Published in final edited form as:

Hum Genet. 2008 August ; 124(1): 63–71. doi:10.1007/s00439-008-0523-7.

Evaluation of a SNP map of 6q24-27 confirms diabetic nephropathy loci and identifies novel associations type 2 diabetes patients enriched with nephropathy from an African American population

Tennille S. Leak, Ph.D.¹, Josyf C. Mychaleckyj, D.Phil.^{2,4,5}, Shelly G. Smith, M.S.¹, Keith L. Keene, Ph.D.¹, Candace J. Gordon, M.S.¹, Pamela J. Hicks, B.S.^{1,3}, Barry I. Freedman, M.D.^{1,2}, Donald W. Bowden, Ph.D.^{1,2,3}, and Michèle M. Sale, Ph.D.^{1,2,4,6,7}

¹Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, NC

²Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC

³Department of Biochemistry, Wake Forest University School of Medicine, Winston-Salem, NC

⁴Center for Public Health Genomics, University of Virginia, Charlottesville, VA

⁵Department of Public Health Sciences, University of Virginia, Charlottesville, VA

⁶Department of Medicine, University of Virginia, Charlottesville, VA

⁷Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA

Abstract

Previously we performed a genome scan for type 2 diabetes (T2DM) using 638 African-American (AA) affected sibling pairs from 247 families; non-parametric linkage analysis suggested evidence of linkage at 6q24-27 (LOD 2.26). To comprehensively evaluate this region we performed a 2-stage association study by first constructing a SNP map of 754 SNPs selected from HapMap on the basis of linkage disequilibrium (LD) in 300 AAT2DM-ESRD subjects, 311 AA controls, 43 European American controls and 45 Yoruba Nigerian samples (Set 1). Replication analyses were conducted in an independent population of 283 AA T2DM-ESRD subjects and 282 AA controls (Set 2). In addition, we adjusted for the impact of admixture on association results by using ancestry informative markers (AIMs). In Stage 1, 137 (18.2%) SNPs showed nominal evidence of association ($P < 0.05$) in one or more of tests of association: allelic ($n=33$), dominant ($n=36$), additive ($n=29$), or recessive ($n=34$) genotypic models, and 2- ($n=47$) and 3-SNP ($n=43$) haplotypic analyses. These SNPs were selected for follow-up genotyping. Stage 2 analyses confirmed association with a predicted 2-SNP “risk” haplotype in the *PARK2* gene. Also, two intergenic SNPs showed consistent genotypic association with T2DM-ESRD: rs12197043 and rs4897081. Combined analysis of all subjects from both stages revealed nominal associations with 17 SNPs within genes; including suggestive associations in *ESR1* and *PARK2*. This study confirms known diabetic nephropathy loci and identifies potentially novel susceptibility variants located within 6q24-27 in AA.

Keywords

Diabetes; linkage disequilibrium; African American

Introduction

Diabetes is a growing epidemic in the United States and worldwide. It is currently estimated that 20.8 million people, or 7% of the total US population, have diabetes, of which 6.2 million (nearly one-third) are undiagnosed. In 2005 there were 1.5 million new cases of diabetes in Americans aged 20 years and older (National Diabetes Information Clearinghouse, 2005). Furthermore, AA have increased rates of many diabetes related complications compared with EA of similar age (Carter et al. 1996; Emanuele et al. 2005; Pugh et al. 1995). Although a genetic component for T2DM has been established and several genome-wide associations have recently been published in EA, only a limited number of genetic variants have shown association with T2DM in AA populations including: insulin promoter factor (*IPF1*) (Karim et al. 2005; Silver and Shetty 2002), betacellulin gene (Elbein et al. 2006), forkhead box O1A (*FOXO1A*) (Karim et al. 2006), serum retinol binding protein 4 (*RBP4*) (Craig et al. 2007) and transcription factor 7-Like 2 (*TCF7L2*) (Sale et al. 2007).

We conducted a genome wide linkage scan in AA families with T2DM which produced evidence for linkage on chromosome 6q24-27 (163.5 cM, LOD 2.26) (Sale et al. 2004). Other linkage studies provide supporting evidence for T2DM susceptibility variants in this region of 6q with measures of insulin resistance, obesity and leptin concentrations in Mexican Americans (Duggirala et al. 2001) and possible paternally derived T2DM alleles in Pima Indians (Lindsay et al. 2001).

Our prior positional candidate gene studies of this region identified significant associations with T2DM for variants in *ESR1* (Gallagher et al. 2007a) and μ -opioid receptor (*OPRM1*) (Gallagher et al. 2006) genes in an AA population, although these did not appear likely to account for the linkage signal. We also investigated three additional positional candidate genes under the linkage peak: the tubby superfamily proteins (*TUSP*), mitogen-activated protein kinase kinase kinase 4 (*MAP3K4*) and superoxide dismutase 2 (*SOD2*), however, significant associations were not observed (unpublished data).

These findings led us to further investigate loci in the 14 Mb region LOD-1 interval on chromosome 6q24-27. The purpose of this study was to investigate this region comprehensively by constructing a SNP map, and identify variants that confer susceptibility to T2DM and/or diabetic end-stage renal disease (T2DM-ESRD) in AA.

