Screening the Genome for Rheumatoid Arthritis Susceptibility Genes

A Replication Study and Combined Analysis of 512 Multicase Families

Damini Jawaheer,[1] Michael F. Seldin,[2] Christopher I. Amos,[3] Wei V. Chen,[3] Russell Shigeta,[2] Carol Etzel,[3] Aarti Damle,[1] Xiangli Xiao,[1] Dong Chen,[2] Raymond F. Lum,[4] Joanita Monteiro,[1] Marlene Kern,[1] Lindsey A. Criswell,[4] Salvatore Albani,[5] J. Lee Nelson,[6] Daniel O. Clegg,[7] Richard Pope,[8] Harry W. Schroeder, Jr.,[9] S. Louis Bridges, Jr.,[9] David S. Pisetsky,[10] Ryk Ward,[11] Daniel L. Kastner,[12] Ronald L. Wilder,[13] Theodore Pincus,[14] Leigh F. Callahan,[15] Donald Flemming,[16] Mark H. Wener,[17] and Peter K. Gregersen,[1] for the North American Rheumatoid Arthritis Consortium

Objective. **A number of non-HLA loci that have shown evidence ($P < 0.05$) for linkage with rheumatoid arthritis (RA) have been previously identified. The present study attempts to confirm these findings.**

Methods. **We performed a second genome-wide screen of 256 new multicase RA families recruited from across the United States by the North American Rheumatoid Arthritis Consortium. Affected sibling pair analysis on the new data set was performed using SIBPAL. We subsequently combined our first and second data sets in an attempt to enhance the evidence for linkages in a larger sample size. We also evaluated the impact of covariates on the support for linkage, using LODPAL.**

Results. **Evidence of linkage at 1p13 (D1S1631), 6p21.3 (the HLA complex), and 18q21 (D18S858) ($P < 0.05$) was replicated in this independent data set. In addition, there was new evidence for linkage at 9p22 (D9S1121 [$P = 0.001$]) and 10q21 (D10S1221 [$P = 0.0002$] and D10S1225 [$P = 0.0038$]) in the current data set. The combined analysis of both data sets (512 families) showed evidence for linkage at the level of $P < 0.005$ at 1p13 (D1S1631), 1q43 (D1S235), 6q21 (D6S2410), 10q21 (D10S1221), 12q12 (D12S398), 17p13 (D17S1298), and 18q21 (D18S858). Linkage at HLA was also confirmed ($P < 5 \times 10^{-12}$). Inclusion of DRB1∗04 as a covariate significantly increased the probability of linkage on chromosome 6. In addition, some linkages on chromosome 1 showed improved significance when modeling DRB1∗04 or rheumatoid factor positivity as covariates.**

Conclusion. **These results provide a rational basis for pursuing high-density linkage and association studies of RA in several regions outside of the HLA region, particularly on chromosomes 1p, 1q, and 18q.**

Rheumatoid arthritis (RA) is a chronic, systemic, inflammatory disease with a complex genetic component. It is the most common cause of inflammatory polyarthritis in adults. The aggregation of RA in families with several autoimmune disorders (1) and the frequent presence of autoantibodies in the serum of RA patients (2) suggests that the disease has an autoimmune pathogenesis, although the precise etiology is unknown. An HLA (human leukocyte antigen) association with RA has been known for over two decades (3) and has been confirmed in numerous population studies (4). As shown by 3 separate genome-wide screens (5–7), genes in the HLA region provide the largest genetic contribution to RA. However, twin and family studies indicate that non-HLA genes are also involved in disease susceptibility (8).

In an attempt to identify the non-HLA loci involved in RA susceptibility, 4 genome-wide scans have been performed by different groups (5–7,9) in studies involving RA multiplex families with affected sibling pairs. These studies have identified a number of non-HLA loci that demonstrate modest evidence of linkage with RA. However, despite a certain degree of overlap between these results, there is no locus outside of the HLA region that has yet been unambiguously identified. This may be due to the fact that the sample sizes have not been large enough to provide statistical power to detect loci that have only modest effects (10).

In our initial genome-wide screen of 257 RA families (6), we identified a number of non-HLA loci (on chromosomes 1, 4, 12, 16, and 17) associated with RA risk ($P < 5 \times 10^{-3}$). We have now performed a second independent genome-wide screen on a new panel of 256 RA multicase families. The results of this second analysis have allowed us to replicate evidence for linkage in several regions outside of the HLA locus. In addition, as suggested by Lander and Kruglyak (11), we have combined our current data set with that from our initial genome screen in an attempt to further enhance the evidence for linkage. The combined data set contains 677 affected sibling pairs (581 independent sibling pairs, using the Hodge correction for multiple affected siblings). Thus, our study constitutes the largest collection of RA sibling pairs reported to date.

SUBJECTS AND METHODS

North American Rheumatoid Arthritis Consortium (NARAC) family collection. Multiplex RA families were recruited nationwide in the United States through the 12 participating recruitment centers of NARAC. Informed consent was obtained from every subject, including all participating family members, and the local institutional review board's approval was secured at every recruitment site prior to the start of enrollment. The criteria for entering a family into the study were 1) ≥2 siblings satisfied the 1987 American College of Rheumatology (formerly, the American Rheumatism Association) criteria for RA (12), 2) at least 1 of the siblings had documented erosions on hand radiographs, and 3) at least 1 of the siblings had disease onset between the ages of 18 and 60 years. The presence of psoriasis, inflammatory bowel disease, or systemic lupus erythematosus (SLE) was an exclusionary criterion in affected individuals.

Radiographs of both hands were obtained from all affected siblings at the time of recruitment into the study, unless films obtained within the prior 2 years were available. All of the radiographs were scored by a single radiologist (DF) who utilized a preliminary, locally developed severity score on a scale of 0–5. The following criteria were utilized for this scale: 0 = no erosion, 1 = subtle erosion, 2 = erosion without joint space loss, 3 = erosion with joint space loss, 4 = complete joint space loss without marked bone destruction or joint deformity, and 5 = marked destruction with or without joint deformity. All patients were examined by trained NARAC personnel and/or consulting rheumatologists at the time of entry and were scored for the presence of joint swelling and tenderness, joint alignment, and range of motion according to standard published methods (13).

