Kidney International, Vol. 66 (2004), pp. 1517–1526

CLINICAL NEPHROLOGY – EPIDEMIOLOGY – CLINICAL TRIALS

A genome scan for diabetic nephropathy in African Americans

DONALD W. BOWDEN, CARLA J. COLICIGNO, CARL D. LANGEFELD, MICHÈLE M. SALE, ADRIENNE WILLIAMS, PAMELA J. ANDERSON, STEPHEN S. RICH, and BARRY I. FREEDMAN

Department of Biochemistry, Department of Internal Medicine, Department of Public Health Sciences, and the Center for Human Genomics, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, North Carolina

A genome scan for diabetic nephropathy in African Americans.

Background. There is substantial evidence for a genetic contribution to diabetic nephropathy susceptibility in the African American population, but little is known about location or identity of susceptibility genes.

Methods. DNA samples were collected from 206 type 2 diabetes (T2DM) and end-stage renal disease (ESRD)/nephropathy-affected sib pairs from 166 African American families (355 affected individuals). A genome scan was performed and data analyzed using nonparametric linkage regression (NPLR) analysis and ordered subsets analysis (OSA) methods.

Results. In initial NPLR analyses no logarithm of odds (LOD) scores >2.0 were observed. Four loci had LOD scores ≥1.0, with LOD = 1.43 at 29 cM on chromosome 7p the highest. NPLR analyses of multilocus interactions detected 6 loci (7p, 12p, 14q, 16p, 18q, and 21q) with LOD scores 1.15 to 1.63. NPLR analyses evaluating phenotypic interactions revealed multiple locations with evidence ($P < 0.05$) for interactions with age-at-onset of ESRD (9 loci), duration of diabetes before onset of ESRD (19 loci), and age-at-onset of diabetes (14 loci). Several loci identified by NPLR analyses were also identified using OSA. OSA revealed evidence for a nephropathy locus at 135 cM on chromosome 3 in an estimated 29% of the families (LOD = 4.55 in the optimal subset). Additional linkage evidence, LOD = 3.59, was observed on chromosome 7p (37% of the families, longer duration of diabetes prior to diagnosis of ESRD), and 18q (max. LOD = 3.72; 64% of the families, early diabetes diagnosis). The 7p linkage has been observed in a recent genome scan of African American type 2 diabetes.

Conclusion. This first genome scan of diabetic nephropathy in African Americans reveals evidence for susceptibility loci on chromosomes 3q, 7p, and 18q. The 7p locus may represent a type 2 diabetes susceptibility locus.

African Americans have a 2.6- to 5.6-fold increased risk of developing end-stage renal disease (ESRD) compared to other racial/ethnic groups in the United States [1], and make up almost half of the individuals on renal replacement therapy. Diabetes-associated nephropathy is the most common source of ESRD and accounts for approximately 40% of cases in the United States. There is now extensive evidence, based upon a variety of approaches [2], that ESRD in the general population, and African Americans specifically, has a genetic component. African Americans have a significantly stronger familial aggregation of ESRD: a close relative with ESRD gives an African American a 9-fold increased risk of developing ESRD [3] compared to an increased risk of only 2.7-fold in Caucasian Americans [4]. These results suggest that there is a stronger familial component to ESRD in the African American population, in addition to a disproportionately higher incidence. The impact on public health is substantial, but the origins of ESRD in the general population are still poorly understood. Only some diabetes-affected individuals will ultimately progress to nephropathy and ESRD, suggesting that, in addition to genetic susceptibility, environmental factors contribute to nephropathy risk.

Many studies have been carried out that have evaluated the contribution of specific "candidate genes" to ESRD and nephropathy susceptibility. A powerful alternative approach, the "genome scan," is a comprehensive genetic survey of the entire genome for chromosomal regions that are coinherited (i.e., linked) with a specific trait. The genome scan approach utilizes linkage analysis of genetic markers evenly spaced over all of the chromosomes in collections of families with multiply affected individuals. The genome scan approach is more difficult, time consuming, and expensive than candidate gene analysis, but has the advantage of being able to comprehensively survey the genome and locate new, potentially as yet undiscovered, genes. The limitation of the genome scan approach is that while it usually has the power to detect major genetic effects, it does not usually have the power to detect loci with small effects. Importantly, the genome scan approach is not limited by prior knowledge of, in this case, renal disease biology. We have performed the first such genome scan of diabetic nephropathy in African Americans in an effort to better understand the genetic contributors to this disorder.

Key words: African Americans, genetics, end-stage renal disease, type 2 diabetes, genome scan, linkage analysis, nephropathy.

Received for publication October 15, 2003
and in revised form February 7, 2004, and April 7, 2004
Accepted for publication April 26, 2004

METHODS

Subjects

DNA samples were collected from self-described African American families with multiple type 2 diabetes mellitus (T2DM) and ESRD or nephropathy-affected members. Briefly, families were originally identified through a proband with ESRD associated with T2DM. 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 in the presence of the following criteria: 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 were available, all subjects had 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 also recruited. Recruitment strategies and selection criteria have been described in detail previously [5–10]. The family set for the genome-wide scan comprised 166 African American families, with 206 (176 full-sibs and 30 half-sibs) DN-affected sibling pairs (ASP) totaling 355 affected individuals. One hundred forty-nine of the families contained 2 affected siblings, 15 families had 3 affected siblings, and 2 families had 4 affected siblings. 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 10 families. Of the affected individuals, 278 had T2DM with ESRD and 77 had diabetes with chronic renal failure (CRF). Sixty-eight individuals in the families had T2DM without a diagnosis of ESRD or CRF, 46 of which were unaffected and 22 had unknown renal status. For the purposes of this manuscript ESRD- and CRF-affected individuals have been treated the same and are described collectively as DN-affected individuals, with the exception of analyses that incorporate age at diagnosis of ESRD and duration of diabetes to ESRD. In these cases, only ESRD cases (and not CRF cases) were incorporated into the models since clear definition of age at onset of CRF was not possible.

Genotyping

DNA extraction was performed using the PureGene system (Gentra Systems, Minneapolis, MN, USA). Through the International Type 2 Diabetes Linkage Analysis Consortium, funded by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), a genome-wide scan was completed by the Center for Inherited Disease Research (CIDR). The marker set used was based on Marshfield Panel 8, with approximately 10% of the markers changed from the previous Marshfield panel. It consisted of largely tetra- and tri-nucleotide repeats, and included 392 primer pairs at an average spacing of 8.9 cM, and no intermarker gaps greater than 18 cM.