Research Design and Methods

Subjects

Two independently recruited groups of T2DM-ESRD patients and control subjects were evaluated. The first group consisted of 294 AA individuals with T2DM-ESRD and 311 AA non-diabetic controls (Set 1). Replication studies were performed on a population of 283 additional cases with T2DM-ESRD and 282 AA non-diabetic controls (Set 2). Set 2 included T2DM-ESRD probands from the original AA genome-wide linkage scan (Sale et al. 2004). This study was conducted under Institutional Review Board approval from Wake Forest University School of Medicine, and adhered to the tenets of the Declaration of Helsinki. Recruitment of AA and European American (EA) patients and controls has been described previously (Sale et al. 2004). Briefly, 577 unrelated AA patients with T2DM, born in North Carolina, South Carolina, Georgia, Tennessee or Virginia, were recruited from dialysis facilities. Probable type 1 diabetes patients were excluded on the basis of a history of diabetic ketoacidosis, or development of diabetes mellitus prior to age 25 and continuous insulin therapy since diagnosis. A diagnosis of T2DM was based upon participants reporting an initial diagnosis of diabetes mellitus after age 35 years, receiving oral hypoglycemic agents or dietary

therapy without insulin for at least one year after initial diagnosis, and active treatment with diabetes medications. Cases had T2DM diagnosed at least 5 years prior to initiating renal replacement therapy, background or greater diabetic retinopathy and/or > 3+ proteinuria on urinalysis in the absence of other causes of nephropathy. Five hundred ninety-three AA controls and 36 EA controls born in the same five states, without a current diagnosis of T2DM or renal disease, were recruited from community sources. DNA extraction was performed using the PureGene system (Gentra Systems, Minneapolis, MN). DNA was also obtained from 44 Yoruba Nigerians (YRI) from the National Institute of General Medical Sciences Human Variation Collection (Coriell Cell Repositories, Camden, NJ).

SNP Selection and Genotyping

Initially, a minimal number of SNPs that distinguished the two most common haplotypes (based on HapMap Yoruban (YRI) data, Release #22) were selected within each of the following candidate genes: low density lipoprotein receptor-related protein 11 (*LRP11*, 3 SNPs); regulator of G-protein signaling 17 (*RGS17*, 5 SNPs); insulin-like growth factor 2 receptor (*IGF2R*, 6 SNPs); peptidylprolyl isomerase (cyclophilin)-like 4 (*PPIL4*, 1 SNP); A-kinase anchor protein 12 (*AKAP12*, 4 SNPs); vasoactive intestinal peptide isoform 1 (*VIP*, 1 SNP); t-complex 1 (*TCPI*, 2 SNPs); solute carrier family 22 members 1, 2 and 3 (*SLC22A1*, 1 SNP; *SLC22A2*, 2 SNPs; *SLC22A3*, 3 SNPs) and *ESR1* (6 SNPs).

Additional SNPs were selected based on a SNP bin tagging approach as described in Carlson et al. (2004) using a threshold LD score of $r^2 \geq 0.8$. All of the YRI genotype, allele frequency, and LD data from HapMap (public data release #22, includes phase 2 data) for SNPs in the genome interval (bld36: 149,134,903-163,348,348) were selected. This yielded 5,112 possible interval SNPs with detailed HapMap data. The SNPs were then filtered, retaining only SNPs with minor allele frequency (MAF) ≥ 0.05 , leaving 3,912 possible SNPs. The SNPs in these sets were scored by Illumina using their proprietary assay designability scoring algorithm. To maximize probability of successful genotyping, the minimum threshold Illumina design score was set to 1.0. The algorithm used selected an Illumina-designable SNP to tag each bin of SNPs (capturing both Illumina-designable and non-designable SNPs). This resulted in 734 intervals of mean length 19.3 kb (total map length 14.2 Mb), including 612 bins containing 2 or more SNPs tagged by a single Illumina-designable SNP, but with 171 physical intervals uncaptured or non-overlapping any tag bin. To expand the relative map coverage in areas of low LD, the 612 tagSNPs were pruned to 522 based on physical map and allele frequency proximity. If two tag SNPs were closer than 19.3 kb and had a MAF difference < 0.1 , one of the pair was removed. Remaining SNPs were selected to provide physical coverage in extended map regions of low LD devoid of bins; yielding a total of 768 SNPs.

Genotyping was performed by Illumina Inc.'s Custom Genotyping Service (San Diego, CA). Seven SNPs in Set 2 were genotyped using iPLEX methodology on a MassARRAY genotyping system (Sequenom Inc., San Diego, CA) (Buetow et al. 2001) due to failed genotyping on the Illumina Platform: rs12197043, rs1498637, rs4870523, rs681442, rs6940976, rs7746991 and rs9397033. Genotyping success rates were >95.9% in cases and >96.7% in controls. For quality control (QC) purposes, each plate contained two duplicate samples and 4 inter-plate controls. Concordance between duplicate samples was >99.9%.

Genotyping for admixture analyses

Seventy biallelic Admixture Informative Markers (AIMs), selected to maximize European and African allele frequency differences and sample all non-acrocentric arms of the autosomal genome, were genotyped by Illumina Inc.'s Custom Genotyping Service (San Diego, CA) in 577 AA cases, 596 AA controls, 44 YRI and 39 EA. Primer sequences are available on request. Genotyping success rates for AIMs were >97.4% in AA cases and >95.6% in AA controls.

Statistical Methods

All SNPs were tested for departures from Hardy Weinberg equilibrium (HWE) using a 1 degree of freedom (df) chi-squared test at a significance level of <0.01 .