Laboratory procedures. Blood samples were collected from the affected siblings and, if available, from their parents, for analysis of DNA and serum. DNA was isolated from the peripheral blood mononuclear cells using a salting-out kit from BIO-101 (Carlsbad, CA). Rheumatoid factor was measured in each affected sibling at the University of Washington, Department of Laboratory Medicine Immunology Division, by latex-enhanced nephelometric assay (Behring Diagnostics, San Jose, CA), which uses human and rabbit IgG coated on latex beads as antigen. This assay was calibrated to the World Health Organization international standard for rheumatoid factor (14). Broad-level HLA–DRB1 typing and high-resolution DRB1*04 typing were accomplished by initial polymerase chain reaction (PCR) amplification of groups of alleles (all DRB1 alleles for broad-level typing, and group-specific amplification for DRB1*04 alleles) using biotinylated PCR primers, followed by hybridization to immobilized sequence-specific oligonucleotide probes in a linear array format. Positive hybridization reactions were detected using a streptavidin–horseradish peroxidase conjugate and a soluble colorless substrate, 3,3′,5,5′-tetramethylbenzidine (15). A computer algorithm based on the sequence-specific oligonucleotide probe hybridization pattern and the Anthony Nolan 1999 HLA sequence database (available at the World Wide Web site http://www.ebi.ac.uk/imgt/hla/nomen.html) were used to assign genotypes to each sample.

Of the 512 multiplex RA families used in this study, 256 families had been analyzed previously in our first

genome-wide screen, which we will refer to as "screen 1" (6). (Although we originally reported on the analysis of 257 families in our first genome screen, reevaluation of hand radiographs from 1 family led us to exclude that family from the current combined analysis due to the absence of clear-cut erosions in either affected sibling.) DNA samples from affected sibling pairs and parents (where available) from the 256 new RA families were genotyped in the present study using markers spaced at ~10-cM intervals throughout the entire genome; we will refer to this second genome-wide screen as "screen 2." All markers used were the same as those analyzed in screen 1. These were from the set8A combo list from the Marshfield World Wide Web site (available at http://www.marshmed.org/genetics/). Some additional markers were added to the set at certain chromosomal locations, such as the HLA complex. The entire marker set consisted of 47 panels (379 markers), with each panel containing markers pooled together according to size and fluorescent label (6-FAM, HEX, and NED). Reaction conditions were standard for all markers used, as described in the Marshfield PCR protocol (available at the Marshfield Web site). Each panel of pooled markers was electrophoresed on a 3700 DNA Analyzer (PE Biosystems, Foster City, CA).

Data analysis. Semi-automated sizing of alleles was carried out using GeneScan 2.1 software (PE Biosystems), and individual genotypes were assigned with the help of the Genotyper version 2.1 software (PE Biosystems). To ensure accuracy of the genotypes, 2 individuals manually checked each genotype. The alleles were then "binned" using a program implemented by us, and consistency in the inheritance of alleles within families was verified using the PedCheck program (16). After correction of any inconsistencies in the data, all allele sizes were downcoded for analysis. We also ensured that this data set did not contain any monozygotic twin pairs by searching the database for siblings with identical genotypes. Three such twin pairs were identified. We also verified the genetic relationships among individuals by applying the program Relative (17) to all of the marker data, to determine if the data set contained any half-sibling pairs or unrelated pairs that were originally ascertained as full-sibling pairs. We identified 16 half-sibling pairs and 2 unrelated pairs in this manner. We used Simwalk to perform marker-to-marker analysis and to check the positions of markers against published maps. By using Simwalk (18) to check genetic distances against published distances, we were able to identify one region of chromosome 18 for which the earlier map that was applied in our first genome scan was incorrect (6). The order and distances among markers have been corrected in the current analysis.

Analysis of the data was performed in 2 stages: 1) The data from the 256 new RA families were analyzed as an independent genome-wide screen (screen 2) for the current study. 2) The data from screen 1 and screen 2 were pooled and analyzed as a combined screen of 512 multicase RA families. Descriptive statistics for demographic and clinical parameters were produced using the SAS package (Chicago, IL). A model-free linkage analysis was carried out using SIBPAL from the SAGE version 3.1 (1997) package (19) for all autosomal marker data. We included all data from affected sibling pairs to implement 1-sided statistical t-tests for evaluating whether the identical by descent (ibd) sharing departs from expectation, assuming Mendelian inheritance of the marker. The pairs were analyzed as though they were independent, but a correction was made to the degrees of freedom to allow for correlation among pairs of siblings from larger families (20).

We used the program ASPEX (available at ftp://lahmed.stanford.edu/pub/aspex/index.html) to perform genetic linkage analysis of the X chromosome and to provide estimates of the average probability that sibling pairs share 0 alleles ibd (the ASPEX software was slightly modified to obtain more significant digits for the probability of sharing 0 alleles ibd). The GeneHunter Plus program was used to perform multipoint analysis of the data for regions that showed significant evidence for linkage using SIBPAL. For the multipoint analysis, we used the Kong and Cox (21) modified version of the nonparametric linkage score to identify regions showing evidence for genetic linkage between marker alleles and the presence of disease in families. The test that we performed evaluates whether there is excess sharing of alleles ibd among pairs of affected relatives. The Kong and Cox approach provides the most weight to relative pairs for which the inheritance of marker alleles can be inferred without error, and provides decreasing weight to marker data according to how poorly one can infer the ibd sharing at a locus among the relatives at a particular locus, given the available marker data from all of the relatives.

For regions that showed significant evidence for linkage using SIBPAL, we also evaluated the impact of covariates on the support for linkage (22). The program LODPAL evaluates the evidence for linkage using a model-free approach that allows the genotypic risk for disease to vary according to pair-specific covariates. As implemented by us (and using the defaults), the program fitted a single regression per covariate, thus assuming that the genotypic effects are approximately additive. Standard output from LODPAL indicates the support for linkage at each marker location by forming a likelihood ratio, in which the numerator models the covariate effects and evidence for linkage, while the denominator only models covariate effects. The distribution of 4.6 times this $\log_{10}$ likelihood ratio is conservatively approximated by a chi-square distribution with 2 degrees of freedom (22). A likelihood ratio statistic to evaluate for significant covariate effects can be obtained by multiplying by 4.6 the difference between the log likelihood ratio for a model including the covariate and one excluding it. In the presence of linkage, this likelihood ratio statistic is asymptotically distributed as a chi-square variate with 1 degree of freedom. The covariates that were examined included DRB1*04, presence of erosions on hand radiographs, rheumatoid factor >100 IU, presence of nodules, and presence of a male affected sibling in the family.