Each pedigree was examined for consistency of familial relationships using Pedigree RElationship Statistical Test (PREST) [11]. 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. A total of 28 pedigrees (17%) exhibited probable incorrect familial relationships and were modified as above, with 96% (27 of 28 families) of these changes being from a full-sibling to half-sibling relationship. Each genetic marker was also examined for Mendelian inconsistencies using PedCheck [12], and sporadic problem genotypes converted to missing. Allele frequency estimates were derived from the genome scan genotyping data from the families through computing the maximum likelihood methods implemented in the software Recode (D. Weeks, personal communication). Map distances were based on the Marshfield genetic map [13].

Linkage analyses

Multipoint linkage analyses were carried out using nonparametric linkage (NPL) regression analyses using the NPL_{pairs} statistics outputted from a modified version of Genehunter [14–17]. The NPL regression approach is a conditional logistic regression analysis in which the family-specific NPL statistic (e.g., NPL_{pairs}) at one or more loci is the predictor variable. Consider a sample of m independent pedigrees and a chromosomal region with one or more markers and a locus of interest. Let τ_i be the pedigree-specific contribution to the NPL statistic at the locus of interest. The likelihood function for a conditional logistic regression with τ_i as a predictor is

$$Lik(\beta; y_i, \tau) \propto \prod_{i=1}^{m} \left[\frac{\exp\{y_i \tau_i \beta\}}{1 + \exp\{\tau_i \beta\}} \right].$$

Here, $y_i = 1$ for all i, and β is the conditional logistic regression parameter. It can be shown that the score test from this likelihood is asymptotically equivalent to Whittemore and Halpern's class of tests [18]. Although unaffected individuals can be used to help estimate the possible inheritance vectors for that pedigree, an NPL regression analysis is an "affected's only" analysis. The primary advantage of the NPL regression approach is that it allows us to evaluate simultaneously, either by joint or conditional hypothesis tests, the effects of multiple loci (i.e., heterogeneity) and test for interactions among sets of loci (e.g., epistasis). To test for an interaction between

two loci, the two locations and their statistical interaction were included into the model and the 1 degree of freedom test of the interaction coefficient was computed. We also tested for interactions between the degree of sharing (IBD) at a location and (*1*) the mean age of diagnosis of ESRD, (*2*) mean duration of diabetes before diagnosis of ESRD, and (*3*) the mean age of diagnosis of diabetes.

Ordered subsets analysis

If a subset of pedigrees that are phenotypically more homogeneous can be identified, it should be possible to improve the power of linkage analysis. Ordered subset analyses (OSA) [19] were computed to investigate the influence of a pedigree's mean age of diagnosis of ESRD, mean age of diagnosis of diabetes, 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 ESRD diagnosis, each pedigree was ranked from lowest to highest for age at ESRD diagnosis. The family with the lowest mean age at ESRD diagnosis entered into the analysis, and the corresponding LOD score was computed on the target chromosome (e.g., chromosome 1) for that family. Next, a second linkage analysis on the target chromosome 1 was computed, combining the two families with the two lowest mean ages at ESRD diagnosis values. The i^{th} OSA analysis proceeds by computing a linkage analysis on the target chromosome using the subset of families with the i^{th} lowest mean ages at ESRD 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.

In several cases with the OSA we compared the demographic characteristics of the optimal subset of families showing evidence for linkage to the characteristics of the remaining families. To compare phenotypes between pedigrees that maximize evidence of linkage and the remaining pedigrees we computed a generalized estimating equation (GEE1) analysis, assuming exchangeable correlation and a robust variance estimation [20].

Table 1. Characteristics of individuals in African American DN families

Trait	Mean	Median	SD	Range Low	Range High
% Female	62.5%	NA	NA	NA	NA
Age at diabetes diagnosis *years*	40.1	40.0	11.4	14	76
Age at ESRD diagnosis *years*	55.9	56.0	9.5	27	85
Diabetes to ESRD duration *years*	17.6	17.0	9.4	0	49
Duration of ESRD at enrollment *years*	3.5	2.0	3.7	0	24
Age at enrollment *years*	58.8	59.0	10.3	24	99
BMI kg/m^2	31.2	30.1	7.3	15.2	64.8
HbA1c %	8.6	8.0	2.1	4.8	16

RESULTS

Clinical and phenotypic data for African American DN individuals

The clinical and phenotypic characteristics for the genotyped DN affected individuals are summarized in Table 1. The genotyped population was 62.5% female, probably reflecting both the increased prevalence of T2DM among African American women [21], survival, and participation bias. The average age at diagnosis of diabetes is relatively early at 40.1 years, with an average duration of diabetes of 17.6 years before the onset of ESRD. The age at diabetes diagnosis and age at ESRD onset are strongly correlated ($r = 0.32$, $P < 0.0001$). There were 24 DN subjects with age at onset of diabetes less than 25 years of age. The average BMI at enrollment into the study of this early diabetes onset group was 31.7 (range 21.9 to 43.8), suggesting that many of these individuals do have T2DM, but we cannot exclude the possibility that some subjects have type 1 diabetes. Ten subjects were diagnosed with ESRD and diabetes simultaneously. The great majority of the ESRD affected subjects was enrolled within 5 years of developing ESRD. Overall, the DN-affected individuals were obese at the time of their enrollment in the study (median BMI >30). In this dataset there is no evidence that duration of ESRD is significantly correlated with BMI in this subject group ($r = 0.006$; $P = 0.18$). The diabetes-affected individuals have relatively poor glucose control (median HbA_{1c} of 8.0% with the normal range being 4.5% to 5.7%).

