Am. J. Hum. Genet. 74:73–82, 2004

Linkage at 12q24 with Systemic Lupus Erythematosus (SLE) Is Established and Confirmed in Hispanic and European American Families

Swapan K. Nath,[1,2] Ana I. Quintero-Del-Rio,[1,3] Jeff Kilpatrick,[1] Lourdes Feo,[1] Maria Ballesteros,[5] and John B. Harley[1,2,3,4]

[1]Arthritis and Immunology Research Program and [2]Genetic Epidemiology Unit, Oklahoma Medical Research Foundation, [3]University of Oklahoma Health Sciences Center, and [4]Department of Veterans Affairs Medical Center, Oklahoma City; and [5]San Lucas Hospital, Ponce, Puerto Rico

Systemic lupus erythematosus (SLE) is a chronic, complex, and systemic human autoimmune disease, with both an environmental component and a heritable predisposition. Clinical studies, reinforced by epidemiology and genetics, show impressive variation in disease severity, expression, prevalence, and incidence by ethnicity and sex. To identify the novel SLE susceptibility loci, we performed a genomewide scan with 318 markers on 37 multiplex Hispanic families, using a nonparametric penetrance-independent affected-only allele-sharing method. Three chromosomal regions (12q24, 16p13, and 16q12-21) exceeded our predetermined threshold ($Z_{lr} > 2.32$; nominal $P < .01$) for further evaluation. Suspected linkages at 12q24, 16p13, and 16q12-21 were tested in an independent data set consisting of 92 European American (EA-1) and 55 African American (AA) families. The linkage at 12q24 was replicated in EA-1 ($Z_{lr} = 3.06$; $P = .001$) but not in AA ($Z_{lr} = 0.37$; $P = .35$). Although neither the 16p13 nor the 16q12-21 was confirmed in EA-1 or AA, the suggestive linkage ($Z_{lr} = 3.06$; $P = .001$) at 16q12-21 is sufficient to confirm the significant linkage, reported elsewhere, at this location. The evidence for linkage at 12q24 in the 129 combined (Hispanic and EA-1) families exceeded the threshold for genomewide significance ($Z_{lr} = 4.39$; $P = 5.7 \times 10^{-6}$; nonparametric LOD = 4.19). Parametric linkage analyses suggested a low-penetrance, dominant model (LOD = 3.72). To confirm the linkage effect at 12q24, we performed linkage analysis in another set of 82 independent European American families (EA-2). The evidence for linkage was confirmed ($Z_{lr} = 2.11$; $P = .017$). Therefore, our results have detected, established, and confirmed the existence of a novel SLE susceptibility locus at 12q24 (designated "SLEB4") that may cause lupus, especially in Hispanic and European American families.

Introduction

Systemic lupus erythematosus (SLE [MIM 152700]) is a complex disease in which immune responses are directed against a multitude of self-antigens. SLE in humans manifests a diverse array of clinical symptoms, variably involving multiple organ systems. The pathogenesis is related to dysregulation of self-reactive B cells, which leads to immune complex activation and complement consumption. The production of autoantibodies against nuclear components is the unifying characteristic of SLE. Antinuclear antibody (ANA) testing is very sensitive for the disease, although not specific, since ANA antibodies are sporadically detected in ⩽5% of the female population of age >40 years. On the other hand, antibodies to double-stranded DNA, ribosomal P protein, or Sm protein are very specific for SLE but are not particularly sensitive.

The overall estimated prevalence of SLE in the United States is 12–64 cases per 100,000 individuals (Hochberg 1997*a;* Lawrence et al. 1998). Significant sex differences are observed in prevalence (female:male = 9:1), age at onset, premorbid conditions, clinical expression, course of illness, response to treatment, and morbid risk. In addition, there are important racial differences in prevalence and disease manifestations. For example, a two- to fourfold higher incidence of SLE is observed in the African American (AA), as compared with European American (EA), population (Petri 1998).

The familial aggregation of SLE suggests an underlying genetic susceptibility; but environmental, stochastic, or epigenetic factors must be important, since even MZ twins are not usually concordant for disease. Substantial evidence has shown that SLE clusters in families, with 7%–12% increased risk among the first- or second-degree relatives of a proband. There is an increased concordance rate in MZ twins (15%–69%), as opposed to DZ twins (2%–5%) (Deapen et al. 1992). The relative risk ratio for the siblings of an affected proband (λ_s) varies from 20 to 40 (Wandstrat and Wakeland 2001).

Received September 10, 2003; accepted for publication October 17, 2003; electronically published December 4, 2003.

Address for correspondence and reprints: Dr. Swapan K. Nath, Arthritis and Immunology Research Program, Oklahoma Medical Research Foundation, 825 N.E. 13th Street, Oklahoma City, OK 73104. E-mail: swapan-nath@omrf.ouhsc.edu

0002-9297/2004/7401-0008$15.00

Table 1

Selected Clinical and Demographic Features of the SLE-Affected Individuals

Population	No. of Families	No. of SLE-Affected Individuals	Sex (male:female)	No. of SLE-Affected Individuals per Family	Age at Onset ± SD (years)	Age at Enrollment ± SD (years)
Hispanic	37	91	85:11	2.46	32.0 ± 12.3	38.6 ± 12.5
AA	55	128	121:7	2.33	34.4 ± 10.7	40.2 ± 11.0
EA-1	92	214	197:17	2.33	35.0 ± 13.5	42.0 ± 13.5
EA-2	82	179	162:17	2.18	36.0 ± 13.7	44.3 ± 13.9

Moreover, the complex pattern of inheritance of SLE suggests multigenic inheritance, which requires interaction of various combinations of contributing genes at multiple loci in individual patients.

For disorders, such as SLE, that have a poorly known biochemical basis, identification of the genes is a prerequisite to understanding its biological basis. Therefore, identification of the genes contributing to susceptibility in SLE will elucidate the development and pathogenesis of the disease and may lead to novel therapeutic interventions. In addition, genetic screening could be used to identify individuals who are at risk so that they can take advantage of early diagnosis and treatment.

Here, we report results of a genomewide search for candidate regions for SLE susceptibility. First, we present the results of an independent first-stage genome scan on a set of 37 Hispanic families multiplex (at least two affected relatives per family) for SLE, and we identify three potential candidate regions for SLE susceptibility. Second, two independent samples consisting of 55 AA and 92 EA (EA-1) families are used to replicate the results. Combined study of the Hispanic and EA-1 families has established the significant evidence of linkage at 12q24. Third, we present a plausible genetic model for SLE at 12q24. Fourth, we have replicated the evidence for linkage at 12q24 in another set of 82 independent EA families (EA-2).