RESULTS

RA families. The structure of the 256 new RA families analyzed for the current genome-wide screen (screen 2), in terms of the number of parents and affected siblings for whom DNA was available and who were therefore genotyped, is shown in Table 1. Unaf-

Table 1. Structure of the genotyping of families in screen 2*

Parents available	Overall (n = 256)	With 2 affected sibs (n = 226)	With 3 affected sibs (n = 27)	With 4 affected sibs (n = 2)	With 6 affected sibs (n = 1)
Both	31	29	2	0	0
One	87	78	9	0	0
None	138	119	16	2	1

* Values are the number of families genotyped. sibs = siblings.

Table 3. Clinical and demographic features of affected individuals in the North American Rheumatoid Arthritis Consortium collection*

Clinical/ demographic features	Screen 1 (n = 546)	Screen 2 (n = 551)	Combined screens 1 and 2 (n = 1,097)
Female, %	77.4	76.0	76.8
Caucasian, %	90.1	92.9	91.2
RF positive, %	82.8	80.7	81.1
HLA–DRB1*04 positive, %	70.4	70.9	70.7
Hand erosions, %	95.7	94.8	95.3
Mean age at disease onset, years	39.1	39.0	39.1
Mean disease duration, years	13.8	14.6	14.3

* RF = rheumatoid factor.

fected siblings were not typed, except in the case of 2 of the families in which parental DNA was not available. At least 1 parent was available for analysis in 46.1% of the families. Table 2 shows the distribution of families after combining the screen 2 data set with that from screen 1.

The clinical and demographic features of the affected siblings who participated in the study are summarized in Table 3. A more detailed analysis of the demographic and clinical features of the patient population is described elsewhere (23). The data from screen 1 differed slightly from the previously published results (6), since the results now included data from a few more affected individuals, and 1 family was dropped for the current analysis (see Patients and Methods). Overall, the demographics and clinical features of the 551 affected siblings in screen 2 were very similar to those in screen 1. In screen 2, the affected siblings were mostly Caucasian (92.9%), had severe RA with high rates of erosions

Table 2. Structure of the genotyping of families in the combined screen*

Parents available	Overall (n = 512)	With 2 affected sibs (n = 453)	With 3 affected sibs (n = 53)	With 4 affected sibs (n = 5)	With 6 affected sibs (n = 1)
Both	66	60†	6	0	0
One	180	162‡	18	0	0
None	266	231§	29¶	5	1

* Values are the number of families genotyped. Unaffected siblings were not typed, except in the case of 2 families for whom parental DNA was not available. sibs = siblings.
† One family had half-siblings.
‡ Nine families had half-siblings.
§ Thirteen families had half-siblings.
¶ Four families had half-siblings.

in the hand (94.8%), had a high frequency of seropositivity for rheumatoid factor (80.7%), and had a mean radiographic severity score of 3.3 on a scale of 0–5. These patients also had a relatively young age at disease onset (39 years) and long disease duration (14.6 years), and 70.9% of them were positive for HLA–DRB1*04. The ratio of female patients to male patients in this data set was 3:1 (76% female), as expected for this disease.

Genome-wide screen. *Screen 2.* The results of a nonparametric analysis (SIBPAL) performed on the genotypic data from the screen 2 data set are shown in Figure 1. Linkage at the HLA complex was confirmed ($P < 0.00005$). Evidence for linkage at the level of $P < 0.05$ was replicated at 1p13 (D1S1631), 6p21.3 (the HLA complex), and 18q21 (D18S858) in this independent data set. There was new evidence for linkage in this new data set at 9p22 (D9S1121; $P = 0.001$) and 10q21 (D10S1221 [$P = 0.0002$] and D10S1225 [$P = 0.0038$]). All markers that showed evidence for linkage at $P < 0.05$ in this second screen are listed in Table 4.