Single-locus linkage results

A multipoint linkage analysis was carried out and multipoint LOD score curves for each chromosome were generated. The maximum LOD score for each chromosome

Table 2. NPL regression analysis of genome scan data: Single locus and multilocus interaction analysis

Chromosome	Position of Position of maximum LOD *cM*	Nearest marker	Single locus analysis			Multilocus analysis	
			LOD	LOD-1 interval	*P* value	LOD	LOD-1 interval
1	73	D1S3721	0.91	38.0-87.5	0.041		
2	248	D2S2968	0.15	231.0-257.5	0.401		
3	6	D3S2387	0.34	pter-50.5	0.212		
4	42.5	D4S391	0.61	16.5-72.0	0.094		
5	130	D5S1505	0.39	65.0-148.5	0.178		
6	165	D6S1035	0.89	115-qter	0.043		
7	29	D7S3051	1.43	10.5-41.0	0.010	1.37	7.5-39
8	32.5	D8S1145	0.90	10.5-114.0	0.041		
9	158.5	D9S1826	0.33	138.5-qter	0.214		
10	149	D10S1656	0.89	79-qter	0.043		
11	43	D11S1392	0.31	26.5-56.5	0.233		
12	55	GATA91H0	1.06	24.5-75.5	0.027	1.38	37.5-76
13	83	D13S779	0.51	61.5-101.5	0.126		
14	118.5	D14S1434	0.60	96.5-qter	0.098	1.23	105.5-qter
15	112	D15S966	0.09	106-qter	0.530		
16	51.5	D16S769	1.00	26.0-79.5	0.032	1.63	37-70
17	125.5	D17S928	0.22	107.5-qter	0.310		
18	99	D18S1364	1.07	69.5-qter	0.026	1.34	90.5-qter
19	45.5	D19S714	0.77	pter-81.5	0.060		
20	44.5	D20S477	0.13	28.5-53.0	0.436		
21	57.5	D21S1446	0.57	43.5-qter	0.105	1.15	20-qter
22	32	D22S685	0.97	pter-qter	0.035		

NPL, nonparametric linkage.

and any secondary linkage peaks are presented in Table 2. Only four regions of the genome yielded LOD scores greater than 1. Chromosome 7 at 29 cM near marker D7S3051 had the strongest evidence for linkage with DN (LOD 1.43). Other regions with evidence for linkage to DN were on chromosome 12 at 55 cM (near GATA91H0, LOD 1.06), on chromosome 16 at 51.5 cM (near D16S769, LOD 1.00), and on chromosome 18 at 99 cM (near D18S1364, LOD 1.07).

Multilocus and interaction linkage analysis results

With evidence that more than one genomic locus may contribute to DN, the data were analyzed using a multilocus NPL regression model (Table 2). Six chromosomal regions in 7p, 12p, 14q, 16p, 18q, and 21q remained statistically significant ($P < 0.05$) after adjusting for the evidence for linkage at the other five chromosomal regions. Specifically, considering these loci within the same model and computing the five 1 degree of freedom tests of significance yielded increases in the LOD scores for chromosomes 12p, 14q, 16p, 18q, and 21q, and a modest decline in the LOD score for 7p relative to the corresponding single-locus models for each locus. In each case with the increased LOD scores, the LOD-1 interval also narrows in the multilocus analysis.

NPL regression analysis: Interaction with phenotypic traits

In all likelihood, multiple genetic loci contribute to ESRD susceptibility. In some cases the presence of such loci may be difficult to detect when assessed in the background of a large, heterogeneous mixture of families. This possibility has been evaluated through the application of the NPL regression analysis method to evaluate interactions with phenotypic traits of age at ESRD diagnosis, duration of T2DM prior to ESRD diagnosis, and age at diagnosis of diabetes. The results of the NPL regression locus-specific linkage are summarized in Table 3, where regions showing statistically significant ($P < 0.05$) interactions with the trait are listed, and the direction and magnitude of the interaction is indicated by Pearson's correlation coefficient. These results suggest loci linked to DN at several regions of the genome are detectable when adjusting for age at ESRD diagnosis, duration of diabetes prior to ESRD, or age at T2DM diagnosis. Analysis of interaction with age at ESRD diagnosis identifies 9 genomic locations that show significant interactions, with loci on 1p at 89 cM and 3q at 140 cM having the strongest evidence for linkage. Effects are observed in both directions (e.g., interactions with families with mean younger age at diagnosis or older age at diagnosis). The negative Pearson's correlation coefficients indicate that interactions were with *lower* mean ages of diagnosis, and the linked families at the chromosome 1 locus were on average over 4.5 years younger, and on chromosome 3, 3.5 years younger at ESRD diagnosis than the unlinked families. A substantial number of loci, 19 in total, displayed significant ($P < 0.05$) interactions with duration of T2DM to onset of ESRD. The magnitude of these interactions were, on average, stronger than for age at ESRD diagnosis, with *P* values ranging up to 0.00013 for interaction on chromosome 16p at 23 cM in a group of families with 4 years longer duration of diabetes before ESRD onset. It

Table 3. Results of NPL regression interaction analyses across the genome for age at diagnosis of ESRD, duration between diagnosis of diabetes and ESRD, and age at diagnosis of type 2 diabetes in African American ESRD families