Material and Methods

Family Ascertainment

Families with SLE were recruited from all regions of the United States. Multiplex families were ascertained for linkage study. All the patients with SLE met the 1997 revised criteria for classification of SLE (Hochberg 1997*b;* Tan et al. 1982). All patient material was gathered, after receiving informed consent from participants, under protocols approved by the Oklahoma Medical Research Foundation and University of Oklahoma Health Sciences Center institutional review boards. Each potential participant with SLE was interviewed and asked to complete an extensive questionnaire that included demographic variables. All available medical records of SLE-affected family members were reviewed to confirm the diagnosis. In some cases, the participant's physician was also interviewed. Family members without SLE completed a less extensive questionnaire to screen for SLE and other autoimmune diseases. Clinical manifestations were defined as either "present" or "absent" on the basis of the definitions in the SLE classification criteria.

Currently, the pedigree collection consists of three self-identified ethnic groups: EA, AA, and Hispanic. The same recruitment procedure was adopted for all families with SLE. From our collection (Lupus Multiplex Registry and Repository), we have ascertained 37 Hispanic families multiplex for SLE. On the basis of the demographic records of the Hispanic families, we found that the majority (76%) are Mexican American; 16% are Puerto Rican. Two independent replication samples consisted of 92 EA families (EA-1) and 55 AA families; another independent replication sample consisting of 82 EA families (EA-2) (table 1) was used to cross-validate the initial linkage results. Whereas all the families from the EA-1 and AA populations have been studied before in SLE genome scans (Moser et al. 1998; Gray-McGuire et al. 2000), the EA-2 families are reported here for the first time. All the linkage analyses were done using the families multiplex for SLE.

Genotyping and Error Checking

Genomic DNA was isolated, by use of standard methods, from peripheral blood cells, buccal cell swabs, mouth wash specimens, or EBV-transformed cell lines. Detailed methods for the genotyping of families had been described elsewhere (Moser et al. 1998; Gray-McGuire et al. 2000). A total of 318 microsatellite markers, with an average marker spacing of 11 cM (range 2–22 cM) and a heterozygosity of 76% (with a range of 56%–94%), was typed. We used the genetic map and intermarker distances from the Marshfield map. Positions of markers not in the Marshfield map were set by interpolation on the basis of physical distances. Prior to linkage analysis, all family relationships were confirmed using RELTEST (Olson 1999).

Statistical Methods for Linkage Analysis

In complex diseases, such as SLE, the parameters of the disease model (mode of inheritance, disease-allele fre-

quency, and penetrances of genotypes) are uncertain. We therefore assessed the evidence for linkage in the initial genome scan with a nonparametric, penetrance-independent, affected-only, allele-sharing method (Kong and Cox 1997). The linkage analysis was performed using the program ALLEGRO (Gudbjartsson et al. 2000), which is a modification of the GENEHUNTER (Kruglyak et al. 1996) package.

The nonparametric LOD score was obtained, in part, by maximizing the likelihood with respect to a scalar parameter, δ, that measures the amount of excess identical-by-descent sharing among affected relatives (with $\delta = 0$, corresponding to the null hypothesis of no linkage [Kong and Cox 1997; Gudbjartsson et al. 2000]). We used the S_{pairs} scoring function (Whittemore and Halpern 1994; Kruglyak et al. 1996) and the exponential allele-sharing model to generate the relevant statistic (Z_{lr}). Among the allele-sharing approaches, this scoring function has been found to perform well for a variety of underlying disease models (McPeek 1999). Finally, the *P* value was computed on the basis of large-sample theory; the distribution of $Z_{lr} = [2 \times \log_e(10)\text{LOD}]$ approximates a standard normal random variable under the null hypothesis (Kong and Cox 1997; Gudbjartsson et al. 2000). Individual family scores were combined to obtain an overall score, by use of a weighting scheme that, on the log scale, is halfway between weighting the families equally (Kruglyak et al. 1996) and weighting the affected relative pairs equally. This scheme gives weights similar to those proposed by Weeks and Lange (1988), as an extension of the scheme that was originally designed for sibships (Hodge 1984).

Upon establishing a linkage by exceeding the threshold for genomewide significance using an allele-sharing method, we fitted a range of parametric models to the data, in an effort to explain the underlying genetic model. We estimated the race-specific marker-allele frequencies from family founders, using FASTLINK (Cottingham et al. 1993). Parametric linkage analysis assuming linkage homogeneity (LOD) and linkage heterogeneity (HLOD) scores were obtained using both GENEHUNTER-PLUS (Kruglyak et al. 1996) and ALLEGRO (Gudbjartsson et al. 2000). Significance of a LOD score was determined by $\chi^2 = 4.6 \times \text{LOD}$, with 1 df. Since HLOD follows a complex statistical distribution, significance for the observed HLOD was first converted to a χ^2, where $\chi^2 = 4.6 \times \text{HLOD}$. *P* value ($P_1$) was then derived for χ^2 by use of the χ^2 distribution, with 1 df. The *P* value for the HLOD score was then $0.5 \times [1 - (1 - P_1)(1 - P_1)]$ (Faraway 1993).

We also performed the simulations to determine the proper significance of linkage results for the confirmatory group. Empirical *P* values for the complete data set were obtained by simulation of 10,000 sets of three markers (representing EA-2) not linked to a susceptibility locus generated by the ALLEGRO program (Gudbjartsson et al. 2000). Pedigree structures, phenotypic classifications, and genetic models were the same as those used in the actual analysis for EA-2, but marker information was generated without regard to the affection status. The best Z_{lr} score over each simulated data set was extracted and compiled into a single distribution. Z_{lr} from the real analysis was compared with this distribution to determine how frequently this result would be observed by chance in an unlinked data set. This was reported as the empirical *P* value.

Results

Clinical Characteristics of the Patients

The comparison of major family and demographic characteristics shows that, overall, the four groups are similar (table 1). We assessed the significance of the quantitative variables, using one-way ANOVA and post hoc Tukey test for multiple pairwise comparisons between groups. The mean age at onset for affected individuals did not vary significantly among the four groups ($F = 1.95$; $P = .12$). To see whether nonindependence or relatedness among the patients is a problem, given the nature of the family data, we randomly selected one patient with SLE per family and recalculated ANOVA. We found similar results ($F = 1.19$; $P = .31$). The post hoc pairwise comparison between groups also resulted in a similar conclusion.

