes HP, Yard BA, Zschocke J, van der Woude FJ. Carnosine as a protective factor in diabetic nephropathy: association with a leucine repeat of the carnosinase gene CNDP1. *Diabetes*. 2005; 54:2320–7. [PubMed: 16046297]
- Langefeld CD, Davis CC, Brown WM. Nonparametric linkage regression. I: Combined Caucasian CSGA and German genome scans for asthma. *Genet Epidemiol*. 2001; 21 1:S136–41. [PubMed: 11793656]
- Langefeld CD, Boehnke M. Multiple trait locus nonparametric linkage regression. *Am J Hum Genet*. 1999; 65:A1140.
- Li YJ, Du Y, Li CX, Guo H, Leung JC, Lam MF, Yang N, Huang F, Chen Y, Fang JQ, Maxwell PH, Lai KN, Wang Y. Family-based association study showing that immunoglobulin A nephropathy is associated with the polymorphisms 2093C and 2180T in the 3' untranslated region of the MEGSIN gene. *J Am Soc Nephrol*. 2004; 15:1739–43. [PubMed: 15213261]
- Lim CS, Kim SM, Oh YK, Joo KW, Kim YS, Han JS, Kim S. MEGSIN 2093T-2180C haplotype at the 3' untranslated region is associated with poor renal survival in Korean IgA nephropathy patients. *Clin Nephrol*. 2008; 70:101–9. [PubMed: 18793525]
- Manunta P, Lavery G, Lanzani C, Braund PS, Simonini M, Bodycote C, Zagato L, Delli Carpini S, Tantardini C, Brioni E, Bianchi G, Samani NJ. Physiological interaction between alpha-adducin and WNK1-NEDD4L pathways on sodium-related blood pressure regulation. *Hypertension*. 2008; 52:366–72. [PubMed: 18591455]
- Matisse TC, Chen F, Chen W, De La Vega FM, Hansen M, He C, Hyland FC, Kennedy GC, Kong X, Murray SS, Ziegler JS, Stewart WC, Buyske S. A second-generation combined linkage physical map of the human genome. *Genome Res*. 2007; 17:1783–6. [PubMed: 17989245]
- McDonough CW, Hicks PJ, Lu L, Langefeld CD, Freedman BI, Bowden DW. The influence of carnosinase gene polymorphisms on diabetic nephropathy risk in African-Americans. *Hum Genet Epub*. 2009 Apr 17.
- McPeck MS, Sun L. Statistical tests for detection of misspecified relationships by use of genome-screen data. *Am J Hum Genet*. 2000; 66:1076–94. [PubMed: 10712219]
- Miyata T, Nangaku M, Suzuki D, Inagi R, Uragami K, Sakai H, Okubo K, Kurokawa K. A mesangium-predominant gene, megsin, is a new serpin upregulated in IgA nephropathy. *J Clin Invest*. 1998; 102:828–36. [PubMed: 9710452]
- O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet*. 1998; 63:259–66. [PubMed: 9634505]
- Ohtomo S, Nangaku M, Izuhara Y, Yamada N, Dan T, Mori T, Ito S, van Ypersele de Strihou C, Miyata T. The role of megsin, a serine protease inhibitor, in diabetic mesangial matrix accumulation. *Kidney Int*. 2008; 74:768–74. [PubMed: 18580857]
- Russo CJ, Melista E, Cui J, DeStefano AL, Bakris GL, Manolis AJ, Gavras H, Baldwin CT. Association of NEDD4L ubiquitin ligase with essential hypertension. *Hypertension*. 2005; 46:488–91. [PubMed: 16103266]
- Scott FL, Eyre HJ, Lioumi M, Ragoussis J, Irving JA, Sutherland GA, Bird PI. Human ovalbumin serpin evolution: phylogenetic analysis, gene organization, and identification of new PI8-related genes suggest that two interchromosomal and several intrachromosomal duplications generated the gene clusters at 18q21-q23 and 6p25. *Genomics*. 1999; 62:490–9. [PubMed: 10644448]
- Sile S, Velez DR, Gillani NB, Alexander CA, George AL Jr, Williams SM. Haplotype diversity in four genes (CLCNKA, CLCNKB, BSND, NEDD4L) involved in renal salt reabsorption. *Hum Hered*. 2008; 65:33–46. [PubMed: 17652939]
- Spray BJ, Atassi NG, Tuttle AB, Freedman BE. Familial risk, age at onset, and cause of end-stage renal disease in white Americans. *J Am Soc Nephrol*. 1995; 5:1806–10. [PubMed: 7787148]
- Tang H, Peng J, Wang P, Risch NJ. Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol*. 2005; 28:289–301. [PubMed: 15712363]