Chromosome	Position *cM*	Nearest marker	Interaction *P* value	Mean ± SD (number of families)		Pearson's corrleation coefficient
				Linked families	Unlinked families	
Interaction with age at diagnosis of ESRD						
1	89	D1S3728	0.007	53.6 ± 7.3 (52)	58.1 ± 7.9 (63)	−0.27 (early onset)
2	247	D2S2968	0.042	55.0 ± 7.5 (54)	57.0 ± 8.3 (51)	−0.19 (early onset)
3	140	D3S4523	0.016	54.6 ± 8.0 (65)	58.0 ± 7.5 (50)	−0.22 (early onset)
4	99	D4S2361/D4S1647	0.045	55.2 ± 7.1 (49)	56.7 ± 8.5 (66)	−0.13 (early onset)
5	0	D5S2488	0.022	54.9 ± 7.7 (47)	57.6 ± 7.8 (68)	−0.21 (early onset)
6	119	D6S474	0.040	58.1 ± 6.5 (52)	54.5 ± 8.6 (63)	0.17 (late onset)
16	23	D16S748	0.040	57.5 ± 6.8 (58)	54.7 ± 8.8 (57)	0.20 (late onset)
16	87	D16S2624	0.042	57.7 ± 8.5 (50)	54.8 ± 7.3 (65)	0.21 (late onset)
17	80	D17S1290	0.032	55.1 ± 8.8 (50)	56.8 ± 7.2 (65)	−0.19 (early onset)
Interaction with duration of diabetes to ESRD						
1	143	D1S3723	0.019	10.0 ± 5.5 (40)	8.7 ± 5.5 (55)	0.25 (long duration)
3	**31**	**D1S4545/D3S1259**	**0.003**	**10.5 ± 6.1 (41)**	**8.2 ± 4.7 (54)**	**0.28 (long duration)**
3	182	D3S3053	0.035	11.4 ± 6.1 (34)	8.0 ± 4.7 (61)	0.20 (long duration)
5	8	D5S2849	0.043	10.0 ± 6.1 (43)	8.6 ± 4.8 (52)	0.19 (long duration)
6	155	D6S2436	0.010	10.9 ± 5.8 (43)	7.9 ± 4.8 (52)	0.27 (long duration)
6	178	D6S1277	0.040	9.6 ± 6.1 (46)	8.88 ± 4.9 (49)	0.22 (long duration)
7	**7**	**D7S3056**	**0.0002**	**11.0 ± 6.0 (40)**	**8.0 ± 4.7 (55)**	**0.39 (long duration)**
7	**33**	**D7S1802**	**0.008**	**10.4 ± 6.5 (48)**	**8.0 ± 3.8 (47)**	**0.22 (long duration)**
7	118	D7S1799	0.003	10.7 ± 6.4 (46)	7.8 ± 3.9 (49)	0.27 (long duration)
7	171	D7S3058	0.007	10.6 ± 6.9 (35)	8.4 ± 4.3 (60)	0.27 (long duration)
08	82	D8S1136	0.042	10.5 ± 6.4 (40)	8.3 ± 4.5 (55)	0.22 (long duration)
12	156	D12S2078/D12S1045	0.031	9.7 ± 5.2 (50)	8.65 ± 5.7 (45)	0.17 (long duration)
13	**9**	**D12S787**	**0.008**	**10.3 ± 6.3 (39)**	**8.5 ± 4.8 (56)**	**0.25 (long duration)**
13	**41**	**D13S325**	**0.005**	**10.9 ± 6.3 (36)**	**8.3 ± 4.7 (59)**	**0.27 (long duration)**
14	12	D14S742	0.014	10.8 ± 5.6 (38)	8.2 ± 5.2 (57)	0.24 (long duration)
15	90	D15S652	0.036	10.2 ± 5.6 (37)	8.58 ± 5.37 (58)	0.20 (long duration)
16	**23**	**D16S748**	**0.00013**	**11.2 ± 6.1 (48)**	**7.2 ± 3.9 (47)**	**0.34 (long duration)**
18	75	D18S851	0.008	10.6 ± 5.4 (46)	7.91 ± 5.3 (49)	0.26 (long duration)
20	90	D20S451	0.006	10.7 ± 5.4 (34)	8.42 ± 5.4 (61)	0.30 (long duration)
Interaction with age at diagnosis of diabetes						
1	**91**	**D1S1665**	**0.004**	**37.4 ± 8.1 (66)**	**41.2 ± 8.3 (81)**	**−0.23 (early onset)**
2	159	D2S1399/D2S1353	0.023	40.8 ± 8.2 (53)	38.8 ± 8.5 (94)	0.16 (late onset)
3	135	D3S2460	0.036	37.7 ± 8.6 (66)	41.0 ± 8.0 (81)	−0.16 (early onset)
4	146	D4S1625	0.005	37.9 ± 7.4 (52)	40.4 ± 8.9 (95)	−0.23 (early onset)
6	55	D6S2427	0.024	38.0 ± 7.8 (70)	40.9 ± 8.8 (77)	−0.18 (early onset)
9	**93**	**D9S283**	**0.0001**	**37.2 ± 7.6 (59)**	**41.1 ± 8.7 (88)**	**−0.24 (early onset)**
10	96	D10S1432	0.041	38.0 ± 8.7 (64)	40.7 ± 8.1 (83)	−0.17 (early onset)
10	171	D10S212	0.003	41.3 ± 8.5 (76)	37.6 ± 8.9 (71)	0.24 (late onset)
12	22	GATA49D12/D12S391	0.009	41.4 ± 9.1 (69)	37.8 ± 7.4 (78)	0.19 (late onset)
13	21	D13S217/D13S1493	0.006	38.0 ± 7.8 (56)	40.5 ± 8.7 (91)	−0.19 (early onset)
13	76	D13S793	0.031	37.7 ± 8.6 (67)	41.1 ± 8.1 (80)	−0.19 (early onset)
14	44	D14S306	0.025	37.9 ± 7.8 (47)	40.3 ± 8.7 (100)	−0.17 (early onset)
17	62	D17S1299	0.020	38.2 ± 7.5 (59)	40.4 ± 9.0 (88)	−0.18 (early onset)
18	**75**	**D18S851**	**0.005**	**36.9 ± 7.2 (67)**	**41.7 ± 8.8 (80)**	**−0.20 (early onset)**

Nominally significant ($P < 0.05$) results are shown. Interaction *P* values ≤ 0.005 are shown in bold type.

is noteworthy that this locus, interacting with longer duration of diabetes, was also detected in the ESRD age at onset interaction analysis with later age at ESRD diagnosis. Other interactions with $P < 0.01$ are on chromosome 3p (31 cM), 7p (at 7 and 33 cM), 7q (171 cM), 13 (at 9 and 41 cM), 18q (75 cM), and 20q (90 cM). There is also evidence of a 3q locus at 182 cM. This locus is distal to and significantly separated from the 3q (140 cM) locus, interacting with age at diagnosis of ESRD. Finally, evaluation of interaction with age at diagnosis of diabetes resulted in 14 loci with significant interactions. The most significant interaction with age at diabetes diagnosis was detected on 9q near D9S283 (91 cM, $P = 0.0001$), in a subgroup with age at diagnosis 3.9 years earlier as indicated by a Pearson's correlation coefficient of −0.24. Additional stronger interactions are detected on chromosome 1p at 91 cM (seen also in ESRD age at diagnosis analysis), chromosome 10 at 171 cM (older age at diagnosis), and chromosome 18q at 75 cM (younger age at diagnosis; also seen in duration of diabetes to ESRD analysis).

Ordered subsets analysis with phenotypic traits

OSA is another analytical approach that can be used to evaluate linkage under assumptions that linkage can be more readily detected in subgroups of families within a population, differentiated by specific phenotypic traits.

Table 4. Ordered subsets analysis (OSA) across the genome of age at diagnosis ESRD, duration of type 2 diabetes to ESRD, or age at diagnosis of diabetes of African American ESRD families

Chromosome	Linked subset	Position *cM*	Nearest marker	Entire sample LOD	Maximized LOD	Mean ± SD Optimal subset	Mean ± SD Remaining families	Empirical *P* value for change	% Peds
Subsetting by age at diagnosis of ESRD									
3	**Early onset**	**135**	**D3S2460**	**1.27**	**4.55**	**46.8 ± 4.6**	**59.7 ± 5.6**	**0.004**	**29**
17	Early onset	67.5	D17S2180	0.04	1.76	41.5 ± 3.3	57.7 ± 6.4	0.047	10
Subsetting by duration of diabetes to ESRD									
7	**Longer duration**	**33**	**D7S1802**	**1.47**	**3.59**	**14.3 ± 4.6**	**6.0 ± 2.9**	**0.011**	**37**
12	Longer duration	158.5	D12S1045	0.78	2.94	14.5 ± 4.6	6.1 ± 2.9	0.025	36
13	Longer duration	17	D13S217	0.06	1.92	18.1 ± 4.2	7.3 ± 3.4	0.024	17
16	Longer duration	23	D16S748	0.04	2.85	17.1 ± 4.3	7.0 ± 3.3	0.013	21
Subsetting by age at diagnosis of diabetes									
3	Early onset	135	D3S2460	0.29	2.52	28.8 ± 2.8	43.9 ± 7.5	0.015	18
10	Late onset	161.5	D10S217	0.41	2.65	54.4 ± 4.50	38.3 ± 6.8	0.035	19
12	Late onset	39	D12S373	0.70	2.86	46.5 ± 6.5	32.7 ± 5.2	0.035	64
18	**Early onset**	**100.5**	**D18S1364**	**1.64**	**3.72**	**35.7 ± 5.4**	**50.7 ± 5.2**	**0.009**	**64**
19	**Late onset**	**21**	**D19S1034**	**0.60**	**3.13**	**51.3 ± 5.0**	**36.3 ± 5.8**	**0.020**	**34**
20	Late onset	39	D20S470	0.027	2.50	56.2 ± 4.5v	39.2 ± 7.3	0.009	13
21	Late onset	57	D21S1446	0.45	2.59	52.2 ± 4.7	36.9 ± 6.1	0.009	29

Maximized LOD scores ≥3.0 are shown in bold type.

This approach was used to search for differential evidence for linkage depending on the age at diagnosis for ESRD, duration of diabetes before ESRD onset, and age at diagnosis of diabetes: the same traits used in the NPL regression interaction analysis. Regions displaying a significant change in chromosome-specific *P* value ($\Delta P < 0.05$) are shown in Table 4. One region on the long arm of chromosome 3 near D3S2460 at 135 cM exhibited an OSA maximum LOD score of 4.55 in an optimum subset analysis on the 48 pedigrees (29%), with the earliest age at diagnosis compared to a LOD score overall of 1.27 (chromosome-wide *P* value, $\Delta P = 0.004$). The optimum linked family set had an age at onset of ESRD almost 13 years earlier than the remaining families. This chromosome 3 locus was also significant when subsetting on age at diabetes diagnosis. Figure 1A shows the LOD score graph for chromosome 3 with the NPL single locus regression and OSA results. Note that the OSA incorporates calculation of an empirical chromosome-specific *P* value (in this case $P = 0.004$) to evaluate whether these results could be randomly expected. Three other loci, one subsetting on longer duration of diabetes to ESRD on 7p (optimal subset LOD = 3.59, 37% of pedigrees), and two loci on 18q (optimal subset LOD = 3.72; 64% of pedigrees) and 19q (optimal subset LOD = 3.13; 34% of pedigrees) subsetting on age at diabetes diagnosis and subsetting on earlier and later onset of diabetes, respectively, also had optimal LOD scores greater than 3. The LOD score graphs for the chromosome 7 and 18 OSA results are shown in Figure 1B and C, respectively, which show NPL multilocus regression results in addition to the single locus NPL regression and OSA. (The multilocus NPL regression did not reach statistical significance for chromosome 3). Eight additional chromosomal regions provided chromosome-wide statistically significant ($\Delta P < 0.05$) increases in the LOD score based upon subset analysis for early diagnosis of ESRD, longer duration of diabetes to ESRD, and either earlier or later diagnosis of diabetes (Table 4).

For the chromosome 3, 7, and 18 OSA results we have assessed whether the characteristics of the families in the optimal subset varied significantly from characteristics in the rest of the families in the study. These results are summarized in Table 5, which shows the mean and standard deviation and medians for linked and unlinked pedigrees, and reports the *P* value from the GEE1 analysis (see **Methods**). The chromosome 3 optimal subset was based on earlier age of diagnosis of ESRD, which is, as expected, significantly different between the two sets of families. Other traits are also significantly different: duration of diabetes to ESRD, age at recruitment, and age at diagnosis of diabetes. Each of these traits is correlated with age of ESRD diagnosis, so this is not surprising. The chromosome 7 optimal subset was based on longer duration of diabetes to ESRD, which is significantly different between the two sets of families. Other traits are also significantly different: age at diagnosis of diabetes strongly, and age and BMI more weakly. Age at diagnosis of diabetes and age are, again, correlated with duration of diabetes. The lower BMI in the optimal subset may reflect the long-term effects of diabetes on the health, and, as such, the BMI of the longer duration of diabetes individuals. Finally, the chromosome 18 optimal subset was based on earlier age of diagnosis of T2DM, which is significantly different between the two sets of families. Other traits are also significantly different: duration of diabetes to ESRD, age at recruitment, and age at diagnosis of ESRD. Again, each of these age-related traits is correlated, at least in

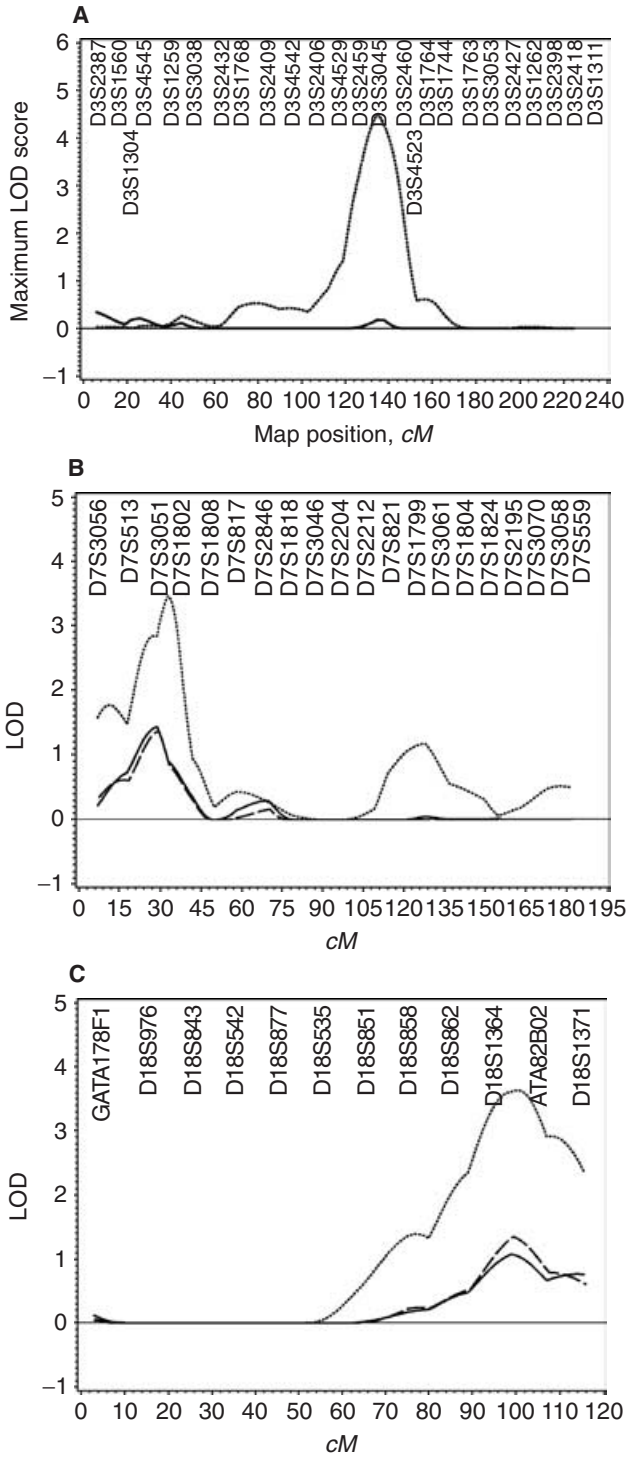

Fig. 1. Logarithm of odds (LOD) score results with (*A*) chromosome 3, earlier mean age at end-stage renal disease (ESRD) pedigrees, (*B*) chromosome 7, longer duration of diabetes to ESRD, and (*C*) chromosome 18, earlier age of diagnosis of diabetes. Marker locations are shown across the top, and map position in cM is shown along the bottom. Solid line = single locus nonparametric linkage (NPL) regression; small dashed line = OSA subset analysis; long dashed line = multilocus NPL regression. The NPL multilocus analysis was not included for chromosome 3 because it did not reach statistical significance. Physical map locations of the peak LOD scores from UCSC Genome Browser (http://genome.ucsc.edu/) are: chromosome 3, 118623238 bp; chromosome 7, 20349565 bp; chromosome 18, 61649324 bp.

part, with the analysis trait (in this case, age of diabetes diagnosis). The means and SDs reported in Tables 4 and 5 differ in the case of chromosomes 3 and 18 modestly, and in the case of chromosome 7 to a greater degree, due to the family means being used in the OSA analysis and individual means being compared in Table 5.

DISCUSSION

To our knowledge this investigation represents the first large-scale effort to map chromosomal locations of genes specifically contributing to T2DM-associated ESRD and nephropathy in the African American population. There are only two other complete reports of genome scans for diabetic nephropathy. Imperatore et al [22] carried out a genome scan for microvascular disease in 98 Pima Indian diabetes-affected sibling pairs and identified evidence for linkage to nephropathy on chromosome 7q, and suggestive evidence for linkage on 3q, 9q, and 20p. Vardarli et al [23] carried out a genome scan in 18 large Turkish families with T2DM and diabetic nephropathy and found strong evidence for linkage (LOD = 6.1) on 18q22.3–23. Evaluation of these loci in the Pima Indian dataset also showed evidence of confirmation ($P = 0.013$–0.006), although this region was not linked in the original Pima genome screen. In one other study, Moczulski et al [24] carried out a focused analysis of several genomic regions and identified evidence for linkage at nephropathy in type 1 diabetes families on the long arm of chromosome 3 in the region of the angiotensin II type 1 receptor (ATR1).

In carrying out this study we proceeded from the assumptions that DN is a genetically complex disease that has both genetic and environmental and lifestyle components. Consequently, a realistic search for ESRD genes requires consideration of both multigenic and phenotypic influences. This study incorporates relatively novel approaches to evaluate these types of interactions (e.g., nonparametric linkage regression multilocus modeling and ordered subsets analysis). We used OSA based on phenotypes such as age at onset of ESRD and diabetes, and duration of diabetes before ESRD onset, in an effort to define more homogeneous subgroups of families that could potentially reveal evidence of linkage.

Only limited evidence of linkage was evident at the first stage at analysis (Table 2) with 4 LOD scores ≥1.0, but less than 2, with the highest LOD score in 7p. In the multilocus analysis, incorporating an evaluation of heterogeneity, 5 chromosomal regions showed evidence of significant interaction in the multilocus models, with the strongest evidence (LOD = 1.63) on 16p. When analytical approaches that incorporate phenotypic trait data were applied, as summarized in Tables 3 and 4, evidence for multiple chromosomal loci contributing to ESRD susceptibility was revealed. In fact, with the NPL regression analysis evaluating interactions with age at ESRD