Genome Scan for SLE Susceptibility

Initially, we performed a genome scan for SLE linkage on 37 Hispanic families on the basis of multipoint nonparametric linkage analysis (fig. 1). Seven different genomic locations yielded the Z_{lr} value >1.65 ($P < .05$) (table 2). Among them, three genomic regions produced evidence of linkage exceeding our predetermined threshold ($Z_{lr} > 2.32$; $P < .01$) for further evaluation. The peak at 12q24 ($Z_{lr} = 3.31$; $P = .00047$) was found on D12S395 at 130 cM, the peak at 16p13 was on D16S764 at 18 cM ($Z_{lr} = 2.62$; $P = .004$), and the peak at 16q12-21 ($Z_{lr} = 3.06$; $P = .001$) was between marker D16S3253 and D16S1624 at 66.8 cM (table 2). The last column of table 2 provides information about the genomic locations reported by other groups.

Linkage at 12q24

In an effort to reevaluate the validity of our initial results at 12q24, 16p13, and 16q12-21, we tested linkage on an independent data set of 147 families with SLE. This data set consisted of 92 EA families (EA-1) and 55 AA families. Since our a priori hypothesis of linkage to SLE at 12q24, 16p13, and 16q12-21 was fixed, we genotyped only three or four markers (peak marker or markers and

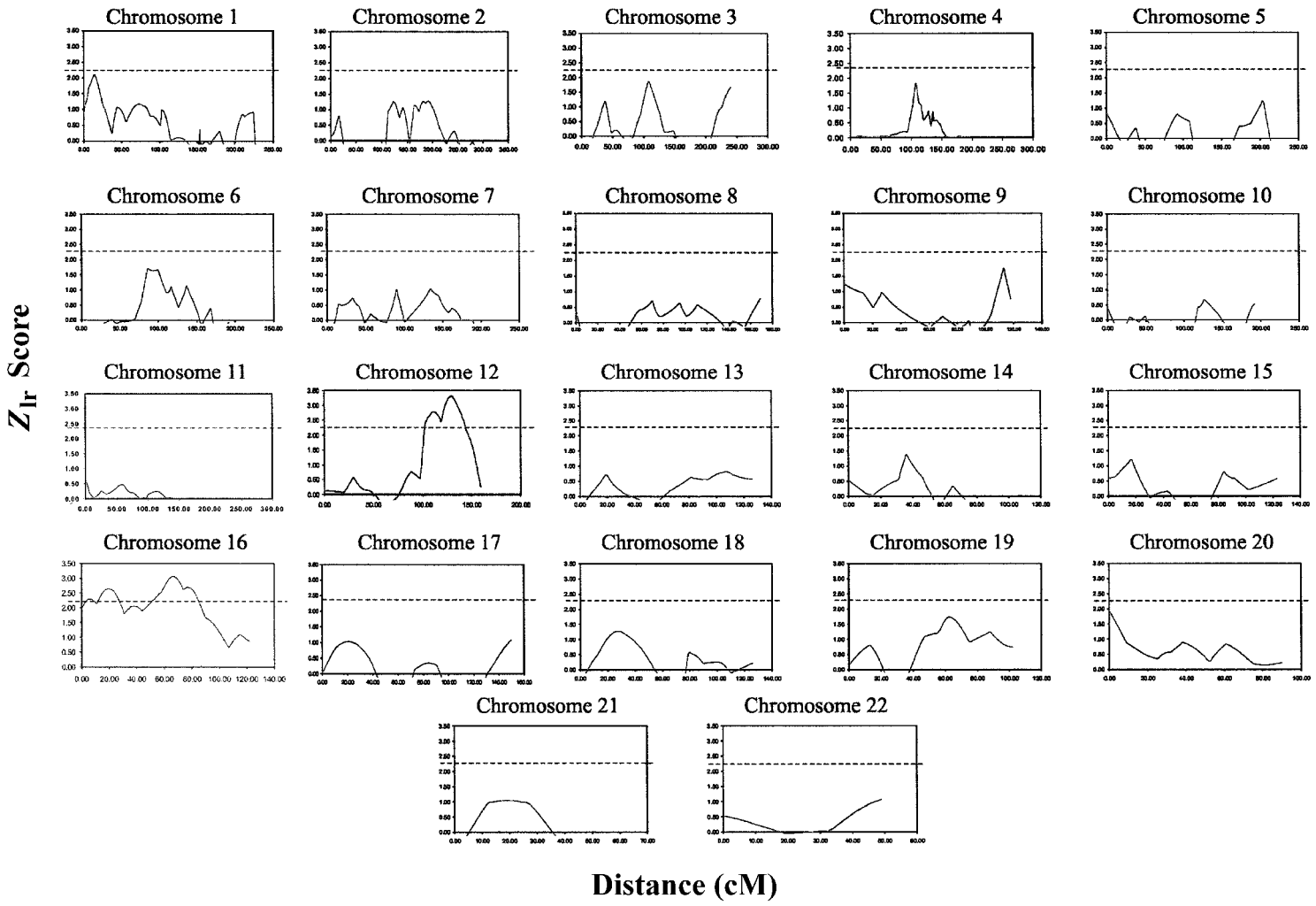

Figure 1 Genomewide scan for multipoint nonparametric linkage analysis for Hispanic families. The *Y*-axis indicates the Z_{lr} values, and the *X*-axis indicates the chromosome length in centimorgans. The horizontal dashed line indicates our predetermined threshold ($Z_{lr} > 2.32$; $P < .01$) for further exploration of linkage evidence in independent families.