Combined screen. A similar nonparametric analysis of the combined data sets from screen 1 and screen 2, using SIBPAL, showed evidence for linkage at the level of $P < 0.005$ at 1p13 (D1S1631), 1q43 (D1S235), 6q21 (D6S2410), 10q21 (D10S1221), 12q12 (D12S398), 17p13 (D17S1298), and 18q21 (D18S858). Linkage at HLA was again confirmed, and the significance level was increased from $P < 5 \times 10^{-6}$ (in screen 1 or screen 2) to $P < 5 \times 10^{-12}$. A list of all markers that showed evidence for linkage at $P < 0.05$ is shown in
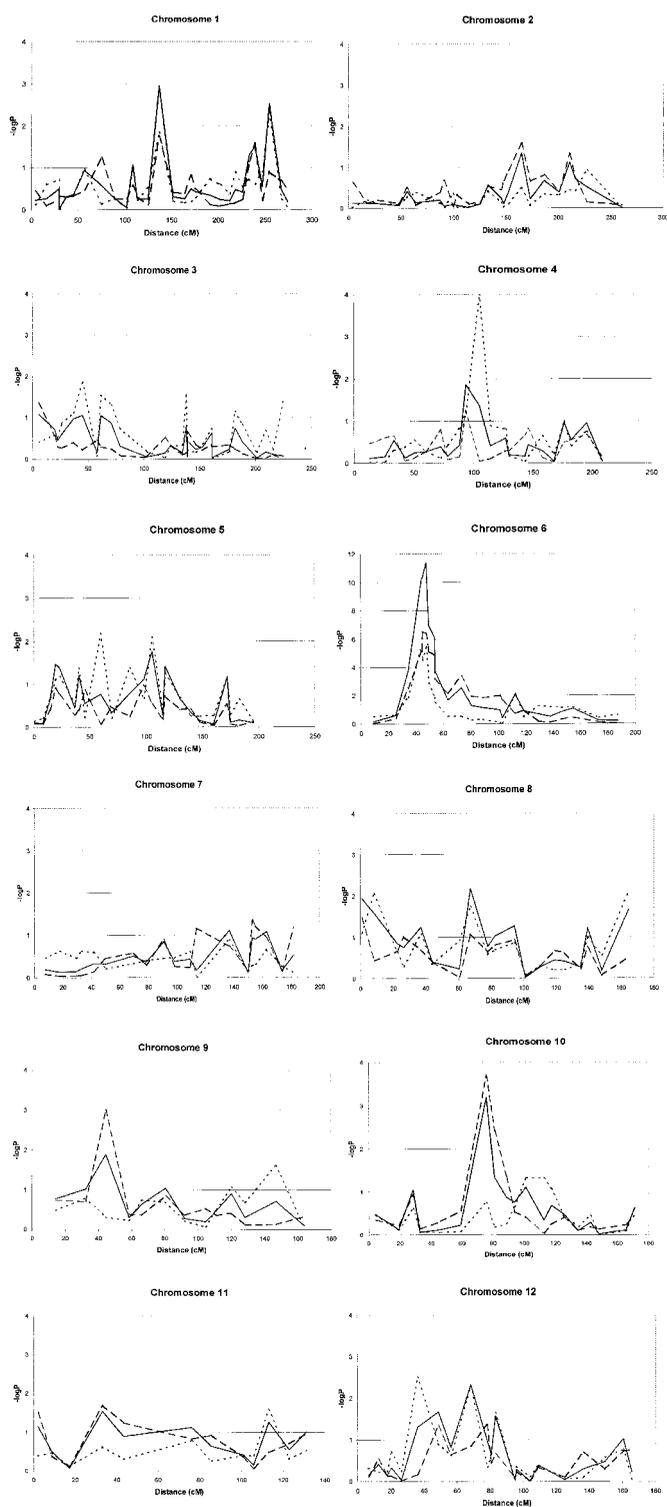
Figure 1. Summary of the sibling pair analyses implemented by SIBPAL, showing evidence for linkage with rheumatoid arthritis. For each chromosome, the negative logarithm of the *P* values (−logP) was plotted against the markers by chromosomal position. For the X chromosome, the analysis was performed using the MLOD method from ASPEX, and the negative logarithm of the *P* values was plotted against the chromosomal position of the markers.
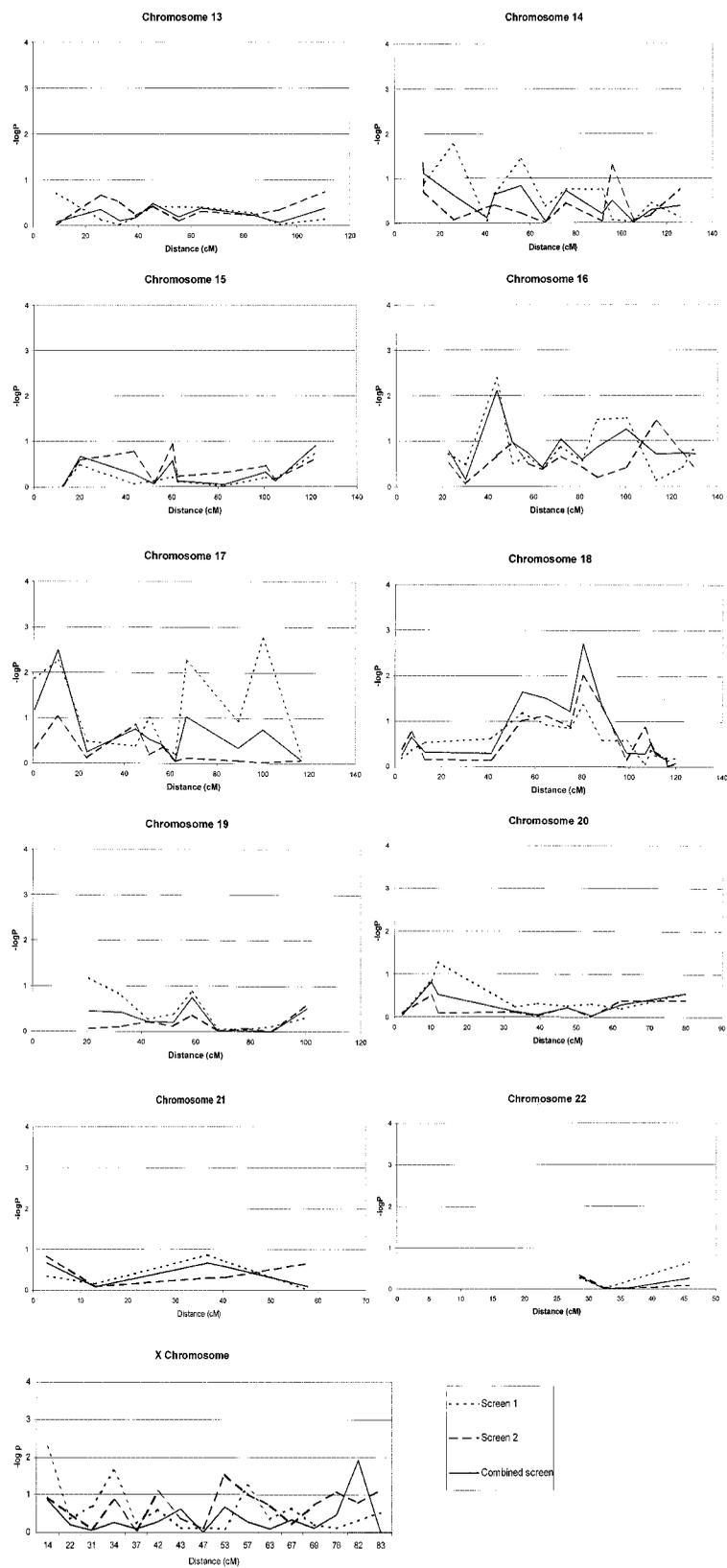
Figure 1 *(Cont'd).*

Table 5, and the results of the SIBPAL analysis are also shown in Figure 1. The results of the SIBPAL analysis performed on the initial data set of 256 families (screen 1) have also been included in Figure 1 for comparison. Analysis of the data from screen 2 and from the combined screen using the GeneHunter program showed evidence for linkage at 1p13, 1q43, 6q21, 16p13.1, and 18q21 (Figure 2).

Table 4. Chromosomal regions with a significance level of P less than 0.05 in screen 2 for the SIBPAL analysis

Locus	Distance (cM)	No. of sibling pairs	Mean sharing	P
Chromosome 1				
D1S1631	136.9	290	0.5348	0.0175
D1S549	239.7	294	0.5315	0.0264
Chromosome 2				
D2S1353	164.5	289	0.5310	0.026
GATA30E06	210.4	268	0.5277	0.0463
Chromosome 3				
D3S2387	5.5	300	0.5278	0.0451
Chromosome 6				
D6S1959	34.2	278	0.5404	0.0029
D6S1641	37	244	0.5629	0.0001
D6S265	44.4	242	0.5766	5.68×10^{-6}
TNFa	44.7	262	0.5772	2.72×10^{-6}
D6S1629	44.9	254	0.5805	2.76×10^{-7}
D6S273	47.7	249	0.5781	3.66×10^{-7}
D6S291	49.5	283	0.5634	6.93×10^{-6}
D6S389	53.8	202	0.5927	1.43×10^{-5}
D6S2427	53.8	293	0.5555	0.0007
D6S1017	63.3	244	0.5365	0.0067
D6S2410	73.1	287	0.5489	0.0004
D6S1053	80.5	282	0.5387	0.0108
D6S1031	88.6	280	0.5356	0.0134
D6S1043	100.9	301	0.5406	0.0098
D6S1056	102.8	263	0.5354	0.0226
Chromosome 7				
TCRB	153	292	0.5301	0.0446
Chromosome 8				
D8S264	0.73	255	0.5316	0.0329
Chromosome 9				
D9S1121	44.3	298	0.5476	0.001
Chromosome 10				
D10S1221	75.6	290	0.5594	0.0002
D10S1225	80.8	293	0.5434	0.0038
GATA121A08	88.4	281	0.5273	0.0443
Chromosome 11				
D11S1984	2.1	296	0.5296	0.0318
ATA34E08	33.0	280	0.5310	0.0205
Chromosome 12				
D12S1042	48.7	254	0.5300	0.0454
D12S1294	78.1	265	0.5313	0.0435
Chromosome 16				
D16S402	113.5	299	0.5295	0.0367
Chromosome 18				
D18S858	80.4	293	0.5366	0.0098
D18S1357	88.6	299	0.5274	0.0472