- Umemura M, Ishigami T, Tamura K, Sakai M, Miyagi Y, Nagahama K, Aoki I, Uchino K, Rohrwasser A, Lalouel JM, Umemura S. Transcriptional diversity and expression of NEDD4L gene in distal nephron. *Biochem Biophys Res Commun.* 2006; 339:1129–37. [PubMed: 16338225]
- US Renal Data System. *USRDS 2008 Annual Data Report: Atlas of End-Stage Renal Disease in the United States.* National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; Bethesda, MD: 2008.
- Vardarli I, Baier LJ, Hanson RL, Akkoyun I, Fischer C, Rohmeiss P, Basci A, Bartram CR, Van Der Woude FJ, Janssen B. Gene for susceptibility to diabetic nephropathy in type 2 diabetes maps to 18q22.3-23. *Kidney Int.* 2002; 62:2176–83. [PubMed: 12427143]
- Wen H, Lin R, Jiao Y, Wang F, Wang S, Lu D, Qian J, Jin L, Wang X. Two polymorphisms in NEDD4L gene and essential hypertension in Chinese Hans - a population-based case-control study. *Clin Exp Hypertens.* 2008; 30:87–94. [PubMed: 18293164]
- Xia Y, Li Y, Du Y, Yang N, Li C, Leung JC, Lam MF, Huang W, Chen S, Maxwell PH, Lai KN, Wang Y. Association of MEGSIN 2093C-2180T haplotype at the 3' untranslated region with disease severity and progression of IgA nephropathy. *Nephrol Dial Transplant.* 2006; 21:1570–4. [PubMed: 16431886]
- Yu H, Anderson PJ, Freedman BI, Rich SS, Bowden DW. Genomic structure of the human plasma prekallikrein gene, identification of allelic variants, and analysis in end-stage renal disease. *Genomics.* 2000; 69:225–34. [PubMed: 11031105]
- Yu H, Bowden DW, Spray BJ, Rich SS, Freedman BI. Linkage analysis between loci in the renin-angiotensin axis and end-stage renal disease in African Americans. *J Am Soc Nephrol.* 1996; 7:2559–64. [PubMed: 8989734]
- Yu H, Bowden DW, Spray BJ, Rich SS, Freedman BI. Identification of human plasma kallikrein gene polymorphisms and evaluation of their role in end-stage renal disease. *Hypertension.* 1998; 31:906–11. [PubMed: 9535413]

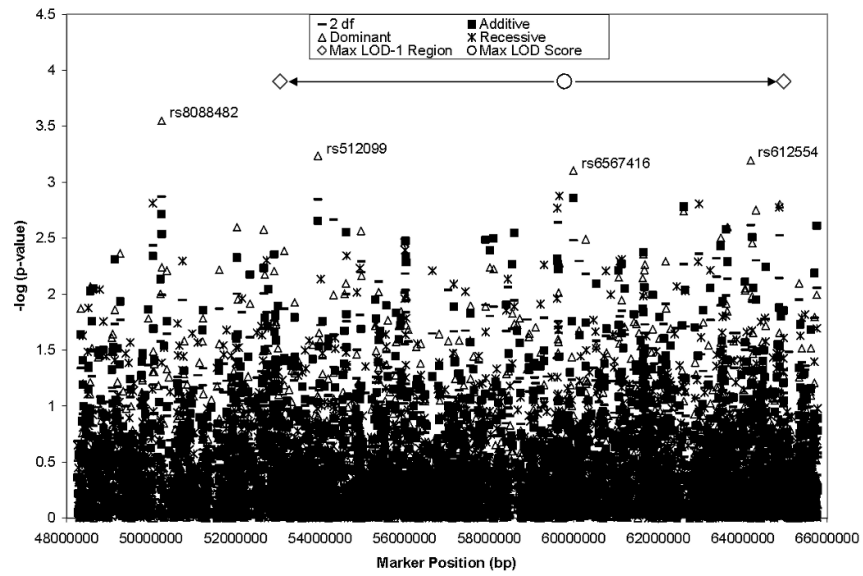
## Abbreviations

<b>cM</b>	centiMorgan
<b>ESRD</b>	end-stage renal disease
<b>LOD</b>	log of the odds
<b>DN</b>	diabetic nephropathy
<b>OSA</b>	ordered subsets analysis
<b>NPL</b>	nonparametric linkage