The pairwise linkage disequilibrium statistic D' and the haplotype block definition of Gabriel et al. (2002) were calculated using Haploview 3.2 (Barrett et al. 2005). D' statistics were visualized and plotted using LDViewer (J. Mychaleckyj and M. Zhong, personal communication).

All SNPs were tested for unadjusted genotypic association in the case-control study under the following models: Cochran-Armitage additive trend (1 df, scores 0,1,2), dominant model (1 df, scores 0,1,1), and recessive model (1 df, scores 0,0,1).

Dandelion, an expectation-maximization (EM) algorithm-based program for haplotype analysis, was used to test for association of haplotypes within the case-control populations. Estimates of case and control haplotype frequencies were obtained using Dandelion, and empirical P-values calculated using a permutation test of the multinomial likelihood ratio statistic initially using 1,000 permutations (Green et al. 2001). A P-value <0.05 was considered statistically significant.

Admixture adjustment of association results

Individual ancestral proportions were estimated using an EM algorithm (FRAPPE) (Tang et al. 2005) under a two population model. Logistic regression tests of genetic (dominant, additive and recessive) models of association included adjustments for individual estimates of African ancestry implemented in the program SNPADMIX (http://www.phs.wfubmc.edu/web/public_bios/sec_gene/downloads.cfm).

Correction for multiple testing

To correct for multiple tests at the gene level we evaluated the association results using a permutation analysis (10000 replicates) to simulate the null joint distribution of marker statistics, and applied MinP step down analysis on the ranked, adjusted P-values (Westfall and Young, 1993). This procedure adjusts for multiple tests of SNPs in the map while maintaining strong control over the Experiment-wise Error Rate (EER) at any desired level of significance.

Results

Population characteristics

Characteristics of the AA case and control populations are shown in Table 1. Controls were significantly younger than cases ($P<0.0001$), although they were significantly older than the mean age at T2DM diagnosis ($P<0.0001$) for both population sets. Age data were unavailable for a significant proportion (25%) of controls recruited during the early phase of the study. BMI data are not presented since measures reflected weight on dialysis in cases and were unavailable for the majority of controls. Similar proportions of females were present among cases (61%) and controls (62%) in Set 2. However, in Set 1 a higher proportion of females (61%) were observed in cases than controls (57.8%).

Estimates of ancestry

FRAPPE (Tang et al. 2005) analysis indicated that the mean $\pm$ SD proportion of African ancestry in Set 1 cases was 0.79 ± 0.16 , and 0.79 ± 0.12 in Set 1 controls ($P=0.57$). In Set 2, the mean African ancestry was 0.84 ± 0.10 in cases and 0.79 ± 0.14 in controls ($P<0.0001$).

Initial association analyses

A total of 754 bi-allelic variants (98.2%) covering a 14Mb region on human chromosome 6q24-27 were successfully genotyped in Set 1. There are 111 known genes in the 14 Mb region of the map. The SNP map had an average SNP density of 1 SNP every 20 kb, with the largest gap 93kb and the smallest 131bp. Sixteen (2.1%) and 13 (1.7%) SNPs were inconsistent with Hardy-Weinberg proportions ($P < 0.01$) in cases and controls respectively. Two SNPs were inconsistent with HWE in both cases and controls: rs9456394 (cases: $P = 0.0002$; controls: $P = 0.00004$), and rs9459056 (cases: $P = 2 \times 10^{-8}$; controls: $P = 5.3 \times 10^{-6}$) located in the synaptotagmin 2 (*SYN2*) gene. Both SNPs are inconsistent with HWE due to an excess of homozygotes, but were retained for exploratory analyses; however there was no evidence for association of these SNPs with T2DM-ESRD. Genotypic frequencies and HWE P-values are shown in Supplementary Table 1.

The LD structure present in the cases and controls was compared by evaluating the total number of blocks and average block size (kb) between cases and controls. Using the Confidence Interval method of block definition (Gabriel et al. 2002) implemented in Haploview (with default settings); there were 106 blocks in cases (average block length 15 kb) and 111 LD blocks in controls (average block size of 14.7kb) across the 14 Mb region.

We note that, although initial linkage analyses showed greater evidence of linkage at 6q in families concordant for diabetes than in the smaller subset concordant for T2DM-ESRD (Sale et al. 2004), the ascertainment scheme of our case-control study does not allow us to distinguish between associations with T2DM and/or diabetic ESRD. While recognizing this as a limitation, both diseases represent a considerable and increasing public health burden, and identification of novel loci for either disease may provide valuable insights. A plot of the genotypic association results for single SNP analyses is shown in Figure 1, and detailed results are presented in Supplementary Table 2. One hundred thirty-seven SNPs (18.2%) showed nominal evidence of association ($P < 0.05$) in one or more of tests of association: allelic ($n = 33$), dominant ($n = 36$), additive ($n = 29$), or recessive ($n = 34$) genotypic models, 2- ($n = 47$) and 3-SNP ($n = 43$) haplotypic analyses. These nominally associated SNPs were subjected to further genotyping in a replication set of AA T2DM-ESRD case-control subjects (Set 2). Strongest associations ($P < 0.005$) in Set 1 were seen with 3 SNPs (2.2%): SNP rs10499266 (dominant $P = 0.001$) located in the A-kinase anchor protein 12 (*AKAP12*), and two SNPs within intergenic regions, rs4870523 (dominant $P = 3.3 \times 10^{-5}$, additive $P = 8.9 \times 10^{-5}$) and rs4897081 (additive $P = 0.004$, recessive $P = 0.0002$).