Table 5. Chromosomal regions with a significance level of P less than 0.05 in the combined screen for the SIBPAL analysis

| Locus	Distance (cM)	P value Screen 1	Screen 2	Combined screen	Lambda
Chromosome 1					
D1S1631	136.9	0.0141	0.0175	0.0011	1.220
D1S2141	233.4	0.1905	0.0701	0.0487	1.096
D1S549	239.7	0.2477	0.0264	0.0318	1.133
D1S235	254.6	0.0048	0.1256	0.003	1.151
Chromosome 2					
D2S1353	164.5	0.3287	0.026	0.0453	1.071
Chromosome 4					
D4S2361	93.5	0.0639	0.0565	0.0141	1.050
D4S1647	104.9	0.0001	0.9198	0.0436	1.056
Chromosome 5					
D5S807	19.0	0.0789	0.1227	0.0336	1.078
D5S817	22.9	0.0663	0.174	0.0421	1.073
D5S1462	105.3	0.0079	0.2887	0.0176	1.175
D5S2501	116.9	0.0559	0.1895	0.0382	1.069
Chromosome 6					
D6S1959	34.2	0.0135	0.0029	1.99×10^{-4}	1.699
D6S265	44.4	2.39×10^{-6}	5.68×10^{-6}	5.33×10^{-11}	1.807
D6S1629	44.9	2.97×10^{-5}	2.76×10^{-7}	5.01×10^{-11}	1.811
D6S273	47.7	3.36×10^{-6}	3.66×10^{-7}	4.19×10^{-12}	1.830
D6S291	49.5	0.0013	6.93×10^{-6}	9.90×10^{-8}	1.822
D6S389	53.8	0.0126	1.43×10^{-5}	7.99×10^{-7}	1.406
D6S2427	53.8	0.031	0.0007	1.46×10^{-4}	1.396
D6S1017	63.3	0.2764	0.0067	0.0197	1.503
D6S2410	73.1	0.2844	0.0004	0.0028	1.347
D6S1021	112.2	0.0075	0.1892	0.0083	1.136
Chromosome 8					
D8S264	0.7	0.0804	0.0329	0.0115	1.225
D8S277	8.3	0.0088	0.3713	0.026	1.242
D8S1110	67.3	0.0187	0.0855	0.0067	1.088
D8S373	164.5	0.0084	0.3234	0.0224	1.101
Chromosome 9					
D9S1121	44.3	0.4972	0.001	0.01303	1.193
Chromosome 10					
D10S1221	75.6	0.1784	0.0002	0.0006	1.176
D10S1225	80.8	0.6662	0.0038	0.0461	1.149
Chromosome 11					
ATA34E08	33.0	0.2418	0.0205	0.0292	1.087
Chromosome 12					
D12S373	36.1	0.0031	0.6866	0.0499	1.129
D12S1042	48.7	0.1247	0.0454	0.0216	1.183
D12S398	68.2	0.0051	0.1429	0.0048	1.110
D12S1052	83.2	0.0227	0.2192	0.0266	1.041
Chromosome 14					
D14S742	12.5	0.1943	0.0524	0.0433	1.024
Chromosome 16					
D16S403	43.9	0.0042	0.2083	0.0076	1.192
Chromosome 17					
D17S1298	10.7	0.0053	0.0933	0.0031	1.000
Chromosome 18					
D18S877	54.4	0.0635	0.0962	0.0228	1.128
D18S535	64.5	0.1172	0.0748	0.0309	1.180
D18S858	80.4	0.0433	0.0098	0.002	1.233
D18S1357	88.6	0.2631	0.0472	0.0494	1.120

Analysis of covariates. The results of the covariate analysis are summarized in Figure 3. On chromosome 6p, inclusion of DRB1*04 genotypes increased the
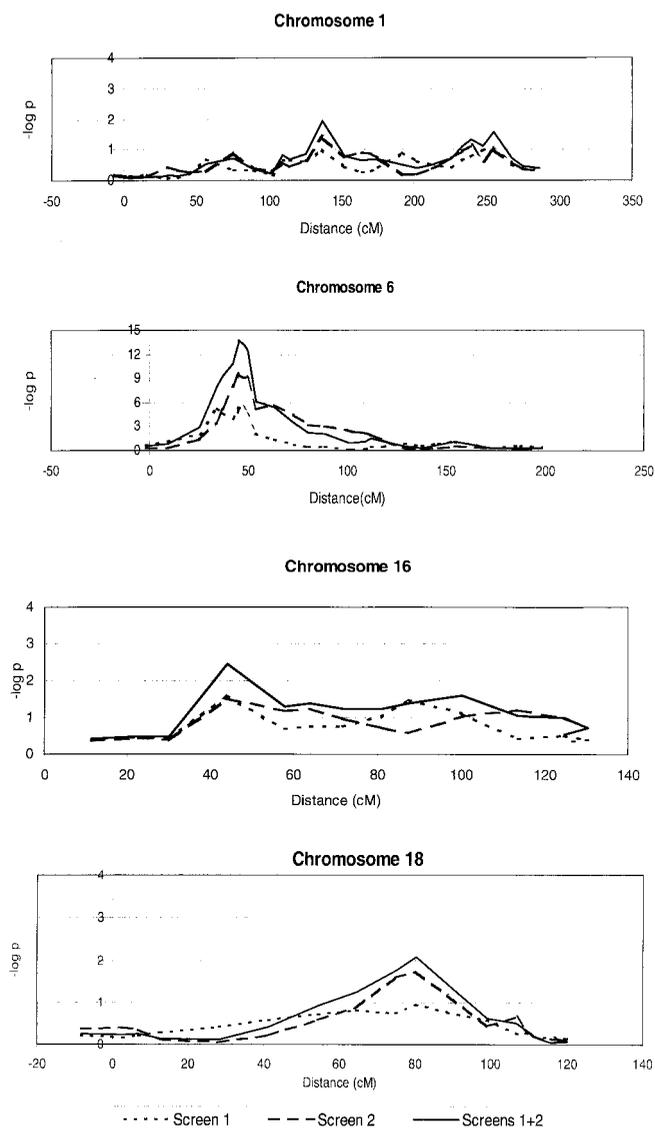
Figure 2. Analysis of the data using the GeneHunter program, showing evidence for linkage at 1p13, 1q43, 6q21, 16p13.1, and 18q21. −log p = negative logarithm of the *P* values.

