

Association Between Single-Nucleotide Polymorphisms in the *SEC8L1* Gene, Which Encodes a Subunit of the Exocyst Complex, and Rheumatoid Arthritis in a Japanese Population

Daisuke Hamada,¹ Yoichiro Takata,² Dai Osabe,³ Kyoko Nomura,³ Syuichi Shinohara,³ Hiroshi Egawa,² Syunji Nakano,² Fumio Shinomiya,⁴ Charles R. Scafe,⁵ Vincent M. Reeve,⁵ Tatsuro Miyamoto,² Maki Moritani,² Kiyoshi Kunika,² Hiroshi Inoue,² Natsuo Yasui,¹ and Mitsuo Itakura²

Objective. To identify rheumatoid arthritis (RA) susceptibility genes in a Japanese population by conducting a large-scale case–control association analysis and linkage disequilibrium (LD) mapping on chromosome 7q31–34, a candidate susceptibility locus identified in a preliminary genome-wide scan in 53 Japanese families, using single-nucleotide polymorphisms (SNPs).

Methods. We prepared 728 dense, evenly spaced SNPs with a minor allele frequency >0.15 in each gene locus on chromosome 7q31–34. Using these SNPs, a 2-stage case–control analysis was performed on 760 RA patients (157 men and 603 women) and 806 non-RA

controls (189 men and 617 women). Haplotypes and LD mapping results were assessed based on SNP genotypes in 380 controls.

Results. Forty-eight SNPs showed allele associations ($P < 0.05$) in the first set of DNA samples (380 RA cases and 380 non-RA controls; first-stage analysis). For 4 of the SNPs in the *SEC8L1* gene, the association was replicated ($P < 0.05$) in the second, independent set of DNA samples (an additional 380 RA cases and 380 non-RA controls; second-stage analysis). When data from the 2 groups were combined, the most significant allele association was observed with SNP 441, an intronic SNP of the *SEC8L1* gene ($P = 0.000059$). The *SEC8L1* SNPs with significant allele associations were all located in a single conserved LD block (block 4). Haplotype analysis revealed the disease-risk ($P = 0.0015$) and disease-protective ($P = 0.0000062$) haplotypes. Resequencing of coding exons within block 4 did not identify any nonsynonymous SNPs. Real-time quantitative polymerase chain reaction revealed that *SEC8L1* was expressed ubiquitously in human tissues, including fibroblast-like synoviocytes from RA patients.

Conclusion. Our locus-wide association and LD analyses identified intronic SNPs and haplotypes in the *SEC8L1* gene that are strongly associated with RA. We propose that *SEC8L1*, which encodes a component of the exocyst complex, is a candidate susceptibility gene for RA in the Japanese population.

Rheumatoid arthritis (RA) is one of the most common autoimmune diseases. It is characterized by chronic inflammation of synovium and subsequent joint

Supported by grants from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (Cooperative Link of Unique Science and Technology for Economy Revitalization [CLUSTER], Special Coordination Funds for Promoting Science and Technology) and the Japan Society for the Promotion of Science (Research for the Future program and grant-in-aid for scientific research). Dr. Hamada's work was supported by the 21st Century Center of Excellence program (Disease Proteomics for Multifactorial Disorders).

¹Daisuke Hamada, MD, Natsuo Yasui, MD, PhD: Institute of Health Biosciences, The University of Tokushima Graduate School, Tokushima, Japan; ²Yoichiro Takata, MD, Hiroshi Egawa, MD, PhD, Syunji Nakano, MD, PhD, Tatsuro Miyamoto, MD, Maki Moritani, PhD, Kiyoshi Kunika, PhD, Hiroshi Inoue, MD, PhD, Mitsuo Itakura, MD, PhD: The University of Tokushima, Tokushima, Japan; ³Dai Osabe, MSc, Kyoko Nomura, MEng, Syuichi Shinohara, BSc: Fujitsu Limited, Tokyo, Japan; ⁴Fumio Shinomiya, MD, PhD: Tokushima Kensei Hospital, Tokushima, Japan; ⁵Charles R. Scafe, BSc, Vincent M. Reeve, BSc: Applied Biosystems, Foster City, California.

Address correspondence and reprint requests to Mitsuo Itakura, MD, PhD, Division of Genetic Information, Institute for Genome Research, Tokushima University, 3-18-15 Kuramoto-Cho, Tokushima-City, Tokushima 770-8503, Japan. E-mail: itakura@genome.tokushima-u.ac.jp.

Submitted for publication August 31, 2004; accepted in revised form January 25, 2005.

destruction. The prevalence is almost constant in many populations throughout the world, ranging from 0.3% to 1% (1). The precise etiology of RA is largely unknown; however, evidence from studies of twins and families suggests that both genetic and environmental factors (e.g., infectious, hormonal, and pregnancy factors) contribute to susceptibility to RA (2,3). An association between the risk of RA and the presence of the HLA locus has been established. In particular, several HLA-DRB1 alleles encoding the "shared epitope" are recognized as disease-risk or disease-severity alleles (4). It is estimated, however, that this association accounts for only one-third of the genetic component (5), and several other, non-HLA genes are believed to contribute to RA susceptibility. Many potential candidate genes, such as tumor necrosis factor α (TNF α), CTLA-4, and cytokine-related genes, have been extensively investigated, but the results are a subject of controversy (6). Genome-wide linkage analyses of RA in affected sibpairs have been performed by at least 4 independent groups (7–12), and although several non-HLA loci showed suggestive evidence of linkage in multiple studies, no common loci aside from the HLA region were identified.

It has been suggested that in common diseases such as RA, linkage analysis alone might not provide the necessary resolution to identify the underlying genes and that fine mapping of the linked regions should be attempted using linkage disequilibrium (LD) analysis with single-nucleotide polymorphisms (SNPs). In fact, such strategies are beginning to reveal novel and complex disease-related genes for type 2 diabetes and for Crohn's disease (13–15). Similar strategies in RA have found a functional haplotype of the peptidyl arginine deiminase 4 (PADI-4) gene to be associated with RA in the Japanese population (16). *PADI4* is located on chromosome 1p36, which 2 previous studies had identified as a candidate locus for RA (7,8). The same group of investigators (17) recently succeeded in identifying an intronic SNP in a runt-related transcription factor 1 (RUNX-1) binding site of the solute carrier family 22 member 4 (SLC22A4) gene that is associated with RA (17).