Table 2

Markers Yielding Multipoint $Z_{lr} > 1.65$ ($P < .05$) from the Hispanic Genome Scan

	Linkage Peak				
Cytogenetic Location	Marker(s)	Position (cM)	Z_{lr}	P	Published Reports for Linkage
1p36	D1S552	14.8	2.11	.017	Gaffney et al. 2000; Johanneson et al. 2002
3p14	D3S2406	108.0	1.87	.031	
9q34	D9S2157	113.2	1.83	.034	
12q24	D12S395	130.3	3.31	.00047	
16p13	D16S764	18.0	2.62	.004	
16q12-21	D16S3253-D16S2624	66.8	3.06	.001	Shai et al. 1999; Gaffney et al. 2000; Tsao et al. 2002
20p13	D20S103	1.8	1.80	.036	Gaffney et al. 1998

the two flanking markers [one from either side]) used in the genome scan. The markers and intermarker distances (cM) used for linkage analysis at three genomic locations are as follows: 12q24 (D12S2070-13.9-D12S395-12.7-D12S2070), 16p13 (D16S748-7-D16S764-14-D16S403), and 16q12-21 (D16S-753-14-D16S3253-16-S2624-13-D16S516). On the basis of multipoint linkage analysis, the initial linkage signal at 12q24 was replicated in EA-1 ($Z_{lr} = 3.06$; $P = .0011$), but virtually no evidence for linkage was found in the AA families ($Z_{lr} = 0.37$; $P = .35$). The corresponding nonparametric LOD for EA-1 and AA were 2.03 and 0.03, respectively. However, on the basis of multipoint linkage analysis, the initial linkage signal at 16p13 was not replicated either in EA-1 families ($Z_{lr} = 0.38$; $P = .35$) or in AA families ($Z_{lr} = -0.59$; $P = .72$). Nonsignificant results were also observed for linkage at 16q12-21 in EA-1 ($Z_{lr} = 0.62$; $P = .27$) and in AA families ($Z_{lr} = 0.69$; $P = .24$).

We then combined our Hispanic families with EA-1 (a total of 129 families) and reanalyzed the linkage effect at 12q24 in the combined families. The evidence of linkage ($Z_{lr} = 4.39$; $P = 5.7 \times 10^{-6}$) at 12q24 exceeded the threshold for genomewide significance recommended in Lander and Kruglyak's (1995) guidelines for interpreting linkage results for complex diseases. The maximum evidence of linkage (peak LOD score) was found at D12S395 at 130.3 cM (fig. 2). The corresponding nonparametric LOD score was 4.19 at this marker. To our knowledge, significant evidence of any SLE susceptibility locus at 12q24 has not been described elsewhere. We designated this novel SLE susceptibility locus "SLEB4."

Parametric Model for Linkage at 12q24

Hodge and Elston (1994) showed that maximization of LOD scores over a range of genetic models (i.e., calculation of maximized LOD scores) is a valid procedure that does not require adjustment of significance levels or a correction for ascertainment. Further, they showed that evaluation for linkage and determination of the most likely genetic model for a data set may proceed simultaneously, provided there is a linked marker. Accordingly, after establishing the SLE linkage at 12q24 by the nonparametric allele-sharing method, we attempted to understand the genetic model (mode of inheritance) of this susceptibility locus by fitting a range of single-locus dominant, recessive, and additive models to the data (table 3).

The results of multipoint parametric linkage, under the assumption of genetic homogeneity and heterogeneity for the best-fitted genetic model, are shown in table 4 for individual and combined group analyses. The best-fitted, most-parsimonious genetic model was obtained as a common dominant allele (disease-allele frequency = 0.2) with low penetrances of 0.0005 for the susceptibility genotypes. The estimated prevalence was 0.00018, which is close to published reports (Hochberg 1997*a;* Lawrence et al. 1998). The other genetic models were substantially inferior to the most-parsimonious model (results not shown). The model separately explained the inheritance in the Hispanic families and EA-1 families surprisingly well. The maximum LOD was virtually identical to the HLOD for both the Hispanic and EA-1 ($\alpha = 1.0$; i.e., 100% of the families were linked to this locus under this model). The peak linkage for both individual ethnic groups, as well as for the combined groups, was at approximately the same position (fig. 2). We also analyzed all 147 (EA-1 and AA) replication families together. The LOD, HLOD, and α decreased to 1.05, 1.37, and 0.66, respectively. This was certainly due to the genetic heterogeneity between the two ethnic groups.

Confirmation of Linkage at 12q24

Independent replication is typically viewed as the sine qua non for acceptance of a hypothesis. Replicating an established linkage applies the scientific method and a gold standard too often not obtained in the linkage studies of genetically complex phenotypes. Lander and Kruglyak (1995) suggested a general guideline that the term "confirmation/replication study" should be reserved for those situations in which significant evidence of linkage has already been obtained in an initial study. Accordingly, our next goal was to confirm this linkage in an independent data set by use of similar parametric and nonparametric