logarithm of odds (LOD) score, with the most significant finding observed at D6S1629, which increased the LOD score from 10.99 to 14.28 (*P* < 0.00009). Significant increases that were evidence for linkages on chromosome 1 were observed when covariates were modeled jointly. Inclusion of DRB1*04 significantly increased the probability of linkage (*P* < 0.003) at D1S1589 (192 cM), yielding an LOD score of 1.91 compared with 0 without the covariate, and inclusion of rheumatoid factor >100 IU as a covariate produced significant increases indica-

tive of linkage (*P* < 0.0005) in the region around D1S551 (114 cM), with the LOD score going from 0.13 to 2.71. However, no significant increase indicating evidence for linkage was observed for chromosome 18 or for any of the other chromosomes. The "nodules" and "erosive" covariates did not significantly contribute to a difference in the level of linkage (data not shown).
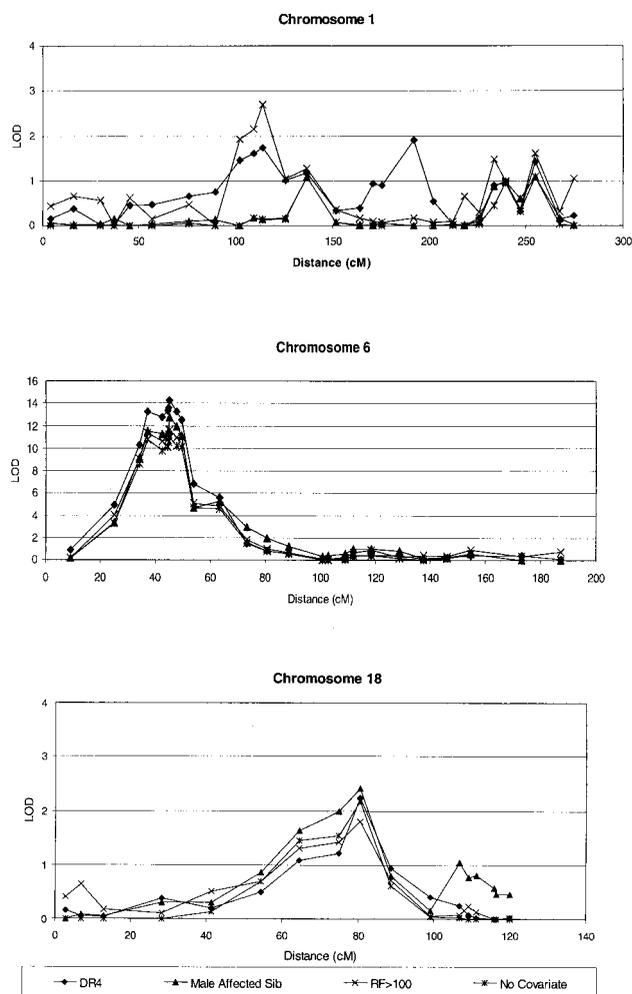
Figure 3. Covariate analyses of the contribution of DRB1*04, male affected siblings (sib), and rheumatoid factor (RF) seropositivity (>100 IU) for selected chromosomal regions, showing evidence for linkage in sibling pair tests. Evidence of a significantly increased likelihood of linkage on chromosome 1 was observed when jointly modeling covariates, including DRB1*04 and RF seropositivity. On chromosome 6p, inclusion of DRB1*04 genotypes increased the logarithm of odds (LOD) score. However, no significant increases that would indicate evidence for linkage were observed for chromosome 18 or for all other chromosomes.

DISCUSSION

In this report, we have presented the results of our second genome-wide screen in 256 new multicase RA families from the NARAC collection. In addition, we have performed a combined analysis of the data from our 2 independent genome-wide screens, in a total of 677 RA sibling pairs (corresponding to 581 independent pairs, using the Hodge correction [24]). Analysis of the demographic and clinical features of the affected siblings from the screen 1 and screen 2 data sets showed that the 2 patient populations were very similar to each other in terms of disease features. This was reassuring, since homogeneity between the 2 populations is clearly desirable if they are going to be used to compare the results of 2 independent genome-wide screens.

The nonparametric analysis of the data from screen 2 revealed several regions that showed evidence for linkage at the level of $P < 0.05$. Of these, 3 regions were replicated from our initial screen, namely, 1p13 (D1S1631), 6p21.3 (the HLA complex), and 18q21 (D18S858). Although a number of other regions had shown some modest evidence for linkage ($P < 0.005$) in screen 1, these were not replicated in screen 2. Similarly, some new loci showed evidence of linkage at $P < 0.05$ in the current data set (screen 2), and not in screen 1. These results may be explained by the limitations of the 2 individual screens in terms of power. Both data sets consisted of similar numbers of affected sibling pairs, and it is likely that these sample sizes do not provide adequate power to unambiguously replicate linkage at loci that have a modest effect in RA. In fact, several thousand affected sibling pairs may be required to demonstrate "definite" linkage for genes of modest effect (11). In the case of RA, it is becoming increasingly clear that pooling results across different genome screens (10) will be required to increase the power to detect true linkage for loci with modest effects. The sample size of 581 independent sibling pairs is adequate to detect locus-specific increased recurrence risks (25) of 1.21, 1.31, and 1.40 at the 0.05, 0.005, and 0.0005 significance levels, respectively.

As a start, we combined the data sets from our 2 independent genome-wide screens, and repeated the analyses that we carried out on each screen. The increased power of this combined data set to detect linkage is reflected in the improvement in the significance level of linkage at the HLA markers, which improved by $\sim10^{-6}$-fold. Thus, it may be extrapolated from this observation that this larger data set has improved power to detect other loci having more modest effects than HLA. Among the non-HLA loci, evidence for linkage at $P < 0.005$ was observed at 1p13 (D1S1631), 1q43 (D1S235), 6q21 (D6S2410), 10q21 (D10S1221), 12q12 (D12S398), 17p13 (D17S1298), and 18q21 (D18S858). Analysis of the data using the Gene-Hunter software showed evidence of linkage (LOD score >2) at 1p13, 1q43, 6q21, and 18q21 (Figure 2). The GeneHunter results also demonstrated linkage at 16p13.1 (LOD score >2), although this locus only reached a $P = 0.007$ level of significance in the SIBPAL analysis.

The analysis of covariates, including DRB1*04, presence of erosions on hand radiographs, rheumatoid factor >100 IU, presence of nodules, presence of a male affected sibling in the family, and ethnicity, did not show a significant increase in linkage, except on chromosomes 1 and 6. On chromosome 6p, the increase in LOD score observed with the DRB1*04 covariate showed that DRB1*04 or a closely linked locus in disequilibrium with DRB1*04 modifies the genetic risk for RA. Further stratified analysis showed that the presence of DRB1*04 in both siblings increased the evidence for linkage; however, among sibling pairs who were both negative for DRB1*04, there remained evidence for linkage ($P = 0.0002$). These data are consistent with our recent observations that the DRB1 locus does not explain all of the HLA associations with RA (26). The results for chromosome 1 showed that there may be multiple genetic factors on chromosome 1, and that the genetic effects may be most pronounced in subsets of RA patients. Because there was no evidence for linkage at D1S1589 when covariate effects were not modeled, the distribution of the test for a significant covariate effect may not be very well approximated by chi-square distribution (22).

Further studies of additional data will help to evaluate whether there is linkage heterogeneity according to DRB1*04 status for D1S1589. Stratifying the data set in terms of presence of rheumatoid factor >100 IU or presence of DRB1*04 may be particularly useful for further fine mapping at the 2 loci that showed an increase in LOD scores when these covariates were included in the analysis.

To date, 3 other genome-wide screens have been performed in RA multicase families (5,7,9). Our combined screen in the present study represents the largest data set among all of these genome screens in RA. The strongest evidence for linkage was observed at the HLA locus in all screens, except one (9). In addition, some of the non-HLA loci that provided modest evidence of linkage ($P < 0.005$) in our combined screen overlap with

regions that have been identified in other RA screens. The 1q43 locus has been linked to RA in both the European (5) and the British (7) data sets. The 6q21 region, however, showed evidence of linkage to RA in the British data set only, whereas linkage at the 18q21 locus was observed in the European study. Studies of candidate regions have implicated the 17q22 locus (27), which is syntenic to the Cia5 locus in the rat, and the 8q13 locus in the vicinity of the CRH gene in RA susceptibility. In our first genome-wide screen (screen 1), the 17q22 locus showed encouraging evidence for linkage ($P < 0.0017$). However, this result was not replicated in our second independent screen, nor was it present in our combined data set of 512 families. The 8q13 locus, on the other hand, was close to the $P < 0.005$ level of significance for linkage in our combined screen (D8S1110; $P = 0.0067$). Further analyses will be required to ascertain whether this modest level of linkage with RA is due to CRH or to some other gene in the region.

The non-HLA regions showing linkage with RA have also been implicated in a number of other autoimmune diseases. The 1p13 region has been linked to SLE (28). The 1q43 region containing the D1S235 marker was first linked to susceptibility to SLE using a candidate approach (29), and this linkage was subsequently confirmed in 3 separate genome-wide screens (28–30). The 18q21 locus has shown evidence for linkage with type I insulin-dependent diabetes mellitus (31–33), SLE (34), and Grave's disease (35) in humans, and the orthologous region in rodents has been implicated in animal models of SLE (36), multiple sclerosis (37), and experimental allergic encephalomyelitis (38).

The linkage data described herein provide only a rough guide as to the location of RA susceptibility genes. A further analysis of linkage with a dense set of markers may allow us to narrow down regions of interest to within 5–10 cM. This is still a large region of the genome, and it is premature to aggressively evaluate a large number of candidate genes in these regions. Given that the etiology of RA is still unknown, the spectrum of reasonable candidate genes can be very broad, even within a particular linked region. Nevertheless, it is interesting to examine these regions for particularly provocative candidate genes. For example, one possible candidate gene on chromosome 18q21 is the TNFRSF11A gene, which encodes receptor activator of nuclear factor κB (RANK). RANK has been shown to be critically involved in the differentiation of osteoclasts in inflamed synovium and is clearly important for the development of bone resorption in inflammatory arthri-

tis (39). Because the NARAC collection is specifically selected for erosive RA, our patient population is well suited to detect genes involved in the pathogenesis of bony erosions. Therefore, we intend to pursue family-based association studies of RANK in our family collection. Association studies on multiplex family data sets can offer more statistical power than that of standard case–control analyses, particularly when association is specifically examined in families with evidence of linkage to the region of interest (40). Nevertheless, a candidate-gene strategy is risky, and it is not an efficient way to thoroughly evaluate the candidate regions of linkage that we have identified.

For RA and other complex autoimmune disorders, the challenge is to move to gene identification in these regions of modest linkage. Further expansion of multiplex family collections will enhance confidence in these linkage results. However, it is likely that only dense haplotype-association mapping will provide sufficiently precise localization to justify a comprehensive analysis of all candidate genes in a region. A combined approach based on positional cloning and candidate-gene analysis has led to gene identification in Crohn's disease (41,42). We are hopeful that an analysis of candidate-linkage regions in RA will provide similar success in the future.

REFERENCES

1. Lin JP, Cash JM, Doyle SZ, Peden S, Kanik K, Amos CI, et al. Familial clustering of rheumatoid arthritis and other autoimmune diseases. Hum Genet 1998;103:475–82.
2. Harris ED Jr. Rheumatoid arthritis: pathophysiology and implications for therapy. N Engl J Med 1990;322:1277–89.
3. Stastny P. Association of the B-cell alloantigen DRw4 with rheumatoid arthritis. N Engl J Med 1978;298:869–71.
4. Ollier W, Thomson W. Population genetics of rheumatoid arthritis. Rheum Dis Clin North Am 1992;18:741–59.
5. Cornelis F, Faure S, Martinez M, Prud'homme JF, Fritz P, Dib C, et al. New susceptibility locus for rheumatoid arthritis suggested by

a genome-wide linkage study. Proc Natl Acad Sci U S A 1998;95: 10746–50.

6. Jawaheer D, Seldin MF, Amos CI, Chen WV, Shigeta R, Monteiro J, et al. A genomewide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases. Am J Hum Genet 2001;68:927–36.