Table 5. Characteristics of families in the ordered subset analyses

Trait	Unlinked # of pedigrees = 82		Linked # of pedigree = 33		
	Mean (SD)	Median (N)	Mean (SD)	Median (N)	*P* value
Chromosome 3 subset: Subsetting by age at diagnosis of ESRD					
Diabetes to ESRD	19.38 (9.32)	19 (183)	15.34 (5.73)	15 (59)	0.002
Age	60.42 (10.03)	61 (331)	51.89 (6.71)	53 (72)	<0.0001
ESRD age of onset	58.83 (8.5)	59 (210)	46.91 (6.06)	48 (70)	<0.0001
T2DM age of diagnosis	41.34 (11.6)	41 (324)	34.38 (8.08)	34 (72)	<0.0001
BMI	30.98 (7.33)	29.88 (328)	32.01 (7.29)	31.57 (68)	0.26
	Unlinked # of pedigrees = 82		Linked # of pedigree = 37		
	Mean (SD)	Median (*N*)	Mean (SD)	Median (*N*)	*P* value
Chromosome 7 subset: Subsetting by duration of diabetes to ESRD					
Diabetes to ESRD	16.16 (7.32)	15 (164)	23.09 (8.95)	22 (78)	<0.0001
Age	58.29 (10.58)	58 (329)	61.59 (6.72)	62 (74)	0.018
ESRD age of onset	55.14 (10.04)	54 (201)	57.65 (7.66)	58 (79)	0.069
T2DM age of diagnosis	41.4 (11.35)	41 (315)	34.91 (9.87)	33 (81)	<0.0001
BMI	31.70 (7.44)	30.96 (322)	28.81 (6.31)	27.68 (74)	0.022
	Unlinked # of pedigrees = 82		Linked # of pedigree = 141		
	Mean (SD)	Median (*N*)	Mean (SD)	Median (*N*)	*P* value
Chromosome 18 subset: Subsetting by age at diagnosis of diabetes					
Diabetes to ESRD	15.15 (7.29)	13 (62)	19.51 (8.95)	19 (180)	0.0003
Age	63.98 (8.76)	64 (137)	56.28 (9.69)	56 (266)	<0.0001
ESRD age of onset	61.21 (8.3)	61 (91)	53.27 (8.94)	53 (189)	<0.0001
T2DM age of diagnosis	49.5 (9.17)	50 (118)	36.08 (9.72)	35 (278)	<0.0001
BMI	30.57 (7.45)	29.19 (135)	31.46 (7.25)	30.50 (261)	0.45
HBA1C	8.69 (2.81)	8.1 (47)	8.83 (2.35)	8.25 (72)	0.99

SD, standard deviation; *N*, number of subjects.

diagnosis, duration of diabetes before ESRD, and age at diabetes diagnosis, a large number of loci: 9, 19, and 14, respectively, showed significant evidence for linkage.

Using the OSA approach (Table 4) to subset families ranked based on these phenotypic traits identified subsets which, in some cases, showed dramatic increases in LOD scores compared with the entire family set. The most dramatic example is the chromosome 3q locus where subsetting based on age at ESRD diagnosis revealed a subset of families with earlier onset of ESRD that had a combined LOD score of 4.55 compared to a LOD score of 1.27 for the entire sample. One challenge of using novel analytical approaches is to assess the significance of the resulting LOD score (in this case of OSA). For the OSA analysis an empiric chromosome specific *P* value has been calculated for chromosome 3 ($P = 0.004$), which suggests that the LOD score does indeed represent significant evidence for linkage.

Faced with a large number of potentially linked loci, it is challenging to evaluate significance and to prioritize which regions are most likely to contain ESRD genes. Overall, there is no overwhelming evidence for linkage in the entire sample but each of the major peaks (chromosomes 3, 7, and 18) in the OSA analyses have LOD scores over 3.5 and, in the case of the chromosome 3 analysis subsetting on age of ESRD diagnosis, a LOD over 4.5. These are, however, maximized LODs under optimal conditions, and conventionally used criteria for evidence of linkage (e.g., Lander and Kruglyak, [25]) do not directly apply. As in any genome scan, one must be mindful that some of these linkage results could represent chance events.

We have weighed their significance using multiple criteria: magnitude of LOD scores, consistent evidence of linkage in multiple analysis approaches, and evidence from other genome scans and other genetics studies. Several of the loci detected here show consistent evidence of linkage using each of these criteria. As outlined above, the highest LOD score was observed in the OSA on chromosome 3q at 135 cM (Table 4) when subsetting on age at ESRD diagnosis. Consistent with this observation, NPL regression analysis detected evidence for linkage to this same locus in the interaction analysis with age at ESRD onset ($P = 0.016$) and age at diabetes diagnosis ($P = 0.036$). Each of these linkage results, as with the ordered subset analysis, is with earlier age at diagnosis. Evidence for linkage to 3q was also observed by Moczulski et al [24] in type 1 diabetes-associated nephropathy, and by Imperatore et al [22] in the Pima Indians, although linkage peak in the Pimas appears to be more distal (approximately 180 cM).

In addition, there is evidence of linkage on 7p at approximately 29 cM in the NPL single locus and multilocus analysis. This locus is also detected in the NPL

interaction analysis with duration of diabetes before ESRD ($P = 0.008$ at 33 cM), and the OSA, where a subset of 37% of the families with longer duration between diabetes diagnosis and ESRD onset had a LOD score of 3.59. Another interesting locus is on 18q at approximately 100 cM. This locus has LOD scores of 1.0 and 1.34 in the single locus and multilocus NPL regression analysis, respectively, and a max. LOD score of 3.72 in families subsetted based on age at diabetes diagnosis in the OSA. This 18q location is the same area as the linkage reported by Vardarli et al [23]. Finally, in an earlier report from our laboratory we detected evidence for linkage with markers in 10q [26, 27] in a collection of families that contained both diabetic and non-diabetic ESRD families. In this study with a larger number of families (including the previously genotyped DN families described in Freedman et al [10]) and restricted to diabetes-associated nephropathy, evidence for linkage was reduced, though specific NPL regression analysis ($P = 0.003$) and OSA (optimal LOD = 2.65) still suggest evidence of linkage on chromosome 10q.

In another approach, we have compared the demographic characteristics of individuals in families that contributed to the evidence for linkage on chromosome 3, 7, and 18 using OSA with subjects from families that were not part of the optimal subset (Table 5). There are no unexpected differences between the subjects. Most of the significant differences are correlated with different measures of age (simple age, diabetes diagnosis, ESRD diagnosis, duration of diabetes to ESRD), all of which are correlated. It is interesting to note however, that there is only modest overlap in the families contributing to the OSA optimal subsets. For example the overlap between the chromosomes 3 and 7 optimal subsets is 7 of 29 and 37 families respectively (data not shown). This suggests that these linkages may indeed represent different DN loci.