LD mapping with SNPs in candidate loci identified through linkage analysis is useful for pinpointing disease susceptibility genes in common multifactorial disorders. In this study, we performed a large-scale case-control association analysis and LD mapping on chromosome 7q31–34, a candidate locus identified through a preliminary genome scan in a Japanese population (18). Our results indicate that the *SEC8L1* gene,

which encodes a subunit of the exocyst complex, is a candidate susceptibility gene in RA.

PATIENTS AND METHODS

Subjects evaluated in the association study. A total of 1,566 Japanese subjects, consisting of 760 RA patients (157 men and 603 women) and 806 healthy subjects (non-RA controls; 189 men and 617 women), were evaluated in the association study. RA patients were recruited through the orthopedic clinic at Tokushima University Hospital and the rheumatology clinic at Tokushima Kensei Hospital. Healthy control subjects were recruited from the general population. All RA patients met the revised criteria of the American College of Rheumatology (19), with a minimum disease duration of 3 years. Clinical information, including the age at disease onset, initially affected joints, medication history, side effects of disease-modifying antirheumatic drugs, and family medical history, was obtained on all study subjects. RA disease status was classified into 3 subsets (mutilating disease, more erosive subset, and less erosive subset) according to the criteria described by Ochi et al (20).

Blood specimens were collected from all study subjects. Genomic DNA was prepared from peripheral blood leukocytes or from Epstein-Barr virus-immortalized lymphoblast cell lines using a standard protocol. Synovial specimens were obtained from the knee joints of 3 RA patients who had undergone arthroplastic surgery. Fibroblast-like synoviocytes (FLS) were derived from the synovial specimens and cultured in Dulbecco's modified Eagle's medium containing 10% fetal calf serum as described previously (21).

The study protocol was approved by the Institutional Review Board of the University of Tokushima. Written informed consent was obtained from all participants prior to blood sampling.

SNP selection and genotyping. To establish a useful set of markers for the association study, we first searched SNPs that were located in an interval between *D7S486* and *D7S2513*, an ~ 1 -logarithm of odds (LOD)-drop interval from *D7S2560*, which showed nominal evidence of linkage in the public databases (Database of Single-Nucleotide Polymorphisms [dbSNP] and The SNP Consortium) and the Celera Discovery System database. Among the numerous candidate SNPs identified, we selected for analysis SNPs that met the following 3 criteria: 1) dense and evenly-spaced (1 SNP every 10–20 kb) gene-centric SNPs in each gene locus spanning 10 kb upstream of the transcription start site of the gene and 10 kb downstream of the stop codon, 2) SNPs suitable for use in developing the TaqMan allele discrimination assay, and 3) common SNPs that showed minor allele frequencies >0.15 in 46 healthy unrelated Japanese control subjects initially tested.

TaqMan minor groove binder probe and primer sets were designed by Applied Biosystems using Primer Express Software (version 2.0; Applied Biosystems, Foster City, CA). We genotyped SNPs using either TaqMan Universal Master Mix (Applied Biosystems) or a QuantiTect Probe polymerase chain reaction (PCR) kit (Qiagen, Stanford, CA) and 5 ng of DNA, 450 nM of each primer, and 100 nM of probe in a 4- μ l reaction volume. PCR conditions were as follows: 95°C for 10 minutes for enzyme activation, followed by 40 cycles of 92°C or

94°C for 15 seconds and 60°C for 1 minute. Thermal cycling was conducted on an ABI GeneAmp PCR System 9700 (Applied Biosystems). Each 384-well plate contained 380 samples of unknown genotype and 4 samples of no-template control. After reaction, the fluorescence of VIC and FAM dyes was detected using an ABI Prism 7900HT Sequence Detector System with SDS software (version 2.1; both from Applied Biosystems).

Case-control association study. To reduce the time and cost of genotyping, we used a 2-stage case-control analysis strategy in 1,520 subjects by assigning them to 2 independent subsets consisting of 380 RA cases and 380 non-RA controls each. All SNPs were genotyped in the first set of DNA samples (first-stage analysis), and SNPs exhibiting significant allele associations with RA (cutoff at $P < 0.05$) were further tested for replication in the independent second set of DNA samples (second-stage analysis). For each stage of analysis, allele frequencies of SNPs were compared in the RA cases and non-RA controls with the use of chi-square tests with 2×2 contingency tables. For SNPs exhibiting significant allele associations, we also tested genotype frequencies for association with disease status using dominant and recessive models. Hardy-Weinberg equilibrium of alleles at individual loci was assessed by chi-square statistics (22). SNPs with a minor allele frequency of <0.15 , genotype success rate of <0.97 , and Hardy-Weinberg equilibrium test result of $P < 0.05$ in control subjects were excluded from analysis.

The distribution of LD in the candidate region and the phase estimation were evaluated using genotypes from 380 control subjects in the first-stage analysis. Pairwise LD was estimated using the formula

$$D = x_{11} - p_1q_1$$

where x_{11} is the frequency of haplotype A_1B_1 , and p_1 and q_1 are the frequencies of alleles A_1 and B_1 at loci A and B, respectively. A standardized LD coefficient, r , was derived using the formula

$$r = D/(p_1p_2q_1q_2)^{1/2}$$

(23), where p_2 and q_2 are the frequencies of the other alleles at loci A and B, respectively. Lewontin's coefficient, D' , was derived using the formula

$$D' = D/D_{\max}$$

where $D_{\max} = \text{minimum}(p_1q_2, p_2q_1)$ when D is <0 , or where $D_{\max} = \text{minimum}(p_1q_1, p_2q_2)$ when D is >0 (24). Haplotype frequencies for multiple loci were estimated using the expectation-maximization method. The statistics including permutation tests were performed with the use of either the Arlequin program-based SNPalyze Pro version 3.2 (Dyna-com, Tokyo, Japan) or our original tool, Fujitsu Disease Susceptibility Gene Discovery System, developed for use in association statistics (Fujitsu, Tokyo, Japan). A graphic overview of the LD was constructed using the Graphical Overview of Linkage Disequilibrium software (25) (online at <http://www.sph.umich.edu/csg/abecasis/GOLD/>).

We calculated the population-attributable risk as follows:

$$\text{population-attributable risk} = (X - 1)/X$$

where $X = (1 - f)^2 + 2f(1 - f)r_1 + f^2r_2$, f is the frequency of the risk allele, and r_1 and r_2 are the estimated genotype risk ratios, and using an RA prevalence of 0.01 (26).

Resequencing of the coding exons in the LD block. PCR primer sets for exons 11–14 of the *SEC8L1* gene were designed as follows: 5'-ATACTCATGTCCTCATGTGCG-3' (exon 11 forward), 5'-TGACACTTGCCAAGTCAATG-3' (exon 11 reverse), 5'-TTCTTGCAAGAGAGCTCAGG-3' (exon 12 forward), 5'-AACCACCATCAGGCTCAAAC-3' (exon 12 reverse), 5'-TCCTGCAGCAAAGGTAGATG-3' (exon 13 forward), 5'-AAACAGATGTGCTTCTCCC-3' (exon 13 reverse), 5'-CTCATGTGTCCTTGTGC-3' (exon 14 forward), and 5'-GACCAGTGTACTCTCATC-3' (exon 14 reverse). DNA samples from 24 subjects (12 RA patients and 12 non-RA controls) were chosen to represent the different haplotypes. PCR was performed with a KOD-Plus DNA polymerase kit (Toyobo, Osaka, Japan) according to the manufacturer's protocol. After PCR, products were treated with ExoSAP-IT (Amersham Biosciences, Piscataway, NJ) at 37°C for 15 minutes, followed by incubation at 80°C for 15 minutes to deactivate enzyme. The purified PCR products were directly sequenced with the primers used for PCR amplification, using BigDye Terminator chemistry on an ABI 3100 Genetic Analyzer (Applied Biosystems).

Real-time quantitative PCR. Total cellular RNA from FLS was extracted with the use of an RNeasy kit (Qiagen). Additional RNA from various tissues was taken from the Human Total RNA Master Panel II kit (BD Biosciences, San Jose, CA). RNA was reverse-transcribed with an oligo(dT) primer and SuperScript II (Invitrogen, Carlsbad, CA). Expression of the *SEC8L1* gene and the control β -actin gene was examined by real-time quantitative reverse transcription-PCR (RT-PCR) using TaqMan Assays-on-Demand gene expression products (Hs00253986_m1 for *SEC8L1* and Hs99999903_m1 for β -actin). Expression of the *SEC8L1* gene in various tissues relative to the bone marrow was calculated by the standard curve method as described by the manufacturer (Applied Biosystems).

RESULTS

Clinical phenotypes of the study subjects. The clinical features of the RA cases and non-RA controls are summarized in Table 1. In the first-stage and second-stage analyses, 78.9% and 79.7%, respectively, of the RA patients were women. The mean \pm SD age at the time of sampling was 61.9 ± 12.2 years in the RA patients and 38.8 ± 15.4 years in the non-RA control subjects. The non-RA control group was adjusted for sex to fit the sex distribution in the RA case group.

Selection of SNPs in the 7q31–34 region and development of the TaqMan assay. Our preliminary genome-wide affected sibpair linkage analysis using 67 sibpairs from 53 Japanese RA families identified nominal evidence of linkage (maximum LOD score >0.74 , corresponding to $P < 0.05$) (27) to the 7q31–34 region, with a maximum LOD score of 1.41 for *D7S2560* (un-

Table 1. Clinical features of the rheumatoid arthritis (RA) cases and the non-RA controls

	RA cases (n = 760)	Non-RA controls (n = 806)
No. of subjects initially screened	–	46
First-stage analysis		
No. of subjects	380	380
% female	78.9	78.9
Second-stage analysis		
No. of subjects	380	380
% female	79.7	77.6
Race, % Japanese	100	100
Age at sampling, mean \pm SD years	61.9 \pm 12.2	38.8 \pm 15.4
Age at RA onset		
<30 years	62	–
30–39 years	115	–
40–49 years	238	–
50–59 years	196	–
60–69 years	114	–
\geq 70 years	35	–
% with family history of RA	16.8	–
RA disease status subgroup, no. of subjects		
Multilating disease	58	–
More erosive subset	279	–
Less erosive subset	423	–

published observations and ref. 18). We therefore considered that the region flanked by markers *D7S486* and *D7S2513*, an \sim 1-LOD-drop interval from *D7S2560*, might contain RA susceptibility or RA severity-modifying genes. The interval spanned \sim 25.3 Mb of the genomic region and contained 179 genes, as shown in the National Center for Biotechnology Information database (Map View Build 34, version 2; statistics for Build 34, version 2 are available online at <http://www.ncbi.nlm.nih.gov/mapview/stats/BuildStats.cgi?taxid=9606&build=34&ver=2>). To establish a useful marker set for positional candidate genes, we searched public and Celera databases and developed TaqMan assays for 728 SNPs. Unfortunately, we were unable to develop SNP assays for 72 genes located in the interval of interest because either no SNPs were available in the databases, the SNPs were not polymorphic in our initial screening of 46 Japanese control subjects, or it was difficult to design the TaqMan probes.

Findings of the case-control association analysis. For the association analysis, we applied a 2-stage case-control analysis strategy in which each stage consisted of 380 unrelated RA patients and 380 non-RA controls. Clinical characteristics, such as sex distribution, average age at disease onset, and RA subtypes, were comparable between the subjects evaluated in the 2 stages. The power to detect an association was calculated to be 0.5125, as determined with the use of the Genetic Power Calculator (28) (online at [http://](http://statgen.iop.kcl.ac.uk/gpc/cc2.html)

statgen.iop.kcl.ac.uk/gpc/cc2.html), based on the following conditions: 380 RA cases and 380 non-RA controls, an RA prevalence of 1%, a relative risk of f_1/f_0 and f_2/f_0 (penetrance for the disease susceptibility allele) of 1.3 and 1.5, respectively, a marker allele frequency of 0.45, and an alpha level of 0.05.

Of the 728 SNPs genotyped in the first set of DNA samples, 69 were excluded from the association analysis either because the minor allele frequency was <0.15 or because the Hardy-Weinberg equilibrium test yielded $P < 0.05$ in control subjects. Of these 659 SNPs, 48 SNPs (6.59% of the initial number of SNPs) in 13 gene regions exhibited a significant association with RA ($P < 0.05$). The 48 candidate SNPs were subsequently genotyped in the second set of 380 RA cases and 380 non-RA controls. Three of the 48 SNPs were excluded from the second-stage association analysis based on the same criteria. Five of the remaining 45 SNPs (SNPs 411, 441, 442, 443, and 445), or 0.69% of the initial number of SNPs, again showed a significant association ($P < 0.05$). When the genotyping data from each stage were combined, SNPs 441, 442, 443, and 445 yielded the smaller P values, while SNP 411 no longer showed a statistically significant association with RA. Although the association with SNP 411 was significant in both the first and the second sets of samples, the allele frequency was reversed in the 2 sets, making the combined association result insignificant.

Table 2. Summary of the association of SNPs in the *SEC8LI* gene between rheumatoid arthritis (RA) cases and non-RA controls

SNP ID*	dbSNP ID or Celera CVID†	Position‡	Allele 1/2	Location	Allele frequency			OR (95% CI)§	No. of subjects	
					Controls	Cases	P		Controls	Cases
SNP 428	hCV30729	132,415,346	G/A	Intron 4	0.330	0.309	0.39	1.10 (0.89–1.36)	379	380
SNP 429	rs6951416	132,508,490	A/G	Intron 7	0.625	0.644	0.45	1.08 (0.88–1.34)	380	379
SNP 430	rs7783217	132,520,329	G/C	Intron 7	0.376	0.354	0.38	1.10 (0.89–1.35)	380	378
SNP 431	hCV3156195	132,538,954	C/G	Intron 7	0.691	0.720	0.21	1.15 (0.92–1.44)	380	379
SNP 432	rs6467491	132,554,973	G/A	Intron 7	0.627	0.644	0.49	1.08 (0.87–1.33)	379	379
SNP 433	rs7785217	132,562,876	T/C	Intron 7	0.434	0.401	0.19	1.15 (0.93–1.40)	380	380
SNP 434	rs6467494	132,591,254	T/C	Intron 9	0.424	0.378	0.067	1.21 (0.98–1.49)	375	376
SNP 435	hCV3291515	132,617,031	T/C	Intron 9	0.697	0.721	0.31	1.12 (0.90–1.40)	380	380
SNP 436	rs12533546	132,668,063	C/A	Intron 9	0.326	0.299	0.25	1.13 (0.92–1.41)	380	380
SNP 437	rs1477422	132,676,423	A/G	Intron 9	0.743	0.763	0.37	1.11 (0.88–1.40)	379	379
SNP 438	rs7801515	132,682,025	A/G	Intron 9	0.315	0.281	0.14	1.18 (0.95–1.47)	379	379
SNP 439	rs3909550	132,684,679	T/G	Intron 9	0.321	0.300	0.38	1.10 (0.89–1.37)	380	380
SNP 440	rs6949187	132,763,209	G/C	Intron 10	0.208	0.184	0.22	1.17 (0.91–1.51)	379	376
SNP 441	rs2911495	132,747,619	C/T	Intron 10	0.533	0.605	0.000059	1.34 (1.16–1.55)	760	756
SNP 442	rs763645	132,759,642	C/T	Intron 10	0.483	0.419	0.00041	1.30 (1.12–1.49)	758	759
SNP 443	rs1014442	132,765,495	A/T	Intron 10	0.472	0.406	0.00023	1.31 (1.14–1.51)	759	760
SNP 444	rs6944149	132,797,579	A/C	Intron 10	0.504	0.571	0.0085	1.26 (1.09–1.46)	379	378
SNP 445	rs1861797	132,830,087	A/G	Intron 10	0.357	0.294	0.00023	1.33 (1.14–1.55)	759	759
SNP 446	rs7804463	132,863,865	C/T	Intron 10	0.489	0.429	0.018	1.28 (1.04–1.56)	380	380
SNP 447	hCV11428778	132,923,084	A/G	Intron 11	0.450	0.408	0.097	1.19 (0.97–1.46)	380	380
SNP 448	rs7800126	132,937,485	C/A	Intron 11	0.528	0.576	0.060	1.22 (0.99–1.49)	380	376
SNP 449	hCV2886729	132,982,552	T/C	Intron 11	0.509	0.455	0.0034	1.24 (1.07–1.43)	758	748
SNP 450	rs6467506	132,988,539	A/G	Intron 11	0.496	0.546	0.0058	1.22 (1.06–1.41)	756	758
SNP 451	rs2042456	132,996,576	A/G	Exon 12	0.503	0.453	0.0053	1.23 (1.06–1.41)	757	748
SNP 452	rs2042455	132,996,674	A/G	Intron 12	0.497	0.549	0.0041	1.23 (1.07–1.42)	759	758
SNP 453	rs2042454	132,996,839	G/A	Intron 12	0.497	0.544	0.0088	1.21 (1.05–1.40)	757	756
SNP 454	rs2042453	132,996,859	G/A	Intron 12	0.503	0.450	0.0040	1.23 (1.07–1.42)	757	755
SNP 455	rs2042452	132,997,222	G/A	Intron 12	0.495	0.547	0.0040	1.23 (1.07–1.42)	759	757
SNP 456	rs6975860	133,005,713	T/G	Intron 12	0.505	0.447	0.024	1.26 (1.03–1.54)	379	379
SNP 457	rs889825	133,012,356	G/A	Intron 12	0.497	0.548	0.0050	1.23 (1.06–1.41)	758	756
SNP 458	rs958405	133,013,280	G/A	Intron 12	0.498	0.548	0.0058	1.22 (1.06–1.41)	759	760
SNP 459	rs10226267	133,017,991	A/G	Intron 12	0.501	0.451	0.0056	1.22 (1.06–1.41)	756	755
SNP 460	rs10263851	133,027,385	C/T	Intron 13	0.502	0.554	0.0039	1.23 (1.07–1.42)	760	755
SNP 461	rs6954610	133,035,115	G/A	Intron 13	0.501	0.556	0.0025	1.25 (1.08–1.44)	759	756
SNP 462	rs3823572	133,096,728	G/A	Intron 14	0.509	0.478	0.22	1.13 (0.93–1.39)	380	380
SNP 463	rs10488419	133,110,024	T/C	Intron 17	0.621	0.596	0.31	1.11 (0.90–1.37)	379	380
SNP 464	hCV11579731	133,120,700	C/A	Intron 17	0.618	0.591	0.28	1.12 (0.91–1.38)	380	379
SNP 465	rs12673178	133,134,959	G/A	Intron 17	0.383	0.412	0.25	1.13 (0.92–1.38)	380	379
SNP 466	rs3735049	133,158,078	A/C	Intron 17	0.256	0.230	0.24	1.15 (0.91–1.45)	379	380

* Single-nucleotide polymorphisms (SNPs) located in the *SEC8LI* gene are shown. ID = identification.

† Celera Validation IDs (CVIDs; generated with the Celera Discovery System) are shown if the SNPs were not available in the Database of Single-Nucleotide Polymorphisms (dbSNP).

‡ Nucleotide positions were obtained from the University of California, Santa Cruz, Genome Bioinformatics site (online at <http://genome.ucsc.edu/>).

§ For odds ratios (ORs) <1, the inverted score is indicated. 95% CI = 95% confidence interval.

The most significant allele association was observed with SNP 441 ($P = 0.000059$, odds ratio [OR] 1.34 [95% confidence interval (95% CI) 1.16–1.55]) (Table 2). The most significant P value was observed with SNP 445 in its dominant trait ($P = 0.0000094$, OR 1.58 [95% CI 1.29–1.94]). In a case–control study such as this in which a large number of SNPs is used, the elimination of false-positive associations due to multiple testing is important. We therefore performed a conservative Bonferroni-type correction for multiple testing, and the association of RA with SNP 441 was still

significant (at $P = 0.035$ when corrected for 659 tests and $P = 0.0026$ when corrected for 45 tests). The population-attributable risk calculated for SNP 441 was 40% at a frequency of 0.57 for the SNP 441 risk allele and at estimated genotype risk ratios r_1 and r_2 of 1.075 and 1.708, respectively, for the C/T genotype and for the C/C genotype relative to the T/T genotype. SNP 441 is located 3.9 Mb from the peak linkage marker *D7S2560*.

SNPs 441, 442, 443, and 445 were all positioned within an ~82-kb genomic region and located in intron 10 of the *SEC8LI* gene, which encodes a subunit of the

Table 3. Other significant results of association analysis in the combined data set

SNP ID*	dbSNP ID†	Gene symbol	Allele 1/2	Trait	Allele frequency		P	OR (95% CI)‡	No. of subjects	
					Controls	Cases			Controls	Cases
SNP 005	rs2270188	CAV2	G/T	Allele frequency	0.302	0.258	0.0065	1.25 (1.06–1.46)	755	757
SNP 006	rs2270189	CAV2	G/A	Allele frequency	0.302	0.257	0.0054	1.25 (1.07–1.47)	756	759
SNP 098	rs41690	FLJ21986	T/C	Allele 1 dominant	0.398	0.350	0.00069	1.43 (1.16–1.76)	757	758
SNP 101	rs7782552	FLJ21986	A/G	Allele 1 recessive	0.417	0.456	0.0045	1.45 (1.12–1.88)	757	760
SNP 368	rs1013368	PODXL	C/T	Allele frequency	0.198	0.242	0.0040	1.29 (1.08–1.53)	756	759
SNP 623	rs999229	ATP6V0A4	C/T	Allele 1 recessive	0.632	0.665	0.0091	1.31 (1.07–1.61)	760	759

* Single-nucleotide polymorphisms (SNPs) with positive associations at $P < 0.01$ are shown. ID = identification.

† dbSNP = Database of Single-Nucleotide Polymorphisms.

‡ For odds ratios (ORs) < 1 , the inverted score is indicated. 95% CI = 95% confidence interval.

exocyst complex. Notably, 12 SNPs (SNPs 449–455 and 457–461) that were located within the *SEC8L1* gene and were negative in the second-stage analysis became positively associated with RA in the analysis of the combined data set (first-stage plus second-stage analysis). The results of association analyses of SNPs in the

SEC8L1 gene are summarized in Table 2. Although SNPs 444, 446, and 456, which showed significant associations in the first set of samples, were genotyped with the second set of samples, the Hardy-Weinberg test P values for the combined data were less than 0.05. Therefore, only the results of the first-stage analysis are

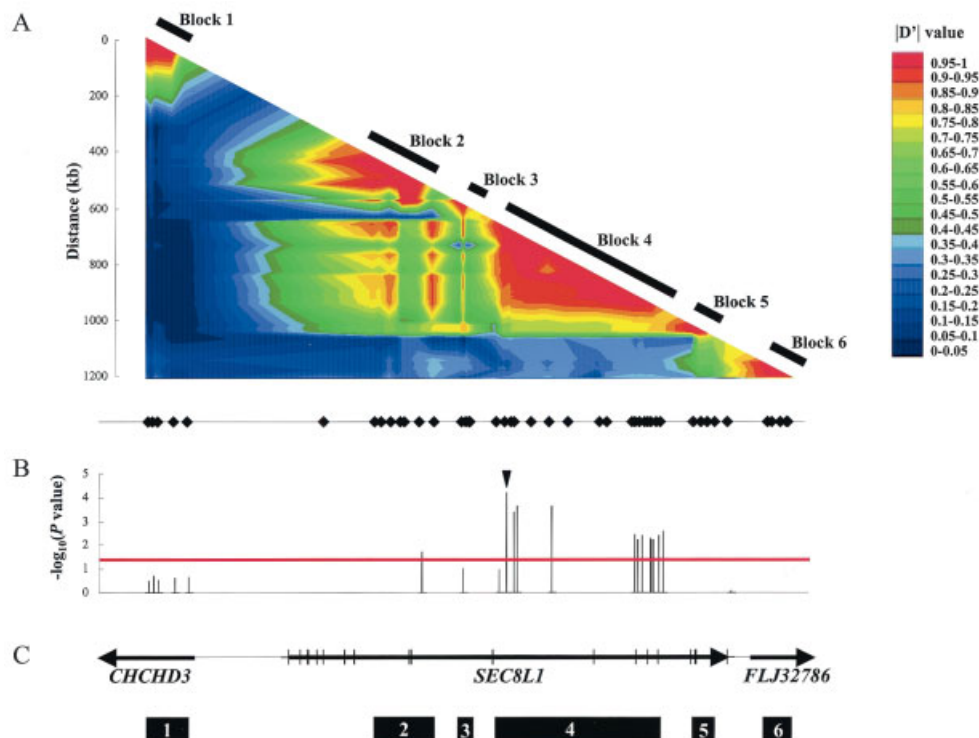


Figure 1. A, Pairwise linkage disequilibrium (LD) between single-nucleotide polymorphisms (SNPs), as measured by the $|D'|$ method, in 380 healthy Japanese control subjects. A physical map of the SNPs and a physical scale are shown below and on the left side, respectively. B, Case-control association plots $-\log_{10}(P \text{ value})$ versus physical location. The red line indicates the cutoff P value ($P = 0.05$). Only the P values of the 21 SNPs genotyped in the 1,520 study subjects (2 different groups consisting of 380 rheumatoid arthritis [RA] cases and 380 non-RA controls each) are plotted. Arrowhead indicates the physical position of SNP 441, which showed the most significant allele association. C, Genomic structure around the *SEC8L1* gene and location of the 6 LD blocks in this region. Vertical lines indicate coding exons of the *SEC8L1* gene.

Table 4. Haplotype frequencies and structures*

	Haplotype frequency		SNP ID					<i>P</i>	Adjusted <i>P</i> †
	Controls	Cases	SNP 440	SNP 441	SNP 445	SNP 447	SNP 449		
Haplotype 1	0.47	0.52	G	C	G	G	C	0.0015	0.001
Haplotype 2	0.16	0.11	G	T	A	A	T	0.0000062	<0.0001
Haplotype 3	0.16	0.15	C	T	A	A	T	0.33	0.33
Haplotype 4	0.09	0.09	G	T	G	A	T	0.94	0.94

* Haplotypes with a frequency >0.05 were compared between rheumatoid arthritis (RA) cases and non-RA controls. SNP = single-nucleotide polymorphism; ID = identification.

† Adjusted for multiple comparisons by permutation testing (10,000 permutations).

shown in Table 2, and these 3 SNPs were excluded from the LD analysis.

Six other SNPs in 4 genes did not show significant allele associations in the second-stage analysis. However, they showed strong associations ($P < 0.01$) with RA in at least 1 trait in the combined data analysis (Table 3).

Findings of the LD and haplotype analyses. The *SEC8L1* gene consists of 18 exons and spans an ~800-kb genomic region. Previous studies found that LD between SNPs usually extends for 10–60 kb for common alleles and may extend several hundred kilobases in some genomic regions. To confirm that the *SEC8L1* gene itself was responsible for the observed associations, we evaluated the patterns of LD around the *SEC8L1* gene and 2 flanking genes, *CHCHD3* and *FLJ32786*. We observed 6 highly structured LD blocks defined at a threshold of $|D'| > 0.9$ (Figure 1A). Almost all SNPs located within block 4 (SNPs 441–461) showed significant associations with RA. This block covered the 300-kb genomic segment between introns 10 and 13 of the *SEC8L1* gene (Figures 1B and C). No other known genes were located in this LD block, suggesting that the observed associations with RA were likely to be due to the *SEC8L1* gene itself.

We assumed that the SNPs in block 4 may act in combination to modify the risk of RA, and we tested this hypothesis by haplotype analysis. Although there were 22 SNPs in block 4 (SNPs 440–461), 3 of them (SNPs 444, 446, and 456) were excluded from the analysis. We divided the 19 remaining SNPs in block 4 into 5 groups that were in strong LD with each other ($r^2 > 0.9$). A haplotype was constructed with 5 SNPs (SNPs 440, 441, 445, 447, and 449) chosen from each group. SNPs 440 and 447, which were originally excluded from second-stage analysis, were genotyped in the second set of samples for haplotype analysis. Four major haplotypes (haplotypes 1–4) were observed (Table 4). Haplotype 1 (G;C;G;G;C) was the most common and was positively associated with RA (disease-risk haplotype) ($P =$

0.0015), while haplotype 2 (G;T;A;A;T) was negatively associated with RA (disease-protective haplotype) ($P = 0.0000062$). P values adjusted for multiple comparisons by permutation testing were as follows: $P = 0.001$ for haplotype 1 and $P < 0.0001$ for haplotype 2.

SEC8L1 SNPs exhibiting significant associations were synonymous (SNP 451) or intronic, with no presumable functional consequences. To search for new SNPs in the *SEC8L1* coding exons, we resequenced exons 11, 12, and 13 of the *SEC8L1* gene within block 4. However, no nonsynonymous or synonymous SNPs, except for SNP 451, were found. Nonsynonymous SNP rs10269237 is recorded in dbSNP in exon 14. Although exon 14 is outside of block 4, we resequenced this exon. The G allele of rs10269237 was not detected in the 48 Japanese chromosomes, however, which proves that the SNP is not polymorphic in the Japanese population.

***SEC8L1* expression in human tissue.** To investigate the expression patterns of *SEC8L1* in human tissues, we performed real-time quantitative PCR analysis on human tissues. The *SEC8L1* gene was ubiquitously expressed in normal tissues, with the highest expression in skeletal muscle. Moderate expression was detected in FLS generated from RA patients (Figure 2).

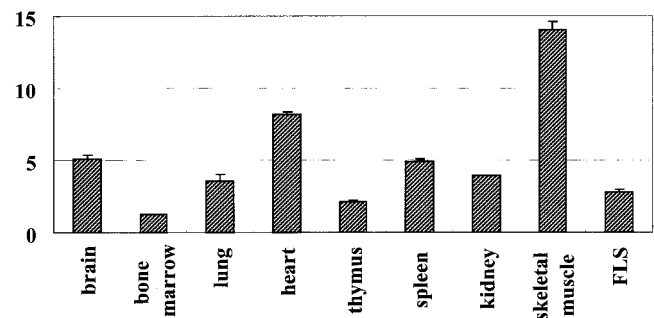


Figure 2. Expression patterns of the *SEC8L1* gene in human tissues relative to the bone marrow. Relative expression of the *SEC8L1* gene is normalized to the amount of β -actin mRNA. FLS = fibroblast-like synoviocytes.

DISCUSSION

In this study, we conducted a SNP-based fine-mapping analysis for RA susceptibility genes in the region of chromosome 7q31–34. We chose this region because chromosome 7q31–34 was one of the potential susceptibility loci in our genome-wide scan of a limited number of samples from 53 Japanese families and because the region has been reported to be a candidate locus for multiple sclerosis and type 1 diabetes (29,30). Familial clustering of different autoimmune diseases (31) and comanifestation of multiple autoimmune diseases in the same individual or family have been reported, and overlapping of susceptibility loci for autoimmune or inflammatory diseases in humans and in animal models has also been reported (32). We therefore presumed the existence of a common susceptibility gene(s) on chromosome 7q31–34 that would predispose a person to the development of autoimmunity.

Because genetic susceptibility to common complex disorders such as RA is supposedly influenced by relatively common polymorphisms (33), we used evenly spaced and dense SNPs (i.e., 1 SNP every 10–20 kb in each gene locus) and common SNPs with a minor allele frequency >0.15 in this study. We applied a 2-stage case-control analysis strategy to reduce the time and cost of genotyping and to avoid false-positive associations, which is an important issue in association studies. Among the 728 SNPs tested, only 4 (SNPs 441, 442, 443, and 445) showed significant and repeated associations with RA in both stages of analysis. Of note, SNP 441, which showed the most significant allele association ($P = 0.000059$), was also statistically significant when a conservative Bonferroni-type correction for multiple testing was applied ($P = 0.039$). SNP 441 is located only 3.9 Mb from the peak linkage marker. This suggests that the results of LD mapping in sporadic RA cases in our study do not contradict those obtained in our previous genome-wide scan in cases of familial RA. Nineteen SNPs (SNPs 441–446 and SNPs 449–461) showed positive ORs of 1.21–1.34. The most significant OR, 1.34, which had a 95% CI of 1.16–1.55, was observed for SNP 441. Despite its modest OR value, but because of its high frequency (57%), the more-common risk allele SNP 441 conferred a relatively large population-attributable risk, accounting for 40% of RA in the study population.

Besides the SNPs in the *SEC8L1* gene that showed strongly positive associations with RA, we detected 6 SNPs in 4 genes that showed positive associations at $P < 0.01$ (ranging from $P = 0.00069$ to $P = 0.0091$), as summarized in Table 3. Although these SNPs

did not satisfy the conservative Bonferroni-type correction for multiple testing, these genes remain as candidate susceptibility genes for RA.

SNPs 441, 442, 443, and 445 were all located in intron 10 of the *SEC8L1* gene, which encodes a subunit of the exocyst complex. Because these SNPs were likely nonfunctional but were in LD with causative susceptibility variants, we performed LD mapping and haplotype analysis to pinpoint the susceptibility locus. We found that associations with RA were confined to a single LD block (block 4) spanning a 300-kb genomic region. Thus far, *SEC8L1* is the only known gene in block 4, although 3 hypothetical genes (BC038554.1, BC040296.1, and BC022408.1) have been deposited in public databases. We therefore used RT-PCR to test their expression in various human tissues. No amplification products were obtained, which suggests that the expression of these 3 genes in specific tissues was limited or that they are nonexistent (data not shown). Haplotype analysis identified disease-risk and disease-protective haplotypes in block 4. The most significant haplotype (at $P = 0.0000062$) was disease-protective haplotype 2, which was 1 order of magnitude more significant than the most significant single-point P value of 0.000059 for SNP 441. Because both P values are highly significant, the single SNP 441 and the haplotype including SNP 441 are supposedly equally important in determining disease susceptibility.

In a study of Alzheimer's disease, association was detected by haplotyping methods using SNPs, even if the functional allele was not typed (34). Because statistically significant SNPs and haplotypes were identified in block 4, it is supposed that functional variants are found in block 4. Resequencing coding exons and exon-intron boundaries of the *SEC8L1* gene within block 4 (exons 11–13) revealed no nonsynonymous (amino acid-altering) substitutions or potential alternative splicing variants. Therefore, we were unable to find the causative susceptibility variants in the coding sequences. The intronic sequences of ~300 kb remain to be searched for polymorphisms in noncoding regions because regulatory sequences, such as transcription factor-binding sites, may cause the disease susceptibility.

The *SEC8L1* gene encodes the SEC-8 protein, a component of the exocyst complex. In budding yeast, the Sec8 complex is essential for exocytosis, and it directs vectorial targeting of Golgi-derived secretory vesicles to the growing bud-tip (35). Like yeast, the mammalian exocyst has been directly implicated in post-Golgi targeting of secretory vesicles to discrete membrane sites (36). In the present study, the *SEC8L1* gene was shown

to be ubiquitously expressed in human tissues. Only moderate expression of the *SEC8L1* gene was observed in FLS from RA patients. FLS is only one component of RA synoviocytes, and other synovial cells also secrete cytokines. The moderate gene expression results do not conflict with an association between *SEC8L1* and RA. Although expression of the *SEC8L1* gene was also detected in immunologic tissues and FLS from RA patients, the precise etiology of how *SEC8L1/Sec8* interacts in the pathogenesis of RA remains unknown.

Macrophages, FLS, and CD4+ T cells, which dominate the synovium in the RA joint, are known to secrete a broad array of cytokines, including TNF α , interleukin-1 β (IL-1 β), IL-6, granulocyte-macrophage colony-stimulating factor, transforming growth factor β , IL-1 receptor antagonist, and others. The cytokines produced by synovial fibroblasts or other infiltrating cells activate either themselves or neighboring cells, such as osteoclasts, through their receptors and lead to the joint destruction in RA.

It has been reported that microvesicle shedding is a major secretory pathway for the rapid release of IL-1 β from activated monocytes (37). Cytokines in mouse macrophages are secreted by small vesicles that originate in the Golgi complex, and they fuse with the cell membrane (38). Based on these observations, we speculate that *SEC8L1/Sec8* might play a role in the secretion of cytokines or in the regulation of plasma membrane receptors in immune cells. This remains to be tested, however.

In conclusion, we performed a large-scale locus-wide association analysis of the candidate region on chromosome 7 in a Japanese population and found strong evidence of an association between the *SEC8L1* gene and susceptibility to RA. We propose that *SEC8L1*, which encodes a component of the exocyst complex, is a candidate susceptibility gene for RA in the Japanese population. To confirm our observations, additional association analyses using an independent sample set or different ethnic groups and functional analysis of the SEC-8 protein will be necessary.

ACKNOWLEDGMENTS

We are grateful to the RA patients and the control subjects for their participation in this study. We greatly appreciate the helpful advice of Hitoshi Kato and the technical assistance of Rumi Katashima and Kazue Tsugawa. We wish to thank members of the Institute for Genome Research for helpful discussions and assistance.

REFERENCES

1. Felson D. Epidemiology of the rheumatic diseases. In: Koopman WJ, editor. *Arthritis and allied conditions: a textbook of rheumatology*. 13th ed. Philadelphia: Williams & Wilkins; 1996. p. 3–34.
2. Lawrence JS. Heberden Oration, 1969: rheumatoid arthritis: nature or nurture? *Ann Rheum Dis* 1970;29:357–79.
3. Del Junco D, Luthra HS, Annegers JF, Worthington JW, Kurland LT. The familial aggregation of rheumatoid arthritis and its relationship to the HLA-DR4 association. *Am J Epidemiol* 1984; 119:813–29.
4. Wordsworth BP, Lanchbury JS, Sakkas LI, Welsh KI, Panayi GS, Bell JI. HLA-DR4 subtype frequencies in rheumatoid arthritis indicate that DRB1 is the major susceptibility locus within the HLA class II region. *Proc Natl Acad Sci U S A* 1989;86:10049–53.
5. Deighton CM, Walker DJ, Griffiths ID, Roberts DF. The contribution of HLA to rheumatoid arthritis. *Clin Genet* 1989;36: 178–82.
6. Barton A, Ollier W. Genetic approaches to the investigation of rheumatoid arthritis. *Curr Opin Rheumatol* 2002;14:260–9.
7. 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.
8. 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.
9. 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.
10. 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.
11. Jawaheer D, Seldin MF, Amos CI, Chen WV, Shigeta R, Etzel C, et al, for the North American Rheumatoid Arthritis Consortium. Screening the genome for rheumatoid arthritis susceptibility genes: a replication study and combined analysis of 512 multicase families. *Arthritis Rheum* 2003;48:906–16.
12. Eyre S, Barton A, Shephard N, Hinks A, Brintnell W, MacKay K, et al. Investigation of susceptibility loci identified in the UK rheumatoid arthritis whole-genome scan in a further series of 217 UK affected sibling pairs. *Arthritis Rheum* 2004;50:729–35.
13. 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.
14. 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.
15. 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.
16. Suzuki A, Yamada R, Chang X, Tokuhiro S, Sawada T, Suzuki M, et al. Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat Genet* 2003;34:395–402.
17. Tokuhiro S, Yamada R, Chang X, Suzuki A, Kochi Y, Sawada T, et al. An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nat Genet* 2003;35:341–8.
18. Takeuchi F, Moritani M, Nabeta H, Mori M, Goto M, Hanyu T, et al. Nine Japanese RA susceptibility genes areas, TIRA1-9, by preliminary genome-wide screening. *Proceedings of the 10th Asia-Pacific League of Associations for Rheumatology Congress*; 2002 Dec 1–6; Bangkok, Thailand.
19. Arnett FC, Edworthy 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.
20. Ochi T, Iwase R, Yonemasu K, Matsukawa M, Yoneda M, Yukioka M, et al. Natural course of joint destruction and fluctuation of serum C1q levels in patients with rheumatoid arthritis. *Arthritis Rheum* 1988;31:37–43.
21. Mishiro T, Nakano S, Takahara S, Miki M, Nakamura Y, Yasuoka S, et al. Relationship between cathepsin B and thrombin in rheumatoid arthritis. *J Rheumatol* 2004;31:1265–73.
22. Nielsen DM, Ehm MG, Weir BS. Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *Am J Hum Genet* 1998;63:1531–40.
23. Hill WG. RA: linkage disequilibrium in finite populations. *Theor Appl Genet* 1968;38:226–31.
24. Lewontin RC. The interaction of selection and linkage. I. General considerations: heterotic models. *Genetics* 1964;49:49–67.
25. Abecasis GR, Cookson WO. GOLD: graphical overview of linkage disequilibrium. *Bioinformatics* 2000;16:182–3.
26. Esterbauer H, Schneitler C, Oberkofler H, Ebenbichler C, Paulweber B, Sandhofer F, et al. A common polymorphism in the promoter of UCP2 is associated with decreased risk of obesity in middle-aged humans. *Nat Genet* 2001;28:178–83.
27. Nyholt DR. All LODs are not created equal. *Am J Hum Genet* 2000;67:282–8.
28. Purcell S, Cherny SS, Sham PC. Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003;19:149–50.
29. Haines JL, Ter-Minassian M, Bazyk A, Gusella JF, Kim DJ, Terwedow H, et al, and the Multiple Sclerosis Genetics Group. A complete genomic screen for multiple sclerosis underscores a role for the major histocompatibility complex. *Nat Genet* 1996;13:469–71.
30. Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, et al. A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 1994;371:130–6.
31. Vyse TJ, Todd JA. Genetic analysis of autoimmune disease. *Cell* 1996;85:311–8.
32. Becker KG. Comparative genetics of type 1 diabetes and autoimmune disease: common loci, common pathways? *Diabetes* 1999;48:1353–8.
33. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet* 2001;17:502–10.
34. Fallin D, Cohen A, Essioux L, Chumakov I, Blumenfeld M, Cohen D, et al. Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE locus variation and Alzheimer's disease. *Genome Res* 2001;11:143–51.
35. Guo W, Sacher M, Barrowman J, Ferro-Novick S, Novick P. Protein complexes in transport vesicle targeting. *Trends Cell Biol* 2000;10:251–5.
36. Hsu SC, Hazuka CD, Foletti DL, Scheller RH. Targeting vesicles to specific sites on the plasma membrane: the role of the sec6/8 complex. *Trends Cell Biol* 1999;9:150–3.
37. MacKenzie A, Wilson HL, Kiss-Toth E, Dower SK, North RA, Surprenant A. Rapid secretion of interleukin-1 β by microvesicle shedding. *Immunity* 2001;15:825–35.
38. Fedorko ME. Proposed role for small cytoplasmic vesicles in cytokine secretion by mouse macrophages. *Med Hypotheses* 2004;62:911–3.