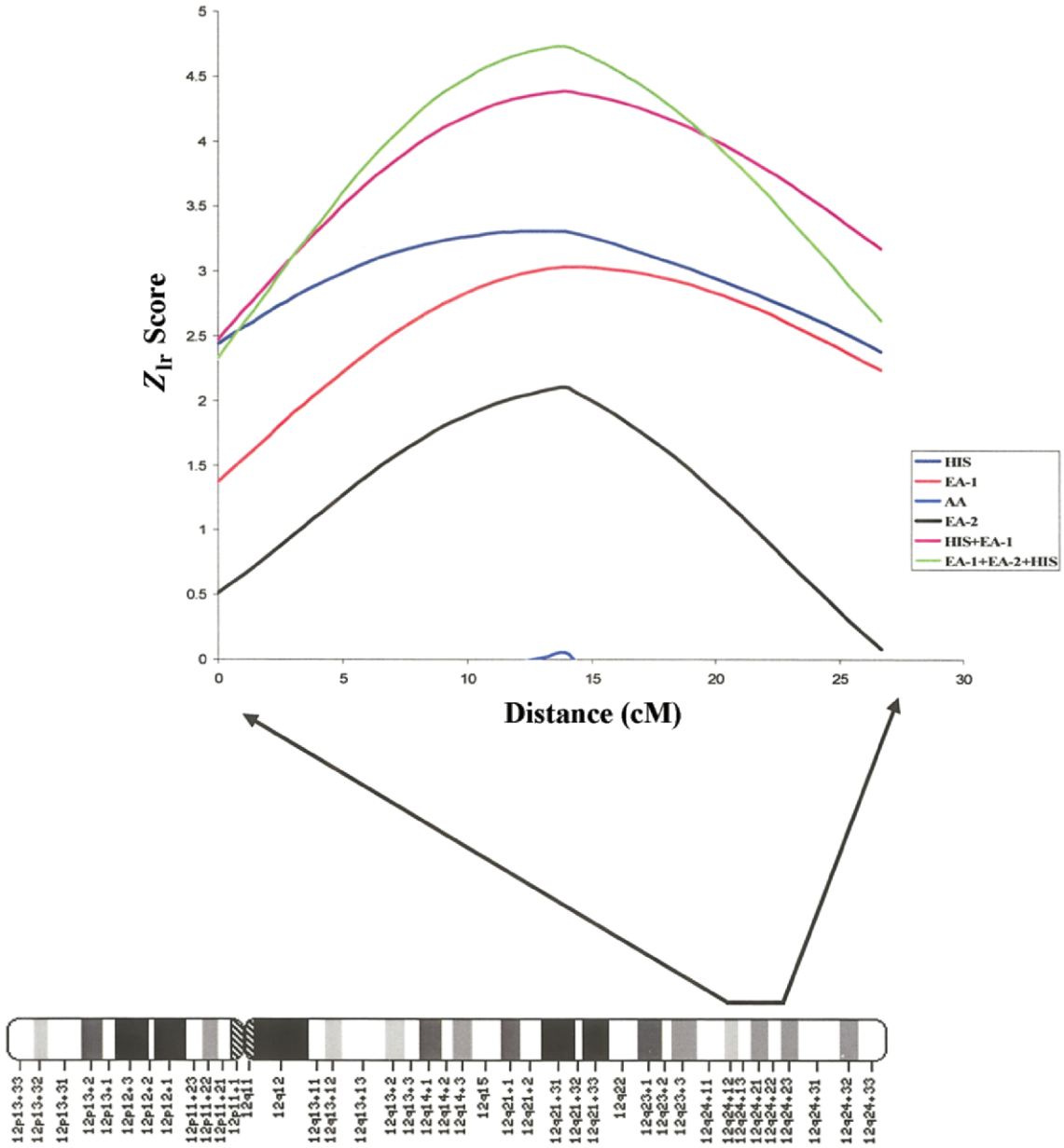

Figure 2 Results of multipoint linkage analysis at 12q24 region for different groups. The *Y*-axis indicates the Z_{lr} score, and the *X*-axis indicates the chromosomal length spanned by three markers. The three markers and their intermarker distances (cM) are as follows: D12S2070-13.9-D12S395-12.7-D12S2078.

analyses. Since we established elsewhere the SLE susceptibility locus at this region (with genomewide significance), our a priori hypothesis of linkage was fixed at 12q24. In this situation, the multiple-testing problem associated with the genomewide search does not apply; hence, to replicate the initial established linkage signal, the significance of linkage would be set for $P < .05$ (equivalent to $Z_{lr} > 1.64$ and LOD > 0.59), if no fine mapping has been performed (Thomson 1994; Lander and Kruglyak 1995; Ott 1999; Robert Elston, personal communication).

For this purpose, we analyzed another independent data set, which consisted of 82 EA families with SLE (EA-2). On the basis of multipoint linkage analysis with three genome scan markers, we detected the maximum evidence of linkage on marker D12S395 in EA-2 by nonparametric ($Z_{lr} = 2.11$; $P = .017$) and parametric (LOD = 0.86, $P = .023$; HLOD = 1.01, $P = .03$; $\alpha = 0.73$) analyses (fig. 2; table 4). On the basis of 10,000 simulations, the empirical P value for the observed Z_{lr} was 0.014, thereby individually replicating the linkage at 12q24.

Table 3

Parameter Values for Five Different Plausible Genetic Models Used in Parametric Linkage Analysis

Genetic Model	Disease allele (A) frequency	Penetrances for Genotypes		
		AA	Aa	aa
High penetrant, rare recessive	.02	.9	.0	.0
Low penetrant, common recessive	.2	.008	.0	.0
High penetrant, rare dominant	.0002	.9	.9	.0
Low penetrant, common dominant	.2	.0005	.0005	.0
Low penetrant, rare additive	.01	.04	.02	.0

Discussion

Genetic approaches now available to evaluate complex diseases such as SLE offer great potential for establishing linkage and then for identifying the disease gene(s). Previous SLE genome scans have proposed several candidate regions for SLE susceptibility loci, some of which have been independently confirmed upon replication. The linkage at 12q24 now joins that group, which includes (along with the associated candidate gene, when known) 1q23 (FcγRIIIA), 2q37 (PCDA-1), 4p15, 6p21 (HLA-DR), and 16q13 (Gaffney et al. 1998, 2000; Moser et al. 1998; Shai et al. 1999; Gray-McGuire et al. 2000; Lindquivst et al. 2000; Edberg et al. 2002; Graham et al. 2002; Kelly et al. 2002; Nath et al. 2002; Prokunina et al. 2002; Tsao et al. 2002). These "confirmed" genetic effects are very unlikely to be false positives, whereas validity of the various other "unconfirmed" genetic linkages that have not achieved this standard are less certain to be true SLE susceptibility loci (Harley 2002; Kelly et al. 2002).

To identify novel susceptibility loci for SLE, we have conducted a whole-genome scan in a group of Hispanic families. Since parametric analyses are model dependent and since misspecification of the disease model may lead to a decrease in power (Clerget-Darpoux et al. 1986), we used model-independent nonparametric analysis for the initial genome scan. Our initial genome scan on Hispanic families yielded two genomic regions (12q24 and 16q12-21) with suggestive evidence ($P \leq .001$) of linkage and a nearly suggestive ($P = .004$) evidence of linkage at 16p13. Following the method of Lander and Kruglyak (1995), we pursued these linkages in extension studies, in which data are combined in an effort to achieve significant evidence for linkage. Assuming that common susceptibility loci would influence the SLE phenotype in other non-Hispanic populations, we used the AA and EA families available to us for replication.

Significant decrease of linkage evidence at 12q24 was found when we analyzed all 147 (92 EA-1 and 55 AA) families with SLE in the replication data set. On the other hand, a significantly increased linkage was present in EA-1, thereby showing linkage heterogeneity at 12q24. When we combined the Hispanic and EA-1 families, the evidence of linkage exceeded the threshold for genome-wide significance.

Once we established significant evidence of linkage at 12q24 by a nonparametric method, we sought a plausible genetic transmission model by parametric linkage analysis. With a complex disease such as SLE, none of these simple models are likely to be exactly true, and the effect of a gene and its variants can be reliably determined only after the at-risk variant(s) has been identified. However, the fitted parametric models approximate the contribution of this particular gene to the familial clustering of the disease. Whereas the power to detect linkage by parametric LOD score analysis is sensitive to misspecification of the genetic model, it has been shown that maximization of LOD scores over multiple genetic models increases the power to detect linkage when the true mode of inheritance of the disease is unknown (Hodge et al. 1997). We explored a range of plausible single-locus genetic models (table 3). The most-parsimonious genetic model was obtained as a low-penetrant, common, dominant allele. As predicted elsewhere (Lander 1996; Chakravarti 1999), the alleles at any single susceptibility gene are more likely to have only small-to-moderate contributions to the total risk of common complex phenotypes, which has been demonstrated by our derived genetic model obtained at 12q24.