This study evaluated families with sibling pairs that were concordant for both diabetes and renal disease. As such, the question arises of whether the evidence of linkage that we observe is to ESRD or diabetes. The families that made up this study are a subset of a significantly larger collection of African American families with multiple cases of type 2 diabetes. The African American diabetes families have also been the subject of a genome scan and subsequent analysis. There is little or no evidence for linkage to the 3q or 18q loci in the diabetes genome scan, so these seem likely to be true ESRD loci. In contrast, a major peak in the diabetes scan was observed on 7p and is associated with early age at diabetes diagnosis and lower BMI. In the DN scan, this 7p peak is not significantly associated with early age at diabetes diagnosis, but is associated with duration of diabetes before diagnosis of ESRD. We have also evaluated this DN peak for interaction with BMI, and similar to the T2DM peak, there is evidence of interaction with lower BMI ($P = 0.035$). Therefore, we cannot exclude the possibility that this represents a T2DM gene. It is possible that a genome scan currently underway in our group in nondiabetic families with ESRD will help clarify which phenotype is being mapped on chromosome 7p.

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 grants R01 DK53591 (DWB) and R01 HL56266 (BIF). 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.

Reprint requests to Donald W. Bowden, Ph.D., Department of Biochemistry, Wake Forest University School of Medicine, Medical Center Blvd., Winston-Salem, NC 27157.
E-mail: dbowden@wfubmc.edu

REFERENCES

1. Us Renal Data System: *USRDS 2003 Annual Data Report: Atlas of End-Stage Renal Disease in the United States*, Bethesda, MD, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, 2003
2. Bowden DW: Genetics of renal disease. *Kidney Int* 63(Suppl 83):S8–S12, 2003
3. Freedman BI, Tuttle AB, Spray BJ: Familial predisposition to nephropathy in African- Americans with non–insulin-dependent diabetes mellitus. *Am J Kidney Dis* 25:710–713, 1995
4. Spray BJ, Atassi NG, Tuttle AB, Freedman BI: Familial risk, age at onset, and cause of end- stage renal disease in white Americans. *J Am Soc Nephrol* 10:1806–1810, 1995
5. Yu H, Bowden DW, Spray BJ, *et al*: Linkage analysis between loci in the renin-angiotensin axis and end-stage renal disease in African Americans. *J Am Soc Nephrol* 7:2559–2564, 1996
6. Freedman BI, Yu H, Spray BJ *et al*: Genetic linkage analysis of growth factor loci and end-stage renal disease in African Americans. *Kidney Int* 51:819–825, 1997
7. Yu H, Bowden DW, Spray BJ, *et al*: Identification of human plasma kallikrein gene polymorphisms and evaluation of their role in end-stage renal disease. *Hypertension* 31:906–911, 1998
8. Yu H, Anderson PJ, Freedman BI, *et al*: Genomic structure of the human plasma prekallikrein gene, identification of allelic variants, and analysis in end-stage renal disease. *Genomics* 69:225–234, 2000
9. Freedman BI, Yu H, Anderson PJ, *et al*: Genetic analysis of nitric oxide and endothelin in end-stage renal disease. *Nephrol Dial Transplant* 15:1794–1800, 2000
10. Freedman BI, Rich SS, Yu H, *et al*: Linkage heterogeneity of end-stage renal disease on human chromosome 10. *Kidney Int* 62:770–774, 2002
11. Mcpeek MS, Sun L: Statistical tests for detection of misspecified relationships by use of genome-screen data. *Am J Hum Genet* 66:1076–1094, 2000
12. O'Connell JR, Weeks DE: PedCheck: A program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63:259–266, 1998
13. Broman KW, Murray JC, Sheffield VC, *et al*: Comprehensive human genetic maps: Individual and sex-specific variation in recombination. *Am J Hum Genet* 63:861–869, 1998
14. Langefeld CD: *The Application of Statistical Methods for Highly Stratified Data to Linkage and Association Analyses of Complex Genetic Traits*, Ann Arbor, MI, University of Michigan, 1999
15. Langefeld CD, Boehnke M: Multiple trait locus nonparametric linkage regression. *Am J Hum Genet* 65:A1140, 1999
16. Langefeld CD, Davis CC, Brown WM: Nonparametric linkage regression. I: Combined Caucasian CSGA and German genome scans for asthma. *Genet Epidemiol* 21(Suppl 1):S136–141, 2001

17. Davis CC, Brown WM, Lange EM, *et al*: Nonparametric linkage regression. II: Identification of influential pedigrees in tests for linkage. *Genet Epidemiol* 21(Suppl 1):S123–129, 2001
18. Whittemore AS, Halpern J: A class of tests for linkage using affected pedigree members. *Biometrics* 50:118–127, 1994
19. Hauser ER, Watanabe RM, Duren WL, *et al*: Ordered subset analysis in genetic linkage mapping of complex traits. *Genet Epidemiol* 27:53–63, 2004
20. Liang KY, Zeger SL: Longitudinal data analysis for discrete and continuous models. *Biometrics* 42:121–130, 1986
21. National Diabetes Information Clearinghouse: Diabetes in African Americans. National Institutes of Health publication 02–3266, 2002
22. Imperatore G, Hanson RL, Pettitt DJ: Sib-pair linkage analysis for susceptibility genes for microvascular complications among Pima Indians with type 2 diabetes. Pima Diabetes Genes Group. *Diabetes* 47:821–830, 1998
23. Vardarli I, Baier LJ, Hanson RL, *et al*: Gene for susceptibility to diabetic nephropathy in type 2 diabetes maps to 18q22.3–23. *Kidney Int* 62:2176–2183, 2002
24. Moczulski DK, Rogus JJ, Antonellis A, *et al*: Major susceptibility locus for nephropathy in type 1 diabetes on chromosome 3q: Results of novel discordant sib-pair analysis. *Diabetes* 47:1164–1169, 1998
25. Lander E, Kruglyak L: Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241–247, 1995
26. Yu H, Sale M, Rich SS, *et al*: Evaluation of markers on human chromosome 10, including the homologue of the rodent *Rf-1* gene, for linkage to ESRD in African Americans. *Am J Kidney Dis* 33:294–300, 1999