Association results from the second round of genotyping

A total of 137 SNPs were successfully genotyped in 283 T2DM-ESRD cases and 282 AA controls in the second round of genotyping (Set 2). All SNPs were examined for departures from HWE ($P < 0.01$) assumptions. Two (1.4%) and 3 (2.2%) SNPs deviated from HWE in cases and controls respectively. Marker rs791190 was inconsistent with HWE in cases ($P = 0.0002$) and controls ($P = 0.0008$); however there was no evidence of association between this SNP and T2DM-ESRD.

Thirty eight (27.7%) SNPs showed nominal evidence of association ($P < 0.05$) in one or more tests of association: allelic ($n = 11$), dominant ($n = 13$), additive ($n = 11$), or recessive ($n = 15$) genotypic models, or 2- ($n = 8$) and 3-SNP ($n = 12$) haplotypic analyses. Genotype frequencies and counts for each SNP are shown in Supplementary Table 3. Single-SNP genotypic association results are presented in Figure 2 and Supplementary Table 4. We compared Set 1 results against Set 2 and considered association confirmed where they met the following criteria: (1) $P < 0.05$ for both sets; (2) associations in the same direction, with similar MAF and odds ratios for each case-control group. Based on this approach we observed two SNPs (Table

2) and one 2-SNP haplotype associated with T2DM-ESRD (Table 3). Genotypic associations were seen with rs12197043 (Set 1: dominant $P=0.025$, admixture-adjusted $P_a=0.034$; Set 2 $P=0.020$, $P_a=0.014$) and rs4897081 (recessive $P=0.0002$, $P_a=0.0001$; $P=0.022$, $P_a=0.035$ in Set 1 and Set 2 respectively) (Table 2). These SNPs are located in an intergenic region on chromosome 6q25.1 within a pseudogene of the *SNRPB2* (small nuclear ribonucleoprotein polypeptide B-double prime; accession number AL603766.4).

A 2-SNP haplotype containing markers rs9458276 and rs11755155, located in the *PARK2* (Parkin) gene, was also consistently associated with T2DM-ESRD (Table 3). Haplotype analysis predicted an uncommon “risk” 2-SNP TA haplotype present at 5.7% in cases, 2.6% in controls (OR 2.26, $P=0.05$) in Set 1, and 6.3% in cases, 2.4% in controls in Set 2 (OR 2.73, $P=0.02$).

Analyses of combined association results

Given the modest power to detect association using each set individually, we also performed association analyses on the combined Set 1 + Set 2 genotypic data. A total of 9 (6.6%) and 3 (2.2%) variants were inconsistent with Hardy-Weinberg proportions ($P<0.01$) in cases and controls respectively. Intergenic SNP rs791190 ($P=1\times 10^{-4}$ in cases; $P=3\times 10^{-6}$ in controls) deviated from HWE, due to deflation in heterozygosity.

We detected a total of 37 (26.8 %) SNPs nominally associated ($P<0.05$) with T2DM-ESRD in one or more test of association (Figure 3). Seventeen of these variants are located within known genes: parkin (*PARK2*; $n=4$), regulator of G protein signaling 17 (*RGS17*; $n=1$), zinc finger DHHC domain containing 14 (*ZDHHC14*; $n=4$), LOC202459 (protein product similar to RIKEN cDNA 2310008M10, $n=1$), C6orf97 (chromosome 6 open reading frame 97; $n=1$), estrogen receptor α (*ESR1*; $n=2$), μ -opioid receptor (*OPRM1*; $n=1$), phosphoinositide-binding protein 3-E (*PIP3E*; $n=1$), radial spokehead like 2 (*RSHL2*; $n=1$) and synaptojanin-2 (*SYNJ2*; $n=1$). Single-SNP genotypic association results are presented in Supplementary Table 5.

Correction for multiple testing

For the additive class of tests, SNPs rs4870523 (adjusted $P=0.048$) was significant at EER $P<0.05$, and the same SNP was significant at EER $P<0.05$ for the recessive class also (adjusted $P=0.015$). The second best recessive model SNP by p-value, rs4879081, was marginally significant at EER $P=0.087$. No SNPs were significant at this EER alpha-level for the dominant class. Under a composite null hypothesis of no association at a SNP under any model, implemented using a minimum P statistic ($\min \{ P_{\text{add}}, P_{\text{dom}}, P_{\text{rec}} \}$), SNP rs4870523 maintained significance at EER $P<0.05$, with composite null map adjusted P value of 0.042.

Discussion

We have carried out a detailed screen for T2DM-ESRD susceptibility variants in the 6q24-27 region using a SNP map of 754 SNPs that were selected to capture the majority of the common variation in an AA population.

The data revealed positive associations with two intergenic SNPs rs12197043 and rs4897081 that appear to have risk (OR 1.47) and protective (OR 0.56) effects, respectively (Table 2). The MAF allele for rs12197043 in the AA cases (0.41) is intermediate between the HapMap YRI (0.37) and HapMap CEU controls (0.45), while the frequency in AA controls (0.36) is lower. The mechanism for the association between the rs12197043 minor “A” allele and T2DM susceptibility is unknown. The more common “protective” G allele appears to be ancestral since it is present in the chimpanzee sequence. The MAF of the rs4897081 “G” allele in AA

controls (0.49) is comparable with the HapMapYRI (0.47) and HapMap CEU controls (0.56) but the frequency of the AA cases was lower (0.43).