Next, we evaluated linkage at 12q24 by another set of 82 independently ascertained families with SLE (EA-2). Replication of initial linkage signals from independent samples is a crucial step toward distinguishing between true positives and false positives (Lander and Kruglyak 1995). However, replication of linkage results for complex traits has been extremely difficult, owing in part to the inability to precisely identify the underlying phenotype, small sample sizes and small multiplex pedigrees, genetic heterogeneity, inaccurate genetic models, and the poor discriminating capacity of the statistical methods employed in analysis (Suarez et al. 1994). Moreover, the considerable heterogeneity between data sets, both within and between populations and ethnic groups, must relatively impede efforts to confirm linkage. None-

Table 4

Results of Multipoint Parametric Model-Based (Most-Parsimonious Model) Linkage Analysis

Population	No. of Families	LOD	HLOD	α[a]
Hispanic	37	1.87	1.87	1.00
EA-1	92	1.86	1.86	1.00
Hispanic + EA-1	129	3.72	3.72	1.00
EA-2	82	0.96	1.01	.73
EA-1 + EA-2	174	2.78	2.80	.90
Hispanic + EA-1 + EA-2	211	4.60	4.61	.95

Note.—Disease-allele frequency and the penetrances for the three genotypes (AA, Aa, and aa, where A is the disease allele) were .2, .0005, .0005, and .000, respectively.

[a] α = proportion of families linked.

theless, we replicated the evidence for linkage at 12q24 by nonparametric, parametric, and simulation analyses. We found that EA-2, unlike the Hispanic or EA-1 families, displayed substantially greater heterogeneity (~27% of EA-2 families are not linked at 12q24).

It is interesting and important that Gaffney et al. (2000) also reported the evidence of linkage at 12q24 in EA families. On the basis of study of the 65 EA families, Gaffney et al. (2000) found evidence of linkage ($Z_{lr} = 2.24$; LOD = 1.09; $P = .01$) at 12q24. The peak they found was at 134.5 cM, whereas our peak is at 130.3 cM. This is not an unexpected shift for a complex disease linkage. Therefore, on the basis of establishing and confirming linkage in our *independent* data set and a second confirmation in the published reports, this linkage at 12q24 should be considered as definite evidence of replication.

The linkage at 12q24 clearly appears to be race specific. The evidence for linkage primarily came from Hispanic and EA families. Virtually no evidence of linkage was found in the AA families. The linkage evidence at 12q24 observed in family EA-2 was smaller than in family EA-1. This can be explained by disease heterogeneity and ascertainment of the families. The disease heterogeneity of SLE, similar to that of many other complex genetics disorders, influences ascertainment; hence, genetic effects are expected to be concentrated in any given sample of the pedigrees on a random basis. Consequently, pedigree contribution to the aggregate result is expected to be smaller in the pedigrees used for replication. Suarez et al. (1994) demonstrated that if the same criteria for significance are applied to localizing a gene and to replicating its localization, approximately five to six times as many families are required for replication as are required for an initial linkage finding. Moreover, the estimated proportion of linkage heterogeneity in family EA-2 is high (~27%). Simulation studies predict that "nonreplications" of true linkages will be common when small, heterogeneous samples are used to detect genes of small effect. For example, 600–1,000 affected sibling pairs would be required to reliably demonstrate locus-specific genetic effects causing a 27%–30% populationwide increase in risk to siblings (Hauser et al. 1996); moreover, multiplicative gene effects are also expected, which should be even more difficult to detect (Rybicki and Elston 2000).

Our previous SLE genome scans identified two major SLE susceptibility loci, 1q23 (Moser et al. 1998) and 4p16 (Gray-McGuire et al. 2000). It is interesting that these evidences are not reproduced in the present study. This may be due to two factors. First, the major linkage effects identified from previous genome scans are mainly based on AA (Moser et al. 1998) or EA (Gray-McGuire et al. 2000) families. The present genome scan was done on Hispanic families. Second, the sample size for Hispanics (37 families) may not be sufficient to detect these loci identified elsewhere, especially for genetically heterogeneous disease (Suarez et al. 1994).

We found the maximum evidence of linkage ($P = .001$) at 66.8 cM in 16q12-21, between markers D16S3253 and D16S2624, in the initial genome scan of Hispanic families. Two earlier findings from Minnesota (Gaffney et al. 2000) and California (Shai et al. 1999), reported the SLE linkage at 16q13. Whereas the Minnesota group found significant evidence ($P = 1.3 \times 10^{-5}$) for linkage at 67.2 cM on marker D16S415 (current Marshfield map) in EA families, the California group found modest evidence ($P = .017$) for linkage to marker D16S3136 at 62.1 cM in Mexican American and EA families. According to the Lander and Kruglyak (1995) criteria for replication and confirmation of a previously established linkage, our results from the Hispanic families confirmed this linkage at 16q12. However, we could not replicate the evidence of linkage at 16q12-21 in our replication group. Presence of genetic heterogeneity is considered to be the most likely explanation. From our study, it has again become clear that, in linkage studies of complex disorders, determining whether a given study has replicated the findings of an initial study is not a trivial task.

In this analysis, the purpose of dividing our family collections into subsets was to increase genetic homogeneity of the families with respect to the specific putative susceptibility locus in which linkage can be established and replicated. Selecting the Hispanic families for this purpose involves a shared cultural experience, implying environmental influences, perhaps, more than a shared genetic population of origin among Hispanics. In the United States, Hispanics are variably admixed with populations originating in Europe, the Americas, and Africa. Consequently, a genetic effect found in Hispanics would be expected to be present in at least one of these originating groups. Finding the 12q24 linkage in EAs may fulfill this prediction.