**Figure 1.** Linkage results. Micro Only: Only microsatellite markers were used in analysis. Combined/ Micro+SNP: All markers, microsatellites and SNPs, were used in analysis. OPT: An optimal subset of pedigrees, identified through OSA, was used in the analysis. Early: A subset of pedigrees with an earlier age at diagnosis of ESRD or diabetes (T2DM) was used in the analysis.



**Figure 2.** Dense SNP map results. Max LOD and LOD-1 region shown correspond to the age at diagnosis of diabetes, Microsatellites Only, optimal slice peak: Max LOD = 3.9, obtained through OSA.

Table 1

Characteristics of the study samples<sup>1</sup>

<i>Family Sample for Linkage Analysis</i>		<i>n=796</i>			
<b>Trait</b>	<b>Mean</b>	<b>SD</b>			
% female	66.10%	NA			
Age at diabetes diagnosis (years)	39	11.3			
Age at ESRD diagnosis (years)	55.9	10.9			
Diabetes to ESRD duration (years)	16.9	10.2			
Duration of ESRD at enrollment (years)	3.2	3.8			
Age at enrollment (years)	59.3	9.5			
BMI (kg/m <sup>2</sup> )	30.6	7.4			
HbA1c (%)	8.3	1.8			
<i>Case-Control Sample for Association Analysis</i>		<i>Cases (n=1029)</i>		<i>Control (n=1027)</i>	
<b>Trait</b>	<b>Mean</b>	<b>SD</b>	<b>Mean</b>	<b>SD</b>	
% female	60.70%	NA	53.40%	NA	
Age at diabetes diagnosis (years)	40.5	11.7	NA	NA	
Age at ESRD diagnosis (years)	57.7	10.9	NA	NA	
Diabetes to ESRD duration (years)	17	11.1	NA	NA	
Duration of ESRD at enrollment (years)	3.6	4.3	NA	NA	
Age at enrollment (years)	61.2	10.3	51.7	11.3	
BMI (kg/m <sup>2</sup> )	29.6	7.2	29.3	6.9	
HbA1c (%)	7.4	1.7	NA	NA	

<sup>1</sup>Total 796 in Family (189 are in Case, 607 are not in either Case or Control)

**Table 2**

NPL regression analysis of fine mapping data. Single locus analysis.

Marker Set	Maximum LOD Position cM	Nearest Marker	LOD	LOD-1 interval	P value
Microsatellites + SNPs	110.6	D18S1371	0.36	87.6-qter	0.1962
<b>Microsatellites Only</b>	<b>121.6</b>	<b>D18S1390</b>	<b>1.08</b>	<b>64.6-qter</b>	<b>0.0258</b>

Table 3

Ordered subsets analysis (OSA) of age at diagnosis of ESRD, and age at diagnosis of diabetes. Maximized LOD scores of  $\geq 3.0$  are shown in bold.

Marker Set	Linked subset	Position cM	Nearest Marker	Entire Sample LOD	Maximized LOD	LOD-1 Interval	Optimal Subset	Mean $\pm$ SD			Empirical P value for change	NPL P value*
								Remaining Families	% Pedigrees			
<i>Subsetting by age at diagnosis of ESRD</i>												
Micro+SNP	Optimal	80.6cM	rs721405	0.04	3.36	76.6-86.6	47.96 $\pm$ 2.22	57.61 $\pm$ 8.17	15	0.0148	0.00008	
Micro Only	Optimal	84.1cM	D18S862	0.52	3.34	79.1-90.1	47.85 $\pm$ 2.0	57.57 $\pm$ 8.16	14	0.0222	0.00009	
<i>Subsetting by age at diagnosis of diabetes</i>												
Micro+SNP	Early onset	73.1cM	D18S851	0	1.2	56.1-qter	33.88 $\pm$ 5.32	48.28 $\pm$ 5.52	61	NA	0.01854	
		96.1cM	rs2032245	0.19	1.82	80.1-119.6	33.88 $\pm$ 5.32	48.28 $\pm$ 5.52	61	NA	0.00383	
		113.6cM	rs552139	0.29	2.12	82.1-118.1	33.88 $\pm$ 5.32	48.28 $\pm$ 5.52	61	NA	0.0018	
Micro Only	Early onset	90.1cM	D18S1364	1.23	3.18	80.6-98.6	33.88 $\pm$ 5.32	48.28 $\pm$ 5.52	61	0.0168	0.00013	
Micro+SNP	Optimal	74.6cM	rs1145315	0	1.36	62.6-qter	35.68 $\pm$ 3.64	43.71 $\pm$ 10.57	50	NA	0.0123	
		90.6cM	rs565973	0.18	2.47	81.6-106.1	35.68 $\pm$ 3.64	43.71 $\pm$ 10.57	50	NA	0.00075	
		113.6cM	rs552139	0.29	1.95	68.1-117.6	35.68 $\pm$ 3.64	43.71 $\pm$ 10.57	50	NA	0.00275	
Micro Only	Optimal	90.1cM	D18S1364	1.23	3.9	78.6-98.1	35.68 $\pm$ 3.64	43.71 $\pm$ 10.57	50	0.0161	0.00002	

NA - No significant increase in the LOD score ( $\Delta P < 0.05$ )

\* NPL P value obtained by performing NPL analysis using the subset of pedigrees identified from OSA



Table 4

The 25 most associated SNPs. P-values  $\leq 0.05$  are shown in bold.