These two SNPs lie within a 14 kb LD block, approximately 45.7 kb from uronyl 2 sulfotransferase (*UST*) and 244 kb from mitogen-activated protein kinase kinase 7 interacting protein 2 (*MAP3K7IP2*) genes. *UST* is involved in transferring sulfate to the 2-position of iduronyl residues in dermatan sulfate and glucuronyl residues in chondroitin sulfate (Kobayashi et al. 1999). In T2DM subjects, increased glycosaminoglycans (GAG) excretion has been detected (Juretic et al. 2002). The *MAP3K7IP2* protein is involved in nuclear factor (NF)- κ B regulation (Takaesu et al. 2000), and SNP rs237025 (M55V) in *MAP3K7IP2* has been shown to be associated with T2DM and diabetic nephropathy (Lin et al. 2007; Noso et al. 2007). SNPs rs12197043 and rs4897081 are approximately 270kb from rs237025. Using HapMap YRI genotype data (HapMap #21, on NCBI B35 assembly), the replicated SNPs fall within two blocks of high LD. Marker rs4897081 is contained within a block of 5 kb containing 14 HapMap SNPs in addition SNP rs12197043 is contained within a 2kb block consisting of 3 HapMap SNPs. It appears unlikely that these SNPs are in LD with rs237025. In addition, two human expressed sequence tags (EST) have been identified in the interval between the intergenic markers: ESTs AV7 17717 and AV739439. These ESTs overlap with *UST* and *MAP3K7IP2* coding regions and may represent alternative isoforms.

We have also identified a 2-SNP “risk” TA haplotype located in *PARK2* that showed consistent association between datasets (Table 3). The seventy genotyped *PARK2* SNPs fall within 15 LD blocks (data not shown). SNPs rs9458276 and rs117551155 are approximately 235kb apart and not contained within defined LD blocks. Parkin is an E3 ubiquitin ligase that has been found to be responsible for an autosomal recessive form of Parkinson disease (Kitada et al. 1998; Shimizu et al. 2000). Single SNP association analysis revealed that rs9458276 (combined $P=0.007$; $P_a=0.007$), and rs1175515 (combined $P=0.038$; $P_a=0.038$) are both nominally associated with T2DM-ESRD. The “TA” risk haplotype was not observed in the African or EA pseudo-ancestral samples (Table 3). It therefore appears that this haplotype containing the African derived rs9458276 “T” allele with the common rs11755155 minor “A” allele enhances risk for T2DM-ESRD. Functional studies have shown that culturing rat islets in high glucose concentrations (16.7mmol/l) results in decreased *PARK2* expression and the formation of amyloid deposits (Lopez-Avalos et al. 2006). Also, recent studies have identified associations with T2DM and augmented risk of Parkinson’s disease (Hu et al. 2007), providing some support for common elements in their molecular etiology.

In the combined analysis of all subjects we detected a total of 37 (29.1%) SNPs nominally associated with T2DM-ESRD. The most significant association was detected with intergenic SNP rs4897081 (recessive $P=0.00001$; $P_a=0.00001$). Additionally, associations were detected with *ESR1* SNPs rs3020411 (dominant $P=0.002$, $P_a=0.008$; additive $P=0.003$, $P_a=0.016$) and rs9340954 (dominant $P=0.001$, $P_a=0.004$; additive $P=0.019$, $P_a=0.064$), and *OPRM1* SNP rs607759 (dominant $P=0.028$, $P_a=0.052$). In candidate gene studies of the same population, significant findings had previously been observed with *ESR1* and *OPRM1* genes.

For the combined sample size of 577 cases and 593 controls, we estimate that this study has 62.9% power to detect an OR of 1.30 based on a MAF of 0.05 with a significance level of 0.05 (Dupont and Plummer 1990), and assuming the associated SNP is in complete LD ($r^2=1$) with the causative SNP.

In the combined analysis, SNP rs685449 in *RGS17* showed nominal significance (dominant $P=0.039$; $P_a=0.051$, additive $P=0.029$; $P_a=0.028$). Interestingly, *RGS17* is involved in the desensitization of *OPRM1* (Rodriguez-Munoz et al. 2007) and, as noted, significant association in the *OPRM1* gene was observed previously in our AA population (Gallagher et al. 2006) and

again in the current study. Interaction analyses of *RGS17* SNP rs685449 and *OPRM1* SNPs rs2075572 (recessive $P=0.035$ for the interaction term; data not shown) and rs497976 (recessive $P=0.023$ for interaction; data not shown), genotyped in previous analyses Gallagher et al. (2006), provides some support for epistasis between these genes.

Association analysis in the remaining high priority genes *AKAP12*, *ESR1*, *TCP1*, *IGF2R*, *SLC22A1*, *SLC22A2* and *SLC22A3* did not reach significance in Set 2 (P -values range from 0.072-0.980; Supplementary Table 4). However coverage of these genes was not comprehensive, so a role in diabetic nephropathy risk cannot be excluded.

Several genome wide association studies (GWAS) performed in European populations have replicated associations with T2DM (Saxena et al. 2007; Scott et al. 2007; Sladek et al. 2007). The reported association on chromosome 6q with rs2673776 in *SYNE1* (Scott et al. 2007) was not investigated in our population, although we genotyped a total of 27 SNPs in *SYNE1*, and imputation analysis indicates that SNP rs719764 was not significantly associated with T2DM-ESRD in our AA population (P -values ranging from 0.20-0.78 in all tests of association; Supplementary Tables 2 and 4). The MAF of the less common “G” allele in the AA cases (0.41) and AA controls (0.39) was comparable to the reported frequency in the Finnish population (Scott et al. 2007). Only three *SYNE1* SNPs (rs10499267, rs7762904 and rs214997) were genotyped in the second round of genotyping; however none reached significance in Set 2 or in the combined subjects (Supplementary Tables 4 and 5). These observations suggest that polymorphisms in this gene may not be involved in T2DM etiology in this population, or we failed to genotype SNPs in sufficient LD with the casual variant.