Our current candidate interval spans a region of ~15 Mb across chromosomal bands 12q24.1 (D12S2070) to 12q24.3 (D12S2078) and contains 120 known genes, several partial transcripts, and several predicted transcripts. Several biologically plausible candidate genes are located within this chromosomal region. (The candidate-gene locations were obtained from the April 2003 freeze at the University of California–Santa Cruz "Golden Path" Web site.) For example, NOS1 (nitric oxide synthase 1), which showed association with asthma and related phenotypes (Grasemann 2001; Immervoll et al. 2001); TCF1 (transcription factor 1) associated with type 2 diabetes (Bulman et al. 2002); and IFN-γ (interferon-γ) associated with asthma and atopy (Barnes et al. 1999) are all located at 12q24. Linkage signals have also been found in this region for other inflammatory diseases, in-

cluding asthma (Yokouchi et al. 2002), diabetes (Mahtani et al. 1996; Ehm et al. 2000), and Graves disease (Siegmund et al. 1998). Identification of the susceptibility gene by the positional candidate cloning approach will require study of many more families and further narrowing of the candidate region.

In conclusion, our results provide strong evidence of the existence of a major SLE susceptibility gene, SLEB4 on 12q24, that influences the SLE phenotypic variations, especially in Hispanics and EAs. In addition, we confirmed the evidence of an already established SLE susceptibility locus located on chromosome 16q12-21. The confirmation of these linkages at 12q24 and 16q12-21 to SLE strongly encourages efforts to identify the responsible susceptibility genes through positional cloning.

Acknowledgments

We thank the participating patients and family members, as well as their referring physicians. We thank Drs. Christopher E. Aston, Michael L. Frigge, and Partha P. Majumder for helpful suggestions and comments. Some of the pedigrees were obtained from the Lupus Multiplex Registry and Repository (supported by National Institutes of Health [NIH] grant AR12253). The study was supported by the NIH grants AR048928, AR42460, AI24717, AR048940, AR049084, AR15577, and AI31584 and by the United States Department of Veterans Affairs.

Electronic-Database Information

URLs for data presented herein are as follows:

Center for Medical Genetics, Marshfield Medical Research Foundation, http://research.marshfieldclinic.org/genetics/

Lupus Multiplex Registry and Repository, http://omrf.ouhsc.edu/

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for SLE)

University of California–Santa Cruz Web Site, http://genome.ucsc.edu/ (for Golden Path)

References

Barnes KC, Freidhoff LR, Nickel R, Chiu YF, Juo SH, Hizawa N, Naidu RP, Ehrlich E, Duffy DL, Schou C, Levett PN, Marsh DG, Beaty TH (1999) Dense mapping of chromosome 12q13.12-q23.3 and linkage to asthma and atopy. J Allergy Clin Immunol 104:485–491

Bulman MP, Harries LW, Hansen T, Shepherd M, Kelly WF, Hattersley AT, Ellard S (2002) Abnormal splicing of hepatocyte nuclear factor 1 alpha in maturity-onset diabetes of the young. Diabetologia 45:1463–1467

Chakravarti A (1999) Population genetics: making sense out of sequence. Nat Genet 21:56–60

Clerget-Darpoux F, Bonaiti-Pellie C, Hochez J (1986) Effects of misspecifying genetic parameters in lod score analysis. Biometrics 42:393–399

Cottingham RW Jr, Idury RM, Schaffer AA (1993) Faster sequential genetic linkage computations. Am J Hum Genet 53:252–263

Deapen D, Escalante A, Weinrib L, Horwitz D, Bachman B, Roy-Burman P, Walker A, Mack TM (1992) A revised estimate of twin concordance in systemic lupus erythematosus. Arthritis Rheum 35:311–318

Edberg JC, Langefeld CD, Wu J, Moser KL, Kaufman KM, Kelly J, Bansal V, Brown WM, Salmon JE, Rich SS, Harley JB, Kimberly RP (2002) Genetic linkage and association of Fcγ receptor IIIA (CD16A) on chromosome 1q23 with human systemic lupus erythematosus. Arthritis Rheum 46:2132–2140

Ehm MG, Karnoub MC, Sakul H, Gottschalk K, Holt DC, Weber JL, Vaske D, Briley D, Briley L, Kopf J, McMillen P, Nguyen Q, Reisman M, Lai EH, Joslyn G, Shepherd NS, Bell C, Wagner MJ, Burns DK, American Diabetes Association GENNID Study Group (2000) Genomewide search for type 2 diabetes susceptibility genes in four American populations. Am J Hum Genet 66:1871–1881

Faraway JJ (1993) Distribution of the admixture test for the detection of linkage under heterogeneity. Genet Epidemiol 10: 75–83

Gaffney PM, Kearns GM, Shark KB, Ortmann WA, Selby SA, Malmgren ML, Rohlf KE, Ockenden TC, Messner RP, King RA, Rich SS, Behrens TW (1998) A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families. Proc Natl Acad Sci USA 95:14875–14879

Gaffney PM, Ortmann WA, Selby SA, Shark KB, Ockenden TC, Rohlf KE, Walgrave NL, Boyum WP, Malmgren ML, Miller ME, Kearns GM, Messner RP, King RA, Rich SS, Behrens TW (2000) Genome screening in human systemic lupus erythematosus: results from a second Minnesota cohort and combined analyses of 187 sib-pair families. Am J Hum Genet 66:547–556

Graham RR, Ortmann WA, Langefeld CD, Jawaheer D, Selby SA, Rodine PR, Baechler EC, Rohlf KE, Shark KB, Espe KJ, Green LE, Nair RP, Stuart PE, Elder JT, King RA, Moser KL, Gaffney PM, Bugawan TL, Erlich HA, Rich SS, Gregersen PK, Behrens TW (2002) Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus. Am J Hum Genet 71:543–553

Grasemann H (2001) Genetics of the neuronal NO synthase (NOS1) in the etiology of bronchial asthma. Pneumologie 55: 390–395

Gray-McGuire C, Moser KL, Gaffney PM, Kelly J, Yu H, Olson JM, Jedrey CM, Jacobs KB, Kimberly RP, Neas BR, Rich SS, Behrens TW, Harley JB (2000) Genome scan of human systemic lupus erythematosus by regression modeling: evidence of linkage and epistasis at 4p16-15.2. Am J Hum Genet 67:1460–1469

Gudbjartsson DF, Jonasson K, Frigge ML, Kong A (2000) Allegro, a new computer program for multipoint linkage analysis. Nat Genet 25:12–13

Harley JB (2002) The genetic etiology of systemic lupus erythematosus: a short dispatch from the combat zone. Genes Immun Suppl 1:1–4

Hauser ER, Boehnke M, Guo SW, Risch N (1996) Affected-sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. Genet Epidemiol 13:117–137