7. MacKay K, Eyre S, Myerscough A, Milicic A, Barton A, Laval S, et al. Whole-genome linkage analysis of rheumatoid arthritis susceptibility loci in 252 affected sibling pairs in the United Kingdom. Arthritis Rheum 2002;46:632–9.

8. Seldin MF, Amos CI, Ward R, Gregersen PK. The genetics revolution and the assault on rheumatoid arthritis. Arthritis Rheum 1999;42:1071–9.

9. Shiozawa S, Hayashi S, Tsukamoto Y, Goko H, Kawasaki H, Wada T, et al. Identification of the gene loci that predispose to rheumatoid arthritis. Int Immunol 1998;10:1891–5.

10. Jawaheer D, Gregersen PK. The search for rheumatoid arthritis susceptibility genes: a call for global collaboration. Arthritis Rheum 2002;46:582–4.

11. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995;11:241–7.

12. Arnett FC, Edworth SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315–24.

13. Spiegel TM, Spiegel JS, Paulus HE. The joint alignment and motion scale: a simple measure of joint deformity in patients with rheumatoid arthritis. J Rheumatol 1987;14:887–92.

14. Anderson SG, Bentzon MW, Houba V, Krag P. International reference preparation of rheumatoid arthritis serum. Bull World Health Organ 1970;42:311–8.

15. Erlich HA, Bugawan T, Gegovich AB, Scharf S, Griffith R, Saiki R, et al. HLA-DR, DQ and DP typing using PCR amplification and immobilized probes. Eur J Immunogenet 1991;18:33–55.

16. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 1998;63:259–66.

17. Goring HH, Ott J. Relationship estimation in affected sib pair analysis of late-onset diseases. Eur J Hum Genet 1997;5:69–77.

18. Sobel E, Lange K. Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 1996;58:1323–37.

19. Statistical analysis for genetic epidemiology, release 3.1. Computer program package available from the Department of Epidemiology and Biostatistics, Rammelkamp Center for Education and Research, MetroHealth Campus, Case Western Reserve University, Cleveland; 1997.

20. Wilson AF, Elston RC. Statistical validity of the Haseman-Elston sib-pair test in small samples. Genet Epidemiol 1993;10:593–8.

21. Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 1997;61:1179–88.

22. Olson JM. A general conditional-logistic model for affected-relative-pair linkage studies. Am J Hum Genet 1999;65:1760–9.

23. Jawaheer D, Lum R, Amos CI, Gregersen PK, Criswell LA. Familial clustering of disease features in rheumatoid arthritis: an analysis of 512 multicase families [abstract]. Arthritis Rheum 2002;46 Suppl 9:572.

24. Hodge SE. The information contained in multiple sibling pairs. Genet Epidemiol 1984;1:109–22.

25. Risch N. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet 1990;46:229–41.

26. Jawaheer D, Li W, Graham RR, Chen WC, Damle A, Xiao X, et

al. Dissecting the genetic complexity of the HLA association with rheumatoid arthritis. Am J Hum Genet 2002;71:585–94.

27. Barton A, Eyre S, Myerscough A, Brintnell B, Ward D, Ollier WE, et al. High resolution linkage and association mapping identifies a novel rheumatoid arthritis susceptibility locus homologous to one linked to two rat models of inflammatory arthritis. Hum Mol Genet 2001;10:1901–6.

28. Gaffney PM, Kearns GM, Shark KB, Ortmann WA, Selby SA, Malmgren ML, et al. A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families. Proc Natl Acad Sci U S A 1998;95:14875–9.

29. Tsao BP, Cantor RM, Kalunian KC, Chen CJ, Badsha H, Singh R, et al. Evidence for linkage of a candidate chromosome 1 region to human systemic lupus erythematosus. J Clin Invest 1997;99: 725–31.

30. Moser KL, Neas BR, Salmon JE, Yu H, Gray-McGuire C, Asundi N, et al. Genome scan of human systemic lupus erythematosus: evidence for linkage on chromosome 1q in African-American pedigrees. Proc Natl Acad Sci U S A 1998;95:14869–74.

31. Davies JL, Kawaguchi Y, Bennett ST, Copemann JB, Cordell JH, Pritchard LE, et al. A genome-wide search for human type 1 diabetes susceptibility genes [see comments]. Nature 1994;371: 130–6.

32. Merriman T, Twells R, Merriman M, Eaves I, Cox R, Cucca F, et al. Evidence by allelic association-dependent methods for a type 1 diabetes polygene (IDDM6) on chromosome 18q21. Hum Mol Genet 1997;6:1003–10.

33. Merriman TR, Cordell HJ, Eaves IA, Danoy PA, Coraddu F, Barber R, et al. Suggestive evidence for association of human chromosome 18q12-q21 and its orthologue on rat and mouse chromosome 18 with several autoimmune diseases. Diabetes 2001; 50:184–94.

34. Shai R, Quismorio FP Jr, Li L, Kwon OJ, Morrison J, Wallace DJ, et al. Genome-wide screen for systemic lupus erythematosus susceptibility genes in multiplex families. Hum Mol Genet 1999; 8:639–44.

35. Vaidya B, Imrie H, Perros P, Young ET, Kelly WF, Carr D, et al. Evidence for a new Graves disease susceptibility locus at chromosome 18q21. Am J Hum Genet 2000;66:1710–4.

36. Kono DH, Burlingame RW, Owens DG, Kuramochi A, Balderas RS, Balomenos D, et al. Lupus susceptibility loci in New Zealand mice. Proc Natl Acad Sci U S A 1994;91:10168–72.

37. Bureau JF, Montagutelli X, Bihl F, Lefebvre S, Guenet JL, Brahic M. Mapping loci influencing the persistence of Theiler's virus in the murine central nervous system. Nat Genet 1993;5:87–91.

38. Dahlman I, Wallstrom E, Weissert R, Stroch M, Kornek B, Jacobsson L, et al. Linkage analysis of myelin oligodendrocyte glycoprotein-induced experimental autoimmune encephalomyelitis in the rat identifies a locus controlling demyelination on chromosome 18. Hum Mol Genet 1999;8:2183–90.

39. Li J, Sarosi I, Yan XQ, Morony S, Capparelli C, Tan HL, et al. TANK is the intrinsic hematopoietic cell surface receptor that controls osteoclastogenesis and regulation of bone mass and calcium metabolism. Proc Natl Acad Sci U S A 2000;97:1566–71.

40. Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, et al. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat Genet 2000;26:163–75.

41. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001;411:599–603.

42. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 2001;411:603–6.