In an attempt to avoid false associations due to admixture, individual ancestral proportions were calculated using the program FRAPPE (Tang et al. 2005) and covariate adjustments using these estimates were performed for all SNPs. The most significantly associated SNPs (rs12197043, rs4897081, rs9340954, rs685449, rs9458276, and rs11755155) remained significant in at least one model of association in the combined set after adjustment; only rs607759 in the *OPRM1* gene was no longer significant following admixture-adjustment (Supplementary Table 5).

A limitation of this study was that the majority of controls were not clinically evaluated for diabetes, thus the control sample likely contained small numbers of cases with undiagnosed T2DM. While this has not impacted our ability to detect positive associations, this is likely to have reduced our power to detect more subtle influences of additional variants, and led to underestimation of OR. As mentioned previously, our ascertainment scheme does not allow us to determine whether the associated genetic variants predispose separately to T2DM, or ESRD as a sequela of T2DM, or both stages of the T2DM-ESRD disease process.

As discussed, only one SNP, rs4870523, maintained map-level significance ($P = 0.042$). The MAF in AA cases (0.38) are comparable with the HapMap YRI (0.36) and HapMap CEU controls (0.38), but contrast with those of the AA controls (0.33). SNP rs4870523 is located in an intergenic region on chromosome 6q25.1 within a pseudogene of the *AMACR* (Alpha-methylacyl-CoA racemase; accession number AL451061.8).

The observed associations across the 14 Mb region on chromosome 6q are few, which implies that genetic susceptibility in this region may be heterogeneous or that greater SNP density may be required. This may also be the result of multiple rare variants that were unable to be detected based on our HapMap based SNP selection approach. The results indicate that this region warrants further examination to investigate the biological relevance of variations in this region. Denser SNP genotyping and direct sequencing should be a priority.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank patients and staff of the Southeastern Kidney Council/ESRD Network 6 and individuals recruited as controls for their participation, Joyce Byers, Mitzie Spainhour, Matt Stiegert, and Mark Hansen and colleagues at Illumina Inc. This work was supported by grants DK066358, DK072550, Wake Forest University General Clinical Research Center M01 RR07122, and a Career Development Award from the American Diabetes Association (MMS).

References

- Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–5. [PubMed: 15297300]
- 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]
- Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;74:106–20. [PubMed: 14681826]
- Carter JS, Pugh JA, Monterrosa A. Non-insulin-dependent diabetes mellitus in minorities in the United States. *Ann Intern Med* 1996;125:221–32. [PubMed: 8686981]
- Craig RL, Chu WS, Elbein SC. Retinol binding protein 4 as a candidate gene for type 2 diabetes and prediabetic intermediate traits. *Mol Genet Metab* 2007;90:338–44. [PubMed: 17174134]
- Duggirala R, Blangero J, Almasy L, Arya R, Dyer TD, Williams KL, Leach RJ, O'Connell P, Stern MP. A major locus for fasting insulin concentrations and insulin resistance on chromosome 6q with strong pleiotropic effects on obesity-related phenotypes in nondiabetic Mexican Americans. *Am J Hum Genet* 2001;68:1149–64. [PubMed: 11283790]
- Dupont WD, Plummer WD Jr. Power and sample size calculations. A review and computer program. *Control Clin Trials* 1990;11:116–28. [PubMed: 2161310]
- Elbein SC, Wang X, Karim MA, Chu WS, Silver KD. Analysis of coding variants in the betacellulin gene in type 2 diabetes and insulin secretion in African American subjects. *BMC Med Genet* 2006;7:62. [PubMed: 16869959]
- Emanuele N, Sacks J, Klein R, Reda D, Anderson R, Duckworth W, Abairra C. Ethnicity, race, and baseline retinopathy correlates in the veterans affairs diabetes trial. *Diabetes Care* 2005;28:1954–8. [PubMed: 16043738]
- 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]
- Gallagher CJ, Gordon CJ, Langefeld CD, Mychaleckyj JC, Freedman BI, Rich SS, Bowden DW, Sale MM. Association of the mu-opioid receptor gene with type 2 diabetes mellitus in an African American population. *Mol Genet Metab* 2006;87:54–60. [PubMed: 16140553]
- Gallagher CJ, Keene KL, Mychaleckyj JC, Langefeld CD, Hirschhorn JN, Henderson BE, Gordon CJ, Freedman BI, Rich SS, Bowden DW, Sale MM. Investigation of the estrogen receptor-alpha gene with type 2 diabetes and/or nephropathy in African-American and European-American populations. *Diabetes* 2007a;56:675–84. [PubMed: 17327435]
- Gallagher CJ, Langefeld CD, Gordon CJ, Campbell JK, Mychaleckyj JC, Bryer-Ash M, Rich SS, Bowden DW, Sale MM. Association of the estrogen receptor-alpha gene with the metabolic syndrome and its component traits in African-American families: the Insulin Resistance Atherosclerosis Family Study. *Diabetes* 2007b;56:2135–41. [PubMed: 17513703]