Rank	Marker	Location	Gene/Gene(s)/Nearby	2 DF	Genotypic Association					
					Dominant		Additive		Recessive	
					P-value	OR (95%CI)	P-value	OR (95%CI)	P-value	OR (95% CI)
1	rs8088482	50,246,989	84.6 kb from C18orf54	<b>0.0014</b>	<b>0.0003</b>	1.40 (1.17-1.68)	0.0019	1.23 (1.08-1.41)	0.3492	1.13 (0.87-1.48)
2	rs512099	53,946,964	NEDD4L	<b>0.0014</b>	<b>0.0006</b>	0.70 (0.57-0.86)	0.0022	0.75 (0.62-0.90)	0.6825	1.17 (0.55-2.49)
3	rs612554	64,190,078	None within 500kb	<b>0.0024</b>	<b>0.0006</b>	0.71 (0.58-0.86)	0.0006	0.73 (0.61-0.87)	0.1590	0.63 (0.33-1.20)
4	rs1431423*	64,870,409	CCDC102B	<b>0.0030</b>	<b>0.0016</b>	1.36 (1.12-1.64)	0.0441	1.14 (1.00-1.29)	0.8942	0.98 (0.79-1.23)
5	rs6567416	59,991,368	250.4 kb from BCO36306	<b>0.0033</b>	<b>0.0008</b>	1.36 (1.14-1.63)	0.0014	1.25 (1.09-1.44)	0.1552	1.25 (0.92-1.69)
6	rs488846	50,041,768	36.6 kb from MBD2; 8.1 kb from POLI	<b>0.0037</b>	<b>0.0496</b>	1.20 (1.00-1.43)	<b>0.0046</b>	1.24 (1.07-1.44)	<b>0.0015</b>	1.91 (1.28-2.84)
7	rs9951478	64,311,444	None within 500kb	<b>0.0050</b>	<b>0.0018</b>	0.75 (0.63-0.90)	0.0112	0.83 (0.72-0.96)	0.9467	0.99 (0.69-1.41)
8	rs1421536	60,129,202	112.6 kb from BCO36306	<b>0.0051</b>	<b>0.0210</b>	0.81 (0.68-0.97)	0.3206	0.94 (0.82-1.07)	0.1441	1.23 (0.93-1.61)
9	rs1720843	59,609,668	SERPINB7	<b>0.0054</b>	0.0826	0.85 (0.71-1.02)	<b>0.0048</b>	0.83 (0.73-0.94)	<b>0.0017</b>	0.65 (0.50-0.85)
10	rs11661057	62,600,331	178.1 kb from CDH19	<b>0.0054</b>	<b>0.0018</b>	1.33 (1.11-1.60)	<b>0.0017</b>	1.23 (1.08-1.40)	0.0674	1.27 (0.98-1.64)
11	rs1814878	54,966,238	SEC11C	<b>0.0064</b>	<b>0.0027</b>	0.76 (0.64-0.91)	0.0205	0.85 (0.74-0.97)	0.9398	1.01 (0.73-1.41)
12	rs8093918	60,284,087	42.2 kb from BCO36306	<b>0.0066</b>	<b>0.0032</b>	1.31 (1.10-1.58)	0.0181	1.20 (1.03-1.41)	0.7286	0.93 (0.60-1.43)
13	rs4021701	64,860,552	CCDC102B	<b>0.0072</b>	0.2796	0.90 (0.74-1.09)	<b>0.0132</b>	0.85 (0.75-0.97)	<b>0.0017</b>	0.69 (0.55-0.87)
14	rs7237658	63,637,491	302.5 kb from DSEL	<b>0.0073</b>	<b>0.0025</b>	0.70 (0.56-0.88)	0.0051	0.73 (0.59-0.91)	0.7525	1.19 (0.40-3.59)
15	rs4275943	65,748,829	CD226	<b>0.0088</b>	<b>0.0101</b>	1.27 (1.06-1.53)	<b>0.0024</b>	1.26 (1.09-1.47)	<b>0.0202</b>	1.55 (1.07-2.25)
16	rs1942872	56,019,345	170.1 kb from MC4R	<b>0.0092</b>	<b>0.0108</b>	1.30 (1.06-1.59)	<b>0.0033</b>	1.31 (1.09-1.57)	<b>0.0230</b>	2.13 (1.11-4.09)
17	rs2576415	52,667,378	WDR7	<b>0.0099</b>	<b>0.0026</b>	1.35 (1.11-1.64)	<b>0.0058</b>	1.20 (1.05-1.36)	0.1678	1.17 (0.94-1.45)
18	rs1789575	52,031,303	21.8 kb from FLJ45743	<b>0.0099</b>	<b>0.0025</b>	1.32 (1.10-1.58)	0.0047	1.21 (1.06-1.38)	0.2100	1.19 (0.91-1.56)
19	rs2874116	54,610,579	42.2 kb from MALT1	<b>0.0100</b>	<b>0.0080</b>	1.30 (1.07-1.57)	<b>0.0028</b>	1.29 (1.09-1.52)	<b>0.0332</b>	1.74 (1.05-2.91)
20	rs9947912	50,250,470	88 kb from C18orf54	<b>0.0102</b>	<b>0.0058</b>	1.33 (1.09-1.63)	<b>0.0029</b>	1.22 (1.07-1.38)	<b>0.0411</b>	1.25 (1.01-1.55)
21	rs9966483	63,601,326	266.3 kb from DSEL	<b>0.0103</b>	<b>0.0121</b>	1.26 (1.05-1.50)	<b>0.0026</b>	1.22 (1.07-1.39)	<b>0.0156</b>	1.40 (1.07-1.83)
22	rs1011371	65,332,722	DOK6	<b>0.0112</b>	<b>0.0080</b>	0.78 (0.65-0.94)	0.0457	0.86 (0.73-1.00)	0.5311	1.14 (0.75-1.74)
23	rs2000835	50,741,329	CCDC68	<b>0.0113</b>	0.9297	1.01 (0.84-1.21)	0.1484	0.91 (0.80-1.03)	<b>0.0050</b>	0.71 (0.56-0.90)

Rank	Marker	Location	Gene/Gene(s) Nearby	Genotypic Association					
				Dominant		Additive		Recessive	
				P-value	OR (95%CI)	P-value	OR (95%CI)	P-value	OR (95% CI)
24	rs12970134	56,035,730	153.8 kb from MC4R	<b>0.0113</b>	1.28 (1.04-1.56)	<b>0.0052</b>	1.29 (1.08-1.55)	<b>0.0164</b>	2.25 (1.16-4.37)
25	rs9958800	58,591,780	PHLPP	<b>0.0114</b>	1.27 (1.04-1.55)	<b>0.0029</b>	1.22 (1.07-1.38)	<b>0.0129</b>	1.32 (1.06-1.65)

\* Out of HWE in the Controls (P=0.002)

^ Indicates Lack of Fit to an Additive Genetic Model

Table 5

Candidate Genes identified from the dense SNP map.

Gene	SNPs Associated				SNP Most Associated within Gene			
	Name	Size (kb)	Number	Percent (%)	Nearby (1-500kb)	Marker	P-value	Model
NEDD4L	353.6	6	8.0	1 (40.1kb)	rs512099	0.0006	dom	
POLJ	25.2	3	75.0	1 (8.1kb)	rs488846	0.0015	rec	
CCDC102B	339.9	6	10.2	4 (11.7-87.6kb)	rs1431423	0.0016	dom	
SERPINB7	30.0	1	9.1	4 (9.6-29.3kb)	rs1720843	0.0017	rec	
CD226	94.0	2	9.5	0	rs4275943	0.0025	add	
SEC11C	18.9	1	14.3	3 (0.7-97.3kb)	rs1814878	0.0027	dom	
MALTI	78.8	1	10.0	4 (3.4-55.9kb)	rs2874116*	0.0028	add	
PHLPP	264.9	4	19.0	5 (37.5-129.3kb)	rs9958800	0.0029	add	
PIGN	142.8	1	9.1	0	rs633490	0.0033	add	
CDH7	130.7	9	27.3	9 (141.7-483.6kb)	rs7232924	0.0043	add	
DCC	1190.1	17	11.2	0	rs1393330	0.0043	dom	
CCDC68	57.8	2	25.0	1 (6.6kb)	rs2000835	0.0050	rec	
WDR7	378.4	7	12.7	7 (50.8-107.8kb)	rs2576415	0.0058	add	
DOK6	441.1	4	4.3	3 (258.1-334.1kb)	rs1011371	0.0080	dom	
CCBE1	263.4	5	7.6	2 (47.1-87.7kb)	rs3097388	0.0095	dom	
ALPK2	147.7	3	13.0	0	rs12954360	0.0102	dom	
FECH	38.5	3	37.5	2 (43.2-49.9kb)	rs8093153	0.0118	dom	
TCF4	360.5	3	6.8	0	rs17089789	0.0139	add	
BCL2	196.1	5	11.9	1 (50.8kb)	rs3810031	0.0198	dom	
TXNDC10	41.4	1	10.0	0	rs167443	0.0318	add	
SERPINB5	28.1	2	20.0	0	rs6567351	0.0319	add	
VPS4B	33.3	1	14.3	0	rs6567341	0.0440	add	
ATP8B1	83.5	1	10.0	0	rs317845	0.0441	add	
DSEL	10.2	0	NA	12 (42.1-452.4kb)	rs7237658*	0.0025	dom	
TXNLI	35.8	0	NA	7 (80.5-399.2kb)	rs1789575*	0.0025	dom	
ZNF532	123.6	0	NA	6 (56-133.8kb)	rs2874116*	0.0028	add	

Gene	SNPs Associated				SNP Most Associated within Gene			
	Name	Size (kb)	Number	Percent (%)	Nearby (1-500kb)	Marker	P-value	Model
PMAIP1	4.3	0	NA	NA	12 (62.1-362.7kb)	rs1942872*	0.0033	add
MC4R	1.0	0	NA	NA	11 (104.3-404.9kb)	rs1942872*	0.0033	add
ST8SIA3	16.4	0	NA	NA	3 (13.6kb-168.2kb)	rs1230706*	0.0041	dom
CDH20	64.6	0	NA	NA	2 (178.7kb-184kb)	rs9956796*	0.0051	all
FAM44C	1.5	0	NA	NA	10 (9.4-190.2kb)	rs4800961*	0.0062	dom
GRP	10.6	0	NA	NA	1 (60.5kb)	rs2125517*	0.0069	dom
ZCCHC2	55.2	0	NA	NA	4 (0.5-91.1kb)	rs435439*	0.0073	rec
RNF152	78.0	0	NA	NA	1 (5.8kb)	rs9966069*	0.0359	dom
ONECUT2	55.6	0	NA	NA	2 (7-13.7kb)	rs576678*	0.0390	add

NA - No SNPs associated within the gene, thus % SNPs Associated could not be calculated

\* Nearby gene (Not in)