Hochberg MC (1997*a*) The epidemiology of systemic lupus erythematosus. In: Wallace DJ, Hahn BH (eds) Dubois' lupus erythematosus. Williams and Wilkins, Baltimore, pp 49–65

——— (1997*b*) Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 40:1725

Hodge SE (1984) The information contained in multiple sibling pairs. Genet Epidemiol 1:109–122

Hodge SE, Abreu PC, Greenberg DA (1997) Magnitude of type I error when single-locus linkage analysis is maximized over models: a simulation study. Am J Hum Genet 60:217–227

Hodge SE, Elston RC (1994) Lods, wrods, and mods: the interpretation of lod scores calculated under different models. Genet Epidemiol 11:329–342

Immervoll T, Loesgen S, Dutsch G, Gohlke H, Herbon N, Klugbauer S, Dempfle A, Bickeboller H, Becker-Follmann J, Ruschendorf F, Saar K, Reis A, Wichmann HE, Wjst M (2001) Fine mapping and single nucleotide polymorphism association results of candidate genes for asthma and related phenotypes. Hum Mutat 18:327–336

Johanneson B, Lima G, von Salomé J, Alarcón-Segovia D, the Collaborative Group on the Genetics of SLE, the BIOMED II Collaboration on the Genetics of SLE and Sjögrens syndrome, Alarcón-Riquelme ME (2002) A major susceptibility locus for systemic lupus erythemathosus maps to chromosome 1q31. Am J Hum Genet 71:1060–1071

Kelly JA, Moser KL, Harley JB (2002) The genetics of systemic lupus erythematosus: putting the pieces together. Genes Immunol Suppl 1:71–85

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Lander ES (1996) The new genomics: global views of biology. Science 274:536–539

Lander ES, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Lawrence RC, Helmick CG, Arnett FC, Deyo RA, Felson DT, Giannini EH, Heyse SP, Hirsch R, Hochberg MC, Hunder GG, Liang MH, Pillemer SR, Steen VD, Wolfe F (1998) Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis Rheum 41: 778–799

Mahtani MM, Widen E, Lehto M, Thomas J, McCarthy M, Brayer J, Bryant B, Chan G, Daly M, Forsblom C, Kanninen T, Kirby A, Kruglyak L, Munnelly K, Parkkonen M, Reeve-Daly MP, Weaver A, Brettin T, Duyk G, Lander ES, Groop LC (1996) Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. Nat Genet 14:90–94

McPeek MS (1999) Optimal allele-sharing statistics for genetic mapping using affected relatives. Genet Epidemiol 16:225–249

Moser KL, Neas BR, Salmon JE, Yu H, Gray-McGuire C, Asundi N, Bruner GR, Fox J, Kelly J, Henshall S, Bacino D, Dietz M, Hogue R, Koelsch G, Nightingale L, Shaver T, Abdou NI, Albert DA, Carson C, Petri M, Treadwell EL, James JA, Harley JB (1998) Genome scan of human systemic lupus erythematosus: evidence for linkage on chromosome 1q in African-American pedigrees. Proc Natl Acad Sci USA 95: 14869–14874

Nath SK, Kelly JA, Reid J, Lam T, Gray-McGuire C, Aston CE, Harley JB (2002) SLEB3 in systemic lupus erythematosus (SLE) is strongly related to SLE families ascertained through neuropsychiatric manifestations. Hum Genet 111:54–58

Olson JM (1999) Relationship estimation by Markov-process models in a sib-pair linkage study. Am J Hum Genet 64:1464–1472

Ott J (1999) Analysis of human genetics linkage, 3rd ed. The Johns Hopkins University Press, Baltimore, pp 321–322

Petri M (1998) The effect of race on incidence and clinical course in systemic lupus erythematosus: the Hopkins Lupus Cohort. J Am Med Womens Assoc 53:9–12

Prokunina L, Castillejo-Lopez C, Oberg F, Gunnarsson I, Berg L, Magnusson V, Brookes AJ, Tentler D, Kristjansdottir H, Grondal G, Bolstad AI, Svenungsson E, Lundberg I, Sturfelt G, Jonssen A, Truedsson L, Lima G, Alcocer-Varela J, Jonsson R, Gyllensten UB, Harley JB, Alarcon-Segovia D, Steinsson K, Alarcon-Riquelme ME (2002) A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. Nat Genet 32:666–669

Rybicki BA, Elston RC (2000) The relationship between the sibling recurrence-risk ratio and genotype relative risk. Am J Hum Genet 66:593–604

Shai R, Quismorio FP Jr, Li L, Kwon OJ, Morrison J, Wallace DJ, Neuwelt CM, Brautbar C, Gauderman WJ, Jacob CO (1999) Genome-wide screen for systemic lupus erythematosus susceptibility genes in multiplex families. Hum Mol Genet 8: 639–644

Siegmund T, Usadel KH, Donner H, Braun J, Walfish PG, Badenhoop K (1998) Interferon-gamma gene microsatellite polymorphisms in patients with Graves' disease. Thyroid 8: 1013–1017

Suarez BK, Hampe CL, Van Eerdewegh P (1994) Problems of replicating linkage claims in psychiatry. In: Gershon ES, Cloninger CR (eds) Genetic approaches to mental disorders. American Psychiatric Press, Washington, DC, pp 23–46

Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield, NF, Schaller JG, Talal, N, Winchester RM (1982) The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 25:1271–1277

Thomson G (1994) Identifying complex disease genes: progress and paradigms. Nat Genet 8:108–110

Tsao BP (2003) The genetics of human systemic lupus erythematosus. Trends Immunol 24:595–602

Wandstrat A, Wakeland E (2001) The genetics of complex autoimmune diseases: non-MHC susceptibility genes. Nat Immunol 2:802–809

Weeks DE, Lange K (1988) The affected-pedigree-member method of linkage analysis. Am J Hum Genet 42:315–326

Whittemore AS, Halpern J (1994) A class of tests for linkage using affected pedigree members. Biometrics 50:118–127

Yokouchi Y, Shibasaki M, Noguchi E, Nakayama J, Ohtsuki T, Kamioka M, Yamakawa-Kobayashi K, Ito S, Takeda K, Ichikawa K, Nukaga Y, Matsui A, Hamaguchi H, Arinami T (2002) A genome-wide linkage analysis of orchard grass-sensitive childhood seasonal allergic rhinitis in Japanese families. Genes Immun 3:9–13