- Green LE, Lange EM, Langefeld CD. Power comparisons of phase-known versus phase-unknown haplotype analyses for case-control designs. *Am. J. Hum. Genet* 2001;69:1948a.
- Guo D, Li M, Zhang Y, Yang P, Eckenrode S, Hopkins D, Zheng W, Purohit S, Podolsky RH, Muir A, Wang J, Dong Z, Brusko T, Atkinson M, Pozzilli P, Zeidler A, Raffel LJ, Jacob CO, Park Y, Serrano-Rios M, Larrad MT, Zhang Z, Garchon HJ, Bach JF, Rotter JI, She JX, Wang CY. A functional variant of SUMO4, a new I kappa B alpha modifier, is associated with type 1 diabetes. *Nat Genet* 2004;36:837–41. [PubMed: 15247916]
- Hu G, Jousilahti P, Bidel S, Antikainen R, Tuomilehto J. Type 2 diabetes and the risk of Parkinson's disease. *Diabetes Care* 2007;30:842–7. [PubMed: 17251276]
- Juretic D, Krajnovic V, Lukac-Bajalo J. Altered distribution of urinary glycosaminoglycans in diabetic subjects. *Acta Diabetol* 2002;39:123–8. [PubMed: 12357296]
- Karim MA, Craig RL, Wang X, Hale TC, Elbein SC. Analysis of FOXO1A as a candidate gene for type 2 diabetes. *Mol Genet Metab* 2006;88:171–7. [PubMed: 16497530]
- Karim MA, Wang X, Hale TC, Elbein SC. Insulin Promoter Factor 1 variation is associated with type 2 diabetes in African Americans. *BMC Med Genet* 2005;6:37. [PubMed: 16229747]
- Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, Yokochi M, Mizuno Y, Shimizu N. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* 1998;392:605–8. [PubMed: 9560156]
- Kobayashi M, Sugumaran G, Liu J, Shworak NW, Silbert JE, Rosenberg RD. Molecular cloning and characterization of a human uronyl 2-sulfotransferase that sulfates iduronyl and glucuronyl residues in dermatan/chondroitin sulfate. *J Biol Chem* 1999;274:10474–80. [PubMed: 10187838]
- Lin HY, Wang CL, Hsiao PJ, Lu YC, Chen SY, Lin KD, Hsin SC, Hsieh MC, Shin SJ. SUMO4 M55V variant is associated with diabetic nephropathy in type 2 diabetes. *Diabetes* 2007;56:1177–80. [PubMed: 17229939]
- Lindsay RS, Kobes S, Knowler WC, Bennett PH, Hanson RL. Genome-wide linkage analysis assessing parent-of-origin effects in the inheritance of type 2 diabetes and BMI in Pima Indians. *Diabetes* 2001;50:2850–7. [PubMed: 11723070]
- Lopez-Avalos MD, Duvivier-Kali VF, Xu G, Bonner-Weir S, Sharma A, Weir GC. Evidence for a role of the ubiquitin-proteasome pathway in pancreatic islets. *Diabetes* 2006;55:1223–31. [PubMed: 16644676]
- National Institute of Diabetes and Digestive and Kidney Diseases. National Diabetes Statistics fact sheet: general information and national estimates on diabetes in the United States, 2005. U.S. Department of Health and Human Services, National Institute of Health; Bethesda, MD: 2005.
- Noso S, Fujisawa T, Kawabata Y, Asano K, Hiromine Y, Fukai A, Ogihara T, Ikegami H. Association of small ubiquitin-like modifier 4 (SUMO4) variant, located in IDDM5 locus, with type 2 diabetes in the Japanese population. *J Clin Endocrinol Metab* 2007;92:2358–62. [PubMed: 17374705]
- Pugh JA, Medina RA, Cornell JC, Basu S. NIDDM is the major cause of diabetic end-stage renal disease. More evidence from a tri-ethnic community. *Diabetes* 1995;44:1375–80. [PubMed: 7589841]
- Rodriguez-Munoz M, Bermudez D, Sanchez-Blazquez P, Garzon J. Sumoylated RGS-Rz proteins act as scaffolds for Mu-opioid receptors and G-protein complexes in mouse brain. *Neuropsychopharmacology* 2007;32:842–50. [PubMed: 16900103]
- Sale MM, Freedman BI, Langefeld CD, Williams AH, Hicks PJ, Colicigno CJ, Beck SR, Brown WM, Rich SS, Bowden DW. A genome-wide scan for type 2 diabetes in African-American families reveals evidence for a locus on chromosome 6q. *Diabetes* 2004;53:830–7. [PubMed: 14988270]
- Sale MM, Smith SG, Mychaleckyj JC, Keene KL, Langefeld CD, Leak TS, Hicks PJ, Bowden DW, Rich SS, Freedman BI. Variants of the Transcription Factor 7-Like 2 (TCF7L2) Gene are Associated with Type 2 Diabetes in an African American Population Enriched for Nephropathy. *Diabetes* 2007;56:2638–42. [PubMed: 17601994]
- Saxena R, Voight BF, Lyssenko V, Burt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J,

- Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007;316:1331–6. [PubMed: 17463246]
- Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007;316:1341–5. [PubMed: 17463248]
- Shimizu N, Asakawa S, Minoshima S, Kitada T, Hattori N, Matsumine H, Yokochi M, Yamamura Y, Mizuno Y. PARKIN as a pathogenic gene for autosomal recessive juvenile parkinsonism. *J Neural Transm* 2000;(Suppl):19–30.
- Silver K, Shetty A. IPF-1 gene variation and the development of type 2 diabetes. *Mol Genet Metab* 2002;75:287–9. [PubMed: 11914043]
- Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007;445:881–5. [PubMed: 17293876]
- Takaesu G, Kishida S, Hiyama A, Yamaguchi K, Shibuya H, Irie K, Ninomiya-Tsuji J, Matsumoto K. TAB2, a novel adaptor protein, mediates activation of TAK1 MAPKKK by linking TAK1 to TRAF6 in the IL-1 signal transduction pathway. *Mol Cell* 2000;5:649–58. [PubMed: 10882101]
- 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]
- Westfall, PH.; Young, SS. Resampling-based multiple testing: Examples and methods for p-value adjustment. John Wiley & Sons Inc.; New York: 1993.

Figure 1. Single SNP genotypic association results for 754 SNPs in Set 1

X axis: Marker positions in kb; Y-axis: $-\log_{10}(\text{global P-value})$. The $P < 0.05$ nominal level of significance is indicated by the dashed line.

Figure 2. Single SNP genotypic association results for 137 SNPs in Set 2

X axis: Marker positions in kb; Y-axis: $-\log_{10}(\text{global P-value})$. The $P < 0.05$ nominal level of significance is indicated by the dashed line. Replicated SNPs (rs12197043 and rs4897081) are represented by enlarged data points.

Figure 3. Single SNP genotypic association results for 137 SNPs in Set 1 and Set 2 combined
 X axis: Marker positions in kb; Y-axis: $-\log_{10}(\text{global P-value})$. The $P < 0.05$ nominal level of significance is indicated by the dashed line. Replicated SNPs (rs12197043 and rs4897081) are represented by enlarged data points.

Table 1
Characteristics of African American study subjects

Trait	Set 1 Cases Mean±SD	Set 2 Cases Mean±SD	T2DM-ESRD Cases (Set 1+ Set 2)	Set 1 Controls Mean±SD	Set 2 Controls Mean±SD	Controls (Set 1 + Set 2) Mean±SD
Female (% (n))	61.2% (175/291)	61% (177/286)	61% (352/577)	57.8% (174/301)	61% (129/288)	51% (306/589)
Age at exam (years, Mean ±SD (n*))	64.4±10.1 (265)	60.1±10.1 (274)	62.2±10.3 (539)	49.3±9.9 (253)	49.1±9.6 (195)	49.3±9.8 (448)
Age at T2DM diagnosis (years, Mean±SD (n*))	43.3±11.9 (287)	40.1±11.1 (255)	41.8±11.6 (542)	-	-	-
Age at ESRD diagnosis (years, Mean±SD (n*))	60.8±10.3 (287)	57.2±10.4 (271)	59.0±10.5 (558)	-	-	-

Table 2
Single SNP genotypic association results for SNPs in the intergenic region between *UST* and *MAP37IP2/SUMO4* showing replicated association with T2DM-ESRD

Marker	Major/Minor alleles	Set	Dominant P-value	OR (95% CI)	Additive P-value	OR (95% CI)	Recessive P-value	OR (95% CI)
rs4897081	A/G	1	0.278	0.83 (0.59-1.17)	0.004	0.72 (0.58-0.90)	0.0002	0.46 (0.31-1.70)
		2	0.580	0.90 (0.62-1.30)	0.088	0.82 (0.65-1.03)	0.022	0.63 (0.42-0.94)
		Combined (Set 1 & 2)	0.316	0.88 (0.68-1.13)	0.003	0.78 (0.66-0.92)	0.00001	0.56 (0.42-0.74)
rs12197043	A/G	1	0.025	1.46 (1.05-2.03)	0.051	1.25 (0.97-1.55)	0.455	1.18 (0.76-1.82)
		2	0.020	1.51 (1.07-2.13)	0.059	1.27 (0.99-1.63)	0.643	1.10 (0.67- 1.81)
		Combined (Set 1 & 2)	0.002	1.47 (1.16-1.87)	0.011	1.25 (1.05-1.48)	0.539	1.11 (0.80-1.54)

Test models refer to the minor allele. **Bold:** P-values <0.05.

Table 3
Predicted haplotype for *PARK2* gene containing rs9458276 and rs1175515

Haplotype	Frequency T2DM-ESRD cases	Frequency Controls	P-Value	OR (95% CI)	Frequency European Americans	Frequency Africans
Set 1						
AA	0.455	0.448	0.866	1.03 (0.82-1.29)	0.577	0.454
AC	0.451	0.500	0.226	0.82 (0.65-1.03)	0.423	0.455
TA	0.057	0.026	0.050	2.26 (1.24-4.13)	-	-
TC	0.037	0.026	0.439	1.43 (0.75-2.72)	-	0.091
Set 2						
AA	0.475	0.468	0.86	1.03 (0.82-1.30)		
AC	0.446	0.468	0.601	0.92 (0.73-1.16)		
TA	0.064	0.024	0.020	2.73 (1.46-5.10)		
TC	0.015	0.041	0.065	0.38 (0.17-0.82)		
Combined						
AA	0.465	0.457	0.788	1.03 (0.88-1.21)		
AC	0.448	0.485	0.209	0.86 (0.73-1.02)		
TA	0.060	0.025	0.003	2.47 (1.60-3.82)		
TC	0.027	0.033	0.529	0.81 (0.50-1.30)		