European Journal of Human Genetics (2006) 14, 69–78

www.nature.com/ejhg

ARTICLE

Haplotype analysis of tumour necrosis factor receptor genes in 1p36: no evidence for association with systemic lupus erythematosus

Sapna Chadha[1], Katie Miller[2], Lisa Farwell[2], Steven Sacks[3], Mark J Daly[2], John D. Rioux[2,4] and Timothy J Vyse*[,1,4]

[1]*Rheumatology Section, Faculty of Medicine, Imperial College, Hammersmith Hospital, London, UK;* [2]*Broad Institute, Centre for Genome Research, Massachusetts Institute of Technology, Cambridge, MA, USA;* [3]*Renal Medicine, Guy's and St. Thomas' Hospital, London, UK*

Systemic lupus erythematosus (SLE) is a systemic autoimmune disease with partially understood aetiology. The 1p36 region has been previously linked with SLE and harbours tumour necrosis factor receptor (*TNFR*) genes. Functional and genetic data implicate their gene products in SLE and other autoimmune diseases. In all, single-nucleotide polymorphisms (SNPs) across *TNFRSF14* (*HVEM*), and 43 SNPs across the *TNFRSF8* (*CD30*) and *TNFRSF1B* (*CD120B*) locus were investigated for linkage disequilibrium (LD) and haplotype analysis in European-Caucasians. Strong LD was observed across *HVEM* and *CD120B*, and little LD and recombination across *CD30*. We also examined the association of SNPs and haplotypes in *HVEM, CD30 and CD120B* with SLE in European-Caucasians. There was no evidence of association for these genes in 456 European-Caucasian families with SLE from UK. Haplotype tagging SNPs are made known across areas of strong LD, which will facilitate analysis for susceptibility in other diseases.
European Journal of Human Genetics (2006) **14,** 69–78. doi:10.1038/sj.ejhg.5201527; published online 23 November 2005

Keywords: SLE; *TNFR*; *HVEM*; *CD30*; *CD120B*; SNP; LD; haplotype

Introduction

Systemic lupus erythematosus (SLE) is a complex autoimmune disease of unclear aetiology with diverse clinical features. Its prevalence is estimated to be about one in 2000 people in some population.[1] There is a strong gender bias, with a female:male ratio of about 9:1 seen between the ages of 15 and 50 years, and ethnicity influences this disease.[1–3] A feature of SLE is the generation of a variety of autoantibodies to cell surface and nuclear antigens. These autoantibodies contribute to end-organ damage via a variety of mechanisms and the formation of immune complexes can result in glomerulonephritis, arthritis, rashes, serositis and vasculitis.

SLE appears to develop through the interaction of largely unknown environment factors and the genetic composition of the individual. Evidence for the genetic component is strongly illustrated by concordance rates in monozygotic twins ranging between 25 and 69%, while the rate in dizygotic twins is only 1–2%.[4] Family-based studies show familial recurrence risks of siblings of probands in the population as a whole (λ_s) is 20 for SLE. Linkage analysis of SLE in ethnically diverse families has identified at least 20 potential susceptibility loci.[5–9]

The 1p36 region has been linked to SLE in four investigations[5,6,7,10] and is orthologous to the murine susceptibility interval on chromosome 4, which harbours *Nba1 and Lbw2* intervals from the NZB/WF1 model.[11]

*Correspondence: Dr TJ Vyse, Rheumatology Section, Faculty of Medicine, Imperial College, Hammersmith Hospital, London W12 0NN, UK. Tel: +44 20 83832339; Fax: +44 20 87433109;
E-mail: t.vyse@imperial.ac.uk
[4]These authors contributed equally to this work
Received 27 April 2005; revised 4 October 2005; accepted 6 October 2005; published online 23 November 2005

Several immunologically relevant genes have been mapped within this interval and include among them the tumour necrosis factor receptor (*TNFR*) superfamily of genes. Members of this superfamily are expressed on diverse cells of the immune system and play a central role in immune regulation and in apoptosis. We selected *TNFRSF14* (*HVEM*), *TNFRSF8* (*CD30*) and *TNFRSF1B* (*CD120B, TNFR2*), as good candidate genes for SLE for study within the 1p36 interval based on position and function.

HVEM is broadly expressed on cells of the immune system and is involved in T-cell activation. Three ligands have been identified for HVEM: the HSV glycoprotein D, and two members of the TNF family, lymphotoxin α3 and the newly described member LIGHT (TL4), which is produced by activated T cells.[12] CD30 is a receptor for CD30 ligand and is expressed on activated but not on resting T and B cells. In T cells, CD30 signalling stimulates proliferation, cytokine production and induces susceptibility to apoptosis. CD30 signalling protects against autoimmunity by controlling the expansion of autoreactive CD8 effector T cells.[13] CD120B is expressed predominantly in cells of stimulated T- and B-lymphocytes and, together with CD120A, constitute the major receptors for tumour necrosis factor alpha. In SLE patients, soluble CD120B has been found to correlate with disease activity.[14]

There have been numerous studies investigating linkage and association of the *CD120B* with autoimmunity in humans. A single-nucleotide polymorphism (SNP) in exon 6 (M196R) has been associated with susceptibility to SLE.[15,16] *CD30* has been implicated for type I diabetes in NOD mice[17] and it has been shown that anti-CD30L monoclonal antibody limits the development of spontaneous diabetes in NOD mice.[18] Since *CD30* neighbours *CD120B* and there are no other known genes between them, we employed a linkage disequilibrium (LD) and haplotype-based approach to study these two genes for linkage and association. This study has defined a high-resolution haplotype structure for *HVEM*, *CD30* and *CD120B*, and examined the association of the individual SNPs and haplotypes with SLE in European-Caucasians from the UK.

Materials and methods

Family collection and DNA

A large collection of SLE nuclear families has been obtained from the UK with predominantly one affected offspring per family. The demographic details of the families studied are summarised in Supplementary Table 6. The families were randomly allocated into cohorts. Cohort 1 was used as an initial screen and cohort 2 was utilised to follow up on SNPs showing hints of association ($P<0.05$) from the initial screen. Samples from both parents were available for 65% of the cases, and siblings were also collected where available. In the case of the single parent families, samples were always taken from siblings. All participants signed informed consents prior to blood and data collection and study protocols were approved by the London multicentre research ethics committee. The clinical manifestations of SLE are variable and diagnosis criteria have been established by the American College of Rheumatology (ACR).[19,20] Patients fulfilling four of the 11 criteria are considered to have SLE. Consequently, SLE-affected individuals have different disease phenotypes. Patients were classified as having renal lupus using the ACR criteria. Further clinical information was obtained from individuals by interview and completion of a health questionnaire.[21] All patients and siblings sera were screened for anti-cardiolipin IgG and were scored as positive if results were >12 arbitrary ELISA units.[22] Thrombosis includes self-reported venous thrombosis, either deep vein thrombosis and/or pulmonary embolism. DNA samples were prepared from 40 ml of blood by phenol–chloroform extraction.[23] DNA concentration was obtained by PicoGreen® quantification. Clinical and genotyping data were stored on a specialised customised database (BioComputing Oy) in a coded anonymised manner.

SNP selection

In silico SNP hunting was employed using published genomic and cDNA sequence data. The largest public databases of SNPs, the SNP consortium (http://snp.cshl.org/), GeneSNPs (http://www.genome.utah.edu/genesnps/), SNPper (http://bio.chip.org:8080/bio), dbSNP (http://ncbi.nlm.nih.gov/SNP), and HGBASE (http://hgbase.interactiva.de/) were explored for SNPs. These resources were constantly changing and were scrutinised on multiple occasions. The majority of these markers in the databases occur outside the coding regions of genes. Initial SNPs were chosen based on coding areas and noncoding SNPs chosen on spacing throughout the genes, validation and submitter information. SNPs that fell in low complexity or repeat regions identified by inputting the sequence area into the program REPEATMASKER (http://ftp.genome.washington.edu/RM/Repeatmasker) were dismissed at this stage. During this study, novel coding SNPs were provided by the SNP Discovery Group at the Whitehead Institute. Reference cluster ID's (rs#) were given for *in silico* SNPs and Whitehead ID's (G6073# and G551#) were used for novel SNPs. SNPs were removed from further analysis if they did not genotype well or failed completely in assay, were monomorphic, did not fall within Hardy–Weinberg equilibrium or had >5 pedigree errors.

Genotyping

The genotyping methodology employed has been described previously.[24,25] In brief, genotyping was performed by multiplexing SNP assays using the Sequenom® MassARRAY™ genotyping system (Sequenom, Hamburg, Germany). Multiplex assays were designed for three to six SNPs

per assay. Primer design for multiplex was carried out using SpectroDESIGNER™1.3.4 and MassARRAY™ Assay Design 2.0 software (Sequenom, Hamburg, Germany).

Statistical analysis

Hardy–Weinberg equilibrium testing was carried out using the exact test courtesy of G Abecasis and J Wigginton (University of Michigan Center for Statistical Genetics). Pedigree checking was performed using PedCheck 1.1.[26] All genotyping data was run through to check for Mendelian inconsistencies in pedigree data. The allele frequency thresholds for significant differences between populations were calculated by dividing $P=0.05$ by the number of singleton SNPs (SNPs not in any LD) multiplied by the number of common haplotypes seen. Two measures of LD, squared correlation coefficient (r^2) and Lewontin's standardised disequilibrium coefficient (D'), were computed between pairs of SNPs from founder chromosomes through use of the Haploview 3.0.[27] Cohort 1 was used to perform the LD analysis. SNPs below 5% minor allele frequency were not included in LD or haplotype analysis. The 95% confidence interval boundaries for pairs of SNPs were used to estimate recombination.[28] Multilocus D' between haplotype blocks is calculated by computing the 2×2 D' score of each allele at first locus with allele at second locus and then taking a weighted average of these values.[29] Haplotype analysis was conducted using the Haploview 3.0. Haplotypes with $<1\%$ frequency were not included in the analysis. Haplotypes were estimated using an accelerated expectation–maximisation (EM)-based algorithm, which can deal with a large number of linked loci that have moderate levels of LD. The output of the EM algorithm is the maximum-likelihood estimate and has highly accurate population frequency estimates of phased haplotypes.[30]

The transmission disequilibrium test (TDT) method evaluates whether the frequency of transmission of alleles from heterozygous parents to their affected child deviates from 50% the expected Mendelian frequency when there is no linkage. TDT analysis was performed on cohort 1, cohort 2 or combined analysis of cohorts 1 and 2. GENEHUNTER (version 3.0) was used for TDT analysis for nuclear simplex families. TRANSMIT (version 2.5.4) was used for TDT analysis of single-parent families. TRANSMIT can deal with transmission of multilocus haplotypes, even if phase is unknown, and parental genotypes may be unknown. Many of the SLE families have only one parent and this programme enables us to obtain more information from our family collection. Data from unaffected siblings may be used to narrow down the range of possible parental genotypes that need to be considered. The pedigree disequilibrium test (PDT) uses data from related nuclear families and discordant sibships from extended pedigrees for TDT analysis.[31] The most discordant siblings were selected based on being female and having negative antinuclear antibody scores. One sibling per family was chosen randomly, if >1 discordant siblings were available. The PDTsum version of the statistic was used to compare allele frequencies between affected individuals and their unaffected discordant siblings within families.

Results

We examined SNP markers across the *TNFR* genes: *HVEM*, *CD30* and *CD120B* in the 1p36 candidate interval for SLE. *HVEM* is located 9.4 Mb upstream of *CD30* and *CD120B* lies

Table 1 *HVEM* SNP summary on cohort 1

			MAF (%)[b]					
SNP	*Marker ID*	*HW[a] P-value*	*E*	*A*	*I*	*Alleles[c]*	*Position[d] from ATG (bp)*	*Position/function*
1	rs2494617	1.00	1.6	6.2	0	T/G	+16981	3′-flanking region
2	rs1555791	0.44	48.4[e]	31.2	30.8	G/C	+10759	3′-flanking region
3	rs2843396	1.00	2.6	0	0	G/C	+8479	3′-flanking region
4	rs2234167	1.00	13.3[e]	28.6	18.8	C/T	+6227	Exon 7, nonsynonymous, V241I
5	rs2234161	0.33	49.1[e]	25	30.8	A/G	+3102	Intron 3
6	rs2234160	1.00	3.9[e]	0	0	C/T	+3061	Intron 3
7	rs2234158	1.00	3.1	0	0	G/A	+1097	Exon 2, synonymous, T35T
8	rs2494614	1.00	0.05	0	0	A/C	+457	Intron 1
9	rs2234156	0.38	47.7[e]	25	26.9	T/C	+50	5′-UTR
10	rs2227313	0.33	49.1[e]	31.2	26.9	G/A	−337	5′-flanking region
11	rs2227312	0.57	49.1[e]	31.2	26.9	G/T	−440	5′-flanking region
12	rs2477684	1.00	4.1[e]	0	0	T/C	−4184	5′-flanking region

[a]HW *P*-value is a measure of deviation from Hardy–Weinberg equilibrium calculated from the observed and predicted heterozygosity in the European-Caucasian founders.
[b]Minor allele frequency (MAF) is representative of all genotypes in the major ethnic groups (European-Caucasian (E), Afro-Caribbean (A) and Indo-Asian (I)).
[c]Common/rare alleles.
[d]Positions are given from the methionine start codon, with A of methionine equivalent to zero base-pair position.
[e]SNPs where allele frequency differs significantly ($P<0.006$) between ethnic groups.

Table 2 *CD30* and *CD120B* SNP summary on cohort 1

SNP	Marker ID	HW[a] *P*-value	MAF (%)[b] E	A	I	Alleles[c]	Position[d] from ATG (bp)	Position/function
1	rs873457	0.33	32.8	20	25	C/G	−77 322	Intron 2 of MFN2 gene
2	rs946461	0.29	28	21.4	30	C/T	−32 447	5′-flanking region
3	rs1208993	0.38	27.2	25	29.2	G/A	−27 789	5′-flanking region
4	rs2002557	0.43	7	16.7	8.3	G/A	+1752	Intron 1 of CD30
5	rs2297875	0.68	44.7	28.6	36.7	G/A	+17 804	Intron 1 of CD30
6	G6073a1	1.00	20.5	30	30	G/A	+40 742	Exon 2, synonymous, A90A
7	rs2297729	0.92	8.6	8.3	3.6	C/T	+46 631	Intron 6 of CD30
8	G6073a4	1.00	2.1	0	0	A/G	+46 632	Intron 6 of CD30
9	rs501525	0.39	38.5	14.3[e]	38.5	A/G	+57 148	Intron 8 of CD30
10	rs664654	1.00	9.2	28.6	0	G/A	+61 379	Intron 10 of CD30
11	rs488733	0.37	37.1	14.3[e]	31.8[e]	C/T	+61 678	Intron 10 of CD30
12	rs685332	0.61	22.3	25	46.2	T/C	+67 543	Intron 12 of CD30
13	rs1148471	0.60	40.9	50	33.3[e]	A/G	+70 955	Intron 12 of CD30
14	rs816046	0.68	28.9	50	45.8	A/C	+73 112	Intron 13 of CD30
15	rs671106	0.89	47.8	28.6[e]	50	G/A	+74 522	Intron 13 of CD30
16	G6073a5	0.56	13.5	21.4[e]	12.5	T/C	+74 640	Exon 11, synonymous, S449S
17	rs755398	0.31	38.5	25	38.5	G/C	+78 438	Intron 14 of CD30
18	rs520311	0	0	14.3[e]	0	G/C	+79 944	3′-flanking region of CD30
19	rs544065	0	0	50[e]	0	G/A	+80 221	3′-flanking region of CD30
20	rs631272	0.29	44.8	21.4[e]	39.3	T/C	+83 098	3′-flanking region of CD30
21	rs1148466	0.65	47.9	41.7	45.8	T/C	+84 125	3′-flanking region of CD30
22	rs599098	1.00	7.2	7.1	0	A/G	+85 703	3′-flanking region of CD30
23	rs1148462	0.70	46.7	42.9	50	G/A	+87 029	3′-flanking region of CD30
24	rs648194	1.00	6.7	0	0	A/G	+89 508	3′-flanking region of CD30
25	rs1148461	0.47	44.4	41.7	50	G/A	+89 680	3′-flanking region of CD30
26	rs1148459	0.28	45	42.9	46.7	C/A	−9298	5′-flanking region of CD120B
27	rs520916	0.95	9.6	28.6[e]	0	T/C	−2697	5′-flanking region of CD120B
28	rs652625	0.48	8	21.4[e]	0	A/T	−1799	5′-flanking region of CD120B
29	rs595254	1.00	1.9	21.4[e]	0	G/C	+1660	Intron 1 of CD120B
30	rs976881	0.36	29.9	35.7	30.8	G/A	+6604	Intron 1 of CD120B
31	rs500734	1.00	31.3	21.4	14.3	T/A	+21 275	Intron 1 of CD120B
32	G551u1	0.84	27.6	14.3	18.2	T/C	+21 792	Exon 11, synonymous, K56K
33	rs1201157	0.49	36.3	42.9	35.7	C/T	+21 881	Intron 2 of CD120B
34	rs653667	0.87	40.8	25	23.3[e]	A/C	+24 658	Intron 3 of CD120B
35	G551u4	0.64	29.8[e]	21.4	14.3	T/G	+25 805	Exon 6, nonsynonymous, M196R
36	G551a6	1.00	6.2	0	0	G/A	+25 912	Exon 6, nonsynonymous, E232K
37	G551u5	1.00	2.2	0	0	C/T	+25 938	Exon 6, nonsynonymous, P240S
38	rs2275416	0.55	24.5	14.3	16.7	G/A	+27 051	Intron 7 of CD120B
39	rs235249	1.00	28.9	16.7	14.3	T/C	+31 081	Intron 8 of CD120B
40	rs235219	0.53	9.2	0	13.3	G/A	+37 205	Intron 9 of CD120B
41	rs235217	0.84	4.9	0	3.3	A/G	+42 182	3′-flanking region of CD120B
42	rs235214	1.00	11.6	14.3	27.3	C/T	+44 353	3′-flanking region of CD120B
43	rs426536	0.99	10.9	50[e]	6.7	G/A	+60 467	3′-flanking region of CD120B

[a]HW *P*-value is a measure of deviation from Hardy–Weinberg equilibrium calculated from the observed and predicted heterozygosity in the European-Caucasian founders.
[b]Minor allele frequency (MAF) is representative of all genotypes in the major ethnic groups (European-Caucasian (E), Afro-Caribbean (A) and Indo-Asian (I)).
[c]Common/rare alleles.
[d]Positions are given from the methionine start codon, with A of methionine equivalent to zero base-pair position.
[e]SNPs where allele frequency differs significantly ($P<0.001$) between ethnic groups.

23 kb downstream of *CD30*. *HVEM* is negatively transcribed, has eight exons and spans 7.5 kb of genomic DNA. *CD30* has 15 exons and spans 81 kb of genomic DNA, and *CD120B* consists of 10 exons and spans 42 kb. Both *CD30* and *CD120B* are positively transcribed.

In an attempt to screen all common haplotypes across *HVEM* for association, 23 SNPs were selected across the gene and its flanking regions. However, 11 of these 23 SNPs were not analysed and were removed from the analyses (Supplementary Table 1). Several factors necessitated the elimination of these 11 SNPs: these included the SNPs failing quality checks, lack of polymorphism or assay failure (as described in Materials and methods). In all, 12 SNPs were taken through analysis stages for *HVEM* (Table 1). We aimed for an SNP density of at least one SNP per 5 kb across *CD120B* and *CD30*. Coding SNPs were prioritised in selection. In this way, 54 SNPs were chosen for genotyping across *CD30* and 58 SNPs were chosen

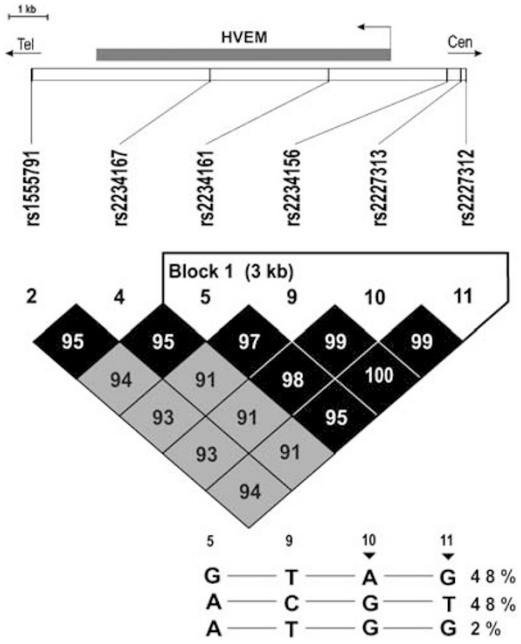

Figure 1 *HVEM* results of LD between SNP pairs and haplotype blocks in Caucasians for cohort 1. Values for D' are presented in each box. Black boxes indicate strong evidence for LD ($D' > 0.75$ with small confidence intervals (CI)) and grey boxes indicate intermediate LD ($D' > 0.75$ with large CI). Horizontal line above the LD diagram representative of chromosome with the location of SNPs indicated. Black rectangle above the chromosome line represents the location of gene on the chromosome. The orientation of the gene in centromeric (Cen) and telomeric (Tel) directions is shown. Below the LD diagram is the haplotype diversity within the haplotype block. Marker numbers are shown across the top. Haplotype tagging SNPs are highlighted with a triangular pointer. Haplotype frequencies are shown next to each haplotype.

across *CD120B*. From these, 69 SNPs were discarded across *CD30* and *CD120B* for similar reasons as listed for *HVEM* (Supplementary Table 1). In total, 25 SNPs were therefore taken through analysis stages for *CD30* and 18 SNPs were selected across *CD120B* (Table 2).

The SNPs that genotyped well (more than 75% individuals successfully genotyped) and fell within Hardy–Weinberg equilibrium were genotyped in an initial arbitrary cohort of 180 families, designated cohort 1 (supplementary Table 6). The allele frequencies for the individual SNPs showed significant differences between the different ethnic groups (Tables 1 and 2). For this study, all haplotype association analyses were performed on European-Caucasian samples.

HVEM LD analysis showed strong LD between SNPs 2 and 11, which encompasses the entire gene (Figure 1). The values of D' were close to 1.0 for most of the SNP pairs across *HVEM*, indicating very little recombination. Using the 95% confidence-bound SNP pair rule (as described in Methods and materials), one haplotype block was found, which spans 3 kb from the 5′-flanking region to intron 3 of *HVEM* (Figure 1). Using less stringent block algorithms such as solid spine or four gamete rule, one haplotype block was formed across the gene. The strength of LD within the haplotype block is reflected by the restricted haplotype diversity observed. Three haplotypes are observed in the block, with two high-frequency haplotypes of 48%. The three haplotypes in the block capture 98% of all diversity seen in our population.

The allele frequencies of individual SNPs within *CD30* and *CD120B* were found to vary significantly between the Afro-Caribbean and European-Caucasian or Indo-Asian populations (Table 2). In comparison, fewer SNP allele frequencies varied significantly between Indo-Asians and European-Caucasian populations. The differences seem to be clustered across *CD30* and *CD120B* and there is no consistent pattern across the entire locus, although Afro-Caribbeans have higher allele frequencies on either side of haplotype block 4. LD analysis across *CD30* and *CD120B* showed considerable recombination across the locus (Figure 2). The SNPs genotyped across this region span 241 kb and previous work in European ancestry samples has shown LD to extend up to 60 kb in such populations.[32] There were only small blocks of LD across *CD30* and *CD120B*, although *CD120B* had the largest LD block of 9 kb. Surprisingly, there is an area of strong LD from SNPs 22 to 26 spanning 8 kb between *CD30* and *CD120B*. This block of LD begins 7 kb after *CD30* and 9 kb before *CD120B*.

Figure 3 shows the allelic composition of the CD30 and CD120B haplotypes for the depicted block sizes in Figure 2. There are three haplotypes in block 1 with one high-frequency haplotype (71%). Haplotype block 2 falls across introns 8 and 10 of *CD30* and contains four haplotypes with one higher-frequency haplotype (52%). Haplotype block 3 comprises of three haplotypes and crosses introns 12 and 13 of *CD30*. Haplotype block 4 falls between *CD30* and *CD120B* and has four haplotypes with two haplotypes of similar frequency seen (44;45%), which span 8 kb. Haplotype block 5 runs from introns 1 to 8 of *CD120B* and six haplotypes with one higher-frequency haplotype (34%) are observed. The haplotypes within blocks 1–4 capture 98–100% of all diversity seen in our samples. Of the 94% haplotype diversity seen above 1% frequency is captured in haplotype block 5. Multilocus D' values between the blocks are <0.30 for haplotype blocks 1–4, indicating very weak LD between blocks and recombination (Figure 3). This can also be seen by the breaking of haplotype structure from each block to the next (Figure 3). There is some LD between blocks 4 and 5 with a multilocus D' of 0.53, but the splitting of haplotypes into greater haplotypes suggests recombination is relatively more recent.

Single marker TDT analysis using GENEHUNTER for the 12 SNPs across *HVEM* (Table 3) and the 43 SNPs across *CD30* and *CD120B* (Table 4) was conducted in the SLE families. Association analyses were performed on all of the

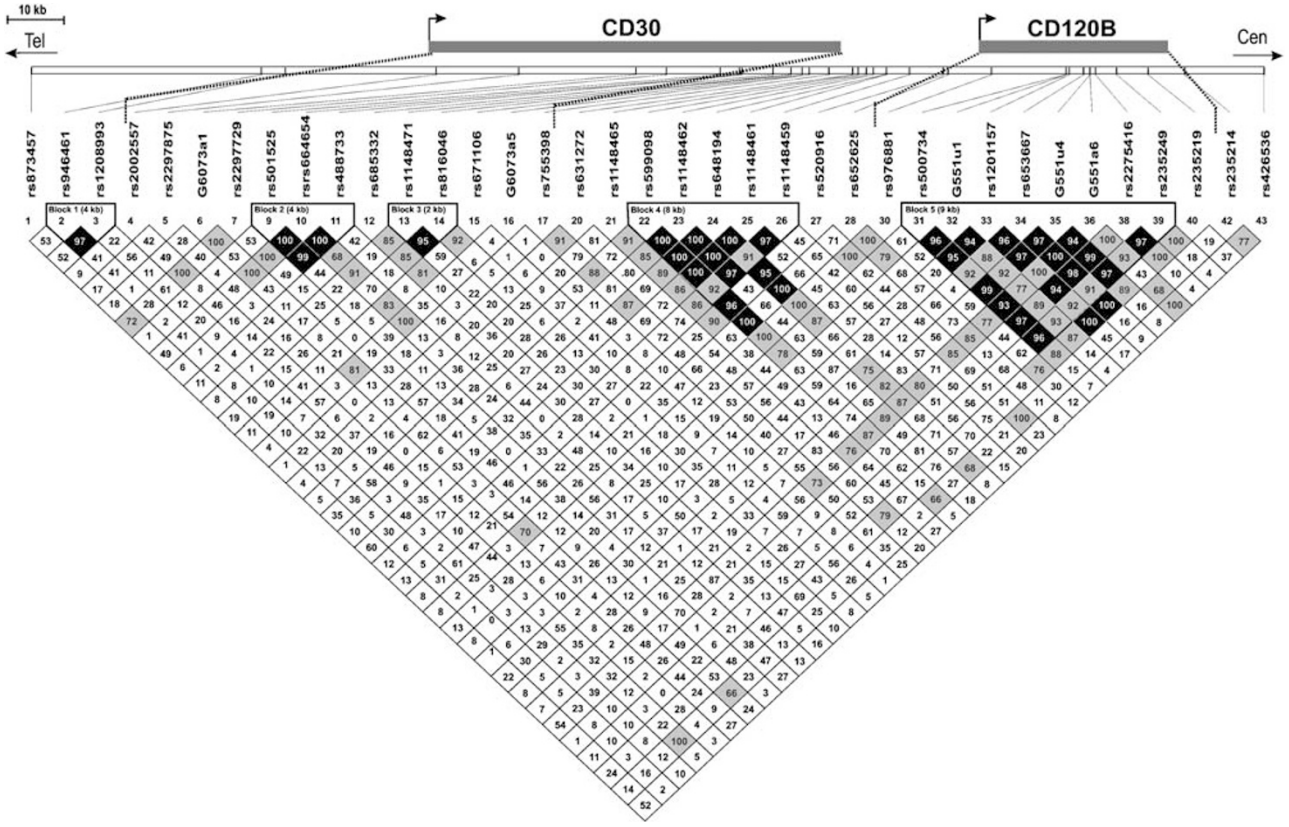

Figure 2 *CD30* and *CD120B* results of LD between SNP pairs and haplotype blocks in Caucasians for cohort 1. Values for *D*′ are presented in each box. Black boxes indicate strong evidence for LD ($D' > 0.75$ with small confidence intervals (CI)), grey boxes indicate intermediate LD ($D' > 0.75$ with large CI) and white boxes indicate inconclusive LD or evidence for recombination ($D' < 0.75$, $r^2 < 0.30$). Horizontal line above the LD diagram representative of chromosome with the location of SNPs indicated. Black rectangle above the chromosome line represents the location of gene on the chromosome. The orientation of the gene is illustrated in centromeric (Cen) and telomeric (Tel) directions.

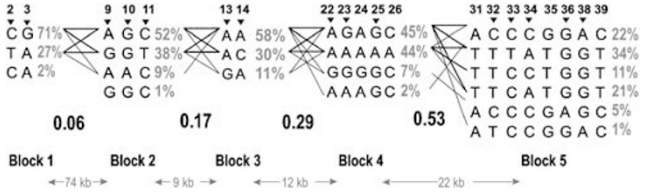

Figure 3 *CD30* and *CD120B* haplotype diversity within haplotype blocks. These were created based on the 95% confidence interval cutoff. Marker numbers are shown across the top. Haplotype tagging SNPs are highlighted with a triangular pointer. Haplotype frequencies are shown to the right of each haplotype. Lines show the most common crossings from one block to the next, with thicker lines showing more common crossings than thinner lines. Shown beneath the crossing lines is multilocus *D*′, which is a measure of the LD between two blocks. Interblock distances are indicated at the bottom of the diagram.

families in cohort 1. Selected SNPs were also analysed in cohort 2. Taking into account the haplotype data generated from analysing cohort 1, SNPs were selected for typing in the second cohort if they tagged a haplotype or showed any trend for association in the first cohort ($P < 0.10$). Several low-frequency SNPs (eg SNPs 35–37) were also typed in both cohorts if they were coding and altering the primary amino-acid sequence of the gene product. TDT analysis on cohort 2 only is shown in Supplementary Tables 2 and 3. Family-based association studies were conducted in European-Caucasian families separately because of racial differences in haplotype structure. No association was found in Afro-Caribbean or Indo-Asian families, although numbers of these families available were very low. Single-marker TDT provided no evidence for association with any of the SNPs screened with SLE (Tables 3 and 4). Furthermore, TRANSMIT TDT and discordant sib analysis using PDT confirmed the lack of association for the SNPs in *HVEM* and across *CD30* and *CD120B* with SLE (data not shown). Stratified TDT analyses were carried out on *HVEM*, *CD30* and *CD120B* SNPs in families with SLE patients selected for renal disease, thrombosis or anti-cardiolipin (aCL-G) antibody production as affected status and there was no evidence of association with these clinical subtypes when data were corrected for multiple

Table 3 Results of *HVEM* GENEHUNTER TDT in European-Caucasians

SNP	*Marker ID*	*T*[a]	*NT*[b]	*Overtransmitted allele*	χ^2	*P-value*[c]
1	rs2494617[d]	3	0	G	3.00	0.08
2	rs1555791[d]	69	52	G	2.39	0.12
3	rs2843396	4	3	G	0.14	0.71
4	rs2234167[d]	49	41	C	0.71	0.40
5	rs2234161	39	31	G	0.91	0.34
6	rs2234160	7	4	T	0.82	0.37
7	rs2234158	5	1	G	2.67	0.10
8	rs2494614	2	0	A	2.00	0.16
9	rs2234156[d]	91	78	T	1.00	0.32
10	rs2227313	30	31	A	0.91	0.34
11	rs2227312	39	33	G	0.50	0.48
12	rs2477684[d]	11	10	C	0.05	0.83

[a]Transmitted alleles.
[b]Nontransmitted alleles.
[c]*P*-value (1 df).
[d]Combined TDT results from cohorts 1 and 2.

testing (data not shown, Supplementary Tables 4 and 5). We selected renal disease as a distinct subset of more severe SLE for analysis. In addition, we chose to analyse patients with thrombosis and aCL-G antibodies, as previous studies had suggested an association of *CD120B* with vascular disease.[33] Haplotype TDT analysis using GENEHUNTER on haplotypes shown in Figures 1 and 2 for *HVEM*, *CD30* and *CD120B* did not reveal any association of haplotypes with SLE (data not shown).

Discussion

The tumour necrosis factor (TNF) cytokine superfamily plays a central role in immune regulation through a network of ligands and receptors. Linkage studies had identified the 1p36 region as a susceptibility locus for SLE,[5–9] and we selected the *TNFR* genes *HVEM*, *CD30* and *CD120B* as good candidates within this interval to examine. Our study aimed to characterise the haplotype structure of these genes and investigate the association of markers with SLE.

The majority of SNPs genotyped in this study were located in introns and untranslated region sequences. Two coding SNPs were successfully genotyped in *HVEM* (rs2234167, rs2234158). Two synonymous coding SNPs in *CD30* (G6073a1, G6073a5) were taken through analysis in European-Caucasians. Four coding sequence SNPs were genotyped in *CD120B* (G551u1, G551u4, G551a6, G551u5), the latter three are nonsynonymous (Table 2). We did not observe any association with SLE for *HVEM*, *CD30* and *CD120B* using family-based tests of association. We used clinical information on our cohorts based on previous reports for lupus nephritis, thrombosis and anti-cardiolipin IgG antibodies, and no evidence for association was found. There have been no association studies previously described for *HVEM* and only one study of *CD30* and type I diabetes, which found no association.[34]

CD120B has been extensively studied in SLE and other autoimmune diseases.[15,33,16,35–48] The M196R polymorphism in exon 6 of *CD120B* has been examined in European-Caucasians, Koreans, Japanese and African-Americans. There have been associations of this SNP with SLE, hypertension, hypercholesterolemia, Crohn's disease, bone mineral density, FCHL, RA, hyperandrogenism and polycystic ovary syndrome.[15,33,40-42,44,48,49] In SLE, there have been two positive associations and six negative associations of polymorphisms in *CD120B*.[15,16,35–39,50] However, no comprehensive haplotype-based studies have been performed. We found no association with SLE of M196R (G551u4) in either cohort 1 and 2 or in combined cohort results. The allele frequency of G551u4 (M196R) shown here was similar to that reported by groups looking at European-Caucasians and Oriental populations. The Afro-Caribbean allele frequency seen in this study is also comparable to a study on African-Americans and RA.[51]

We have characterised the LD pattern and haplotype structure for *HVEM*, *CD30* and *CD120B* in European-Caucasians. There is strong LD across the *HVEM* locus and little recombination. Less rigorous haplotype block algorithms produced one block across *HVEM*. *CD30* and *CD120B* markers span the locus enclosing both genes completely. Within *CD30*, there were two small areas of strong LD but a surprising lack of LD across the gene. Least information and LD was seen from introns 1 to 6 and also in the 3′-end of *CD30* (Figure 2). Between *CD30* and *CD120B*, there is a block of strong LD with no known genes within this area. The 8 kb sequence was queried against EST databases using the Gene2EST web-based program.[52] The EST sequences returned were consequently aligned and visualised with the Artemis program.[53] There was one cluster of overlapping human ESTs spanning 1.2 kb and no

Table 4 Results of *CD30* and *CD120B* GENEHUNTER TDT in European-Caucasians

SNP	*Marker ID*	*T*[a]	*NT*[b]	*Overtransmitted allele*	χ^2	*P-value*[c]
1	rs873457	30	28	G	0.07	0.79
2	rs946461[d]	35	35	—	0.00	1.00
3	rs1208993	28	28	—	0.00	1.00
4	rs2002557[d]	12	9	G	0.43	0.51
5	rs2297875	43	39	G	0.20	0.66
6	G6073a1	22	16	G	0.95	0.33
7	rs2297729	9	7	C	0.25	0.62
8	G6073a4	7	0	A	7.00	0.01
9	rs501525[d]	61	52	G	0.72	0.40
10	rs664654[d]	14	9	G	1.09	0.30
11	rs488733[d]	26	22	T	0.33	0.56
12	rs685332[d]	18	12	T	1.20	0.27
13	rs1148471[d]	64	52	A	1.24	0.27
14	rs816046	21	11	A	3.13	0.08
15	rs671106[d]	74	58	G	1.94	0.16
16	G6073a5	22	16	T	0.95	0.33
17	rs755398[d]	52	49	C	0.09	0.77
18	rs520311	0	0	—	—	—
19	rs544065	0	0	—	—	—
20	rs631272[d]	57	51	T	0.33	0.56
21	rs1148466[d]	52	51	C	0.01	0.92
22	rs599098	11	10	A	0.05	0.83
23	rs1148462[d]	64	57	G	0.41	0.52
24	rs648194[d]	13	13	—	0.00	1.00
25	rs1148461	40	38	G	0.05	0.82
26	rs1148459[d]	40	34	C	0.49	0.49
27	rs520916[d]	14	14	—	0.00	1.00
28	rs652625	12	9	A	0.43	0.51
29	rs595254	3	3	—	0.00	1.00
30	rs976881[d]	31	28	A	0.15	0.70
31	rs500734[d]	50	45	A	0.26	0.61
32	G551u1[d]	31	28	T	0.15	0.70
33	rs1201157[d]	58	57	C	0.01	0.92
34	rs653667[d]	51	49	C	0.04	0.84
35	G551u4[d]	36	33	T	0.13	0.72
36	G551a6[d]	13	10	G	0.39	0.53
37	G551u5[d]	9	3	C	3.00	0.08
38	rs2275416	31	26	G	0.44	0.51
39	rs235249[d]	27	21	T	0.75	0.39
40	rs235219[d]	28	20	A	1.33	0.25
41	rs235217[d]	9	5	G	1.14	0.29
42	rs235214[d]	18	15	T	0.27	0.60
43	rs426536	18	12	G	1.20	0.27

[a]Transmitted alleles.
[b]Nontransmitted alleles.
[c]*P*-value (1 df)
[d]Combined TDT results from cohorts 1 and 2.

specific patterns of coding sequence motifs/features were observed. *CD120B* had strong LD across a large part of the gene (introns 1 to 9) and SNP pairs in and around *CD120B* had high *D*′ values, suggesting recombination is relatively newer than for *CD30* (Figure 2*)*. All the common haplotypes within regions of strong LD across *HVEM*, *CD30* and *CD120B* can be identified by genotyping of haplotype tagging SNPs (htSNPs) as reported in the results section.

Haplotypes and LD patterns across *HVEM*, *CD30* and *CD120B* in European-Caucasian and Yoruban-African populations were compared using the International HapMap Project database (http://www.hapmap.org/). The LD across *HVEM*, *CD30* and *CD120B* was very similar to that seen in our cohorts for European-Caucasians. In all, 16 Sixteen SNPs utilised by the HapMap were also genotyped by us and on comparison, allele frequencies were very similar for European-Caucasians and Afro-Caribbeans. LD patterns in our few Afro-Caribbean and Indo-Asian families were examined (data not shown). The Afro-Caribbean LD across *HVEM*, *CD30* and *CD120B* was weak and there were no haplotype blocks seen, and the Indo-Asian LD pattern was similar to that of European-Caucasians across these genes. The HapMap Yoruban-African data across *HVEM* showed

strong LD from the 5′-flanking region to intron 3 and no LD from intron 3 and 7 kb into the 3′-flanking region, which is similar to that seen in our European-Caucasians. Across the *CD30* and *CD120B* locus, the Yoruban-Africans HapMap data were comparable to our Afro-Caribbean samples where there is weak LD.

Haplotype construction across *CD30* and *CD120B* has been previously described in Korean and Japanese.[54] The patterns of LD were similar in the two populations. This group showed strong LD across markers from introns 1 to 7 in Koreans and introns 1 to 6 in Japanese within *CD120B*. The LD pattern across *CD120B* was very similar to that seen in our European-Caucasian samples. The group reported low LD for *CD30* and did not define any haplotypes across *CD30*. Interestingly, this group also had an area of high LD falling in the 3′-flanking region to *CD30*. However, their map was not as dense in this area and full comparison could not be made. This group captured 98% of the haplotype diversity with only three haplotypes in the Koreans and two haplotypes in the Japanese across *CD120B*. In comparison, we had six haplotypes in our European-Caucasian population with three high-frequency haplotypes. The Koreans and Japanese had one high-frequency haplotype of 75 and 82%, respectively, across *CD120B*.

If there were SLE susceptibility polymorphism/s or haplotypes within *HVEM*, *CD30* and *CD120B*, we would have expected to be able to detect an association within areas of strong LD in this European-Caucasian family study. Power calculations were performed based on methods described elsewhere.[55] For regions with strong LD, we have reasonable power (60% cohort 1, 80% cohort 1 and 2) to detect association. For rarer alleles ($<10\%$), our power to rule out association is poor. It is difficult to completely exclude *CD30* because of the lack of LD in this region. We cannot exclude a rare variant occurring within *HVEM*, *CD30* and *CD120B* and association with SLE. However, we have captured more than 94% of the haplotype diversity in areas of high LD and all the common haplotypes and we have analysed all available missense mutations. These studies have been conducted in one of the largest single family collections of SLE available. Our inability to fully exclude a genetic contribution at this locus in SLE indicates the importance of collaboration and replication in complex trait analysis. Although we have shown no association with SLE, the wide range of diseases for which associations have been reported is such that *HVEM*, *CD30* and *CD120B* htSNPs that we have identified, will be useful for screening areas across these genes in genetic association studies in European-Caucasian populations.

Acknowledgements

The work was supported by a Senior Clinical Research Fellowship to TJV and SC from the Wellcome Trust. Lupus UK gave financial support for collection of samples and advertising this study. We would to thank the many physicians who provided patients for this study along with all the patients and their families for providing us with samples. Without their help this study would not have been possible.

References

1 Kotzin BL: Systemic lupus erythematosus. *Cell* 1996; **85**: 303–306.
2 Hochberg MC: Systemic lupus erythematosus. *Rheum Dis Clin N Am* 1990; **16**: 617–639.
3 Blatt NB, Glick GD: Anti-DNA autoantibodies and systemic lupus erythematosus. *Pharmacol Ther* 1999; **83**: 125–139.
4 Deapen D, Escalante A, Weinrib L *et al*: A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis Rheum* 1992; **35**: 311–318.
5 Gaffney PM, Kearns GM, Shark KB *et al*: A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families. *Proc Natl Acad Sci USA* 1998; **95**: 14875–14879.
6 Gaffney PM, Ortmann WA, Selby SA *et al*: 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* 2000; **66**: 547–556.
7 Shai R, Quismorio Jr FP, Li L *et al*: Genome-wide screen for systemic lupus erythematosus susceptibility genes in multiplex families. *Hum Mol Genet* 1999; **8**: 639–644.
8 Moser KL, Neas BR, Salmon JE *et al*: Genome scan of human systemic lupus erythematosus: evidence for linkage on chromosome 1q in African-American pedigrees. *Proc Natl Acad Sci USA* 1998; **95**: 14869–14874.
9 Lindqvist AK, Steinsson K, Johanneson B *et al*: A susceptibility locus for human systemic lupus erythematosus (hSLE1) on chromosome 2q. *J Autoimmun* 2000; **14**: 169–178.
10 Johanneson B, Lima G, von SJ, arcon-Segovia D, arcon-Riquelme ME: A major susceptibility locus for systemic lupus erythemathosus maps to chromosome 1q31. *Am J Hum Genet* 2002; **71**: 1060–1071.
11 Vyse TJ, Kotzin BL: Genetic susceptibility to systemic lupus erythematosus. *Annu Rev Immunol* 1998; **16**: 261–292.
12 Montgomery RI, Warner MS, Lum BJ, Spear PG: Herpes simplex virus-1 entry into cells mediated by a novel member of the TNF/NGF receptor family. *Cell* 1996; **87**: 427–436.
13 Kurts C, Carbone FR, Krummel MF, Koch KM, Miller JF, Heath WR: Signalling through CD30 protects against autoimmune diabetes mediated by CD8T cells. *Nature* 1999; **398**: 341–344.
14 Gabay C, Cakir N, Moral F *et al*: Circulating levels of tumor necrosis factor soluble receptors in systemic lupus erythematosus are significantly higher than in other rheumatic diseases and correlate with disease activity. *J Rheumatol* 1997; **24**: 303–308.
15 Komata T, Tsuchiya N, Matsushita M, Hagiwara K, Tokunaga K: Association of tumor necrosis factor receptor 2 (TNFR2) polymorphism with susceptibility to systemic lupus erythematosus. *Tissue Antigens* 1999; **53**: 527–533.
16 Tsuchiya N, Komata T, Matsushita M, Ohashi J, Tokunaga K: New single nucleotide polymorphisms in the coding region of human TNFR2: association with systemic lupus erythematosus. *Genes Immun* 2000; **1**: 501–503.
17 Siegmund T, Armitage N, Wicker LS, Peterson LB, Todd JA, Lyons PA: Analysis of the mouse CD30 gene: a candidate for the NOD mouse type 1 diabetes locus Idd9.2. *Diabetes* 2000; **49**: 1612–1616.
18 Chakrabarty S, Nagata M, Yasuda H *et al*: Critical roles of CD30/CD30L interactions in murine autoimmune diabetes. *Clin Exp Immunol* 2003; **133**: 318–325.
19 Tan EM, Cohen AS, Fries JF *et al*: The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheumatism* 1982; **25**: 1271–1277.
20 Hochberg MC: Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheumatism* 1997; **40**: 1725.

21 Roberton CA, Vyse TJ: The genetics of systemic lupus erythematosus. *Exp Nephrol* 2000; **8**: 194–202.
22 Loizou S, McCrea JD, Rudge AC, Reynolds R, Boyle CC, Harris EN: Measurement of anti-cardiolipin antibodies by an enzyme-linked immunosorbent assay (ELISA): standardization and quantitation of results. *Clin Exp Immunol* 1985; **62**: 738–745.
23 Sambrook J, Fritsch E, Maniatis T: *Analysis and Cloning of Eukaryotic Genomic DNA, Molecular Cloning: A Laboratory Manual*, Vol. **9**. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 1989, p 16.
24 Tang K, Opalsky D, Abel K *et al*: Single nucleotide polymorphism analyses by MALDI-TOF MS. *Int J Mass Spectrometry* 2003; **226**: 37–54.
25 Chadha S, Miller K, Farwell L *et al*: Haplotype structure of TNFRSF5-TNFSF5 (CD40–CD40L) and association analysis in systemic lupus erythematosus. *Eur J Hum Genet* 2005; **13**: 669–676.
26 O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63**: 259–266.
27 Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
28 Gabriel SB, Schaffner SF, Nguyen H *et al*: The structure of haplotype blocks in the human genome. *Science* 2002; **296**: 2225–2229.
29 Hedrick PW: Gametic disequilibrium measures: proceed with caution. *Genetics* 1987; **117**: 331–341.
30 Qin ZS, Niu T, Liu JS: Partition–ligation–expectation–maximization algorithm for haplotype inference with single-nucleotide polymorphisms. *Am J Hum Genet* 2002; **71**: 1242–1247.
31 Martin ER, Monks SA, Warren LL, Kaplan NL: A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am J Hum Genet* 2000; **67**: 146–154.
32 Reich DE, Cargill M, Bolk S *et al*: Linkage disequilibrium in the human genome. *Nature* 2001; **411**: 199–204.
33 Glenn CL, Wang WY, Benjafield AV, Morris BJ: Linkage and association of tumor necrosis factor receptor 2 locus with hypertension, hypercholesterolemia and plasma shed receptor. *Hum Mol Genet* 2000; **9**: 1943–1949.
34 Ahmed S, Ihara K, Bassuny WM *et al*: Association study between CD30 and CD30 ligand genes and type 1 diabetes in the Japanese population. *Genes Immun* 2002; **3**: 96–101.
35 Al Ansari AS, Ollier WE, Villarreal J, Ordi J, Teh LS, Hajeer AH: Tumor necrosis factor receptor II (TNFRII) exon 6 polymorphism in systemic lupus erythematosus. *Tissue Antigens* 2000; **55**: 97–99.
36 Sullivan KE: Genetics of systemic lupus erythematosus. Clinical implications. *Rheum Dis Clin N Am* 2000; **26**: 229–256, v–vi.
37 Takahashi M, Hashimoto H, Akizuki M *et al*: Lack of association between the Met196Arg polymorphism in the TNFR2 gene and autoimmune diseases accompanied by vasculitis including SLE in Japanese. *Tissue Antigens* 2001; **57**: 66–69.
38 Lee EB, Yoo JE, Lee YJ, Choi YJ, Park KS, Song YW: Tumor necrosis factor receptor 2 polymorphism in systemic lupus erythematosus: no association with disease. *Hum Immunol* 2001; **62**: 1148–1152.
39 Lee YH, Rho YH, Choi SJ, Ji JD, Song GG: The biallelic variable number of tandem repeats of the tumor necrosis factor receptor 2 promoter in systemic lupus erythematosus. *Rheumatol Int* 2003; **23**: 108–111.
40 Sashio H, Tamura K, Ito R *et al*: Polymorphisms of the TNF gene and the TNF receptor superfamily member 1B gene are associated with susceptibility to ulcerative colitis and Crohn's disease, respectively. *Immunogenetics* 2002; **53**: 1020–1027.
41 Geurts JM, Janssen RG, van Greevenbroek MM *et al*: Identification of TNFRSF1B as a novel modifier gene in familial combined hyperlipidemia. *Hum Mol Genet* 2000; **9**: 2067–2074.
42 Albagha OM, Tasker PN, McGuigan FE, Reid DM, Ralston SH: Linkage disequilibrium between polymorphisms in the human TNFRSF1B gene and their association with bone mass in perimenopausal women. *Hum Mol Genet* 2002; **11**: 2289–2295.
43 Spotila LD, Rodriguez H, Koch M *et al*: Association analysis of bone mineral density and single nucleotide polymorphisms in two candidate genes on chromosome 1p36. *Calcif Tissue Int* 2003; **73**: 140–146.
44 Tasker PN, Albagha OM, Masson CB, Reid DM, Ralston SH: Association between *TNFRSF1B* polymorphisms and bone mineral density, bone loss and fracture. *Osteoporos Int* 2004; **15**: 903–908, [E-pub April 8, 2004].
45 Dieude P, Petit E, Cailleau-Moindrault S *et al*: Association between tumor necrosis factor receptor II and familial, but not sporadic, rheumatoid arthritis: evidence for genetic heterogeneity. *Arthritis Rheum* 2002; **46**: 2039–2044.
46 Van Der Helm-Van Mil AH, Dieude P, Schonkeren JJ, Cornelis F, Huizinga TW: No association between tumour necrosis factor receptor type 2 gene polymorphism and rheumatoid arthritis severity: a comparison of the extremes of phenotypes. *Rheumatology (Oxford)* 2004; **43**: 1232–1234.
47 Glossop JR, Nixon NB, Dawes PT, Hassell AB, Mattey DL: No association of polymorphisms in the tumor necrosis factor receptor I and receptor II genes with disease severity in rheumatoid arthritis. *J Rheumatol* 2003; **30**: 1406–1409.
48 Peral B, San Millan JL, Castello R, Moghetti P, Escobar-Morreale HF: Comment: the methionine 196 arginine polymorphism in exon 6 of the TNF receptor 2 gene (TNFRSF1B) is associated with the polycystic ovary syndrome and hyperandrogenism. *J Clin Endocrinol Metab* 2002; **87**: 3977–3983.
49 Barton A, John S, Ollier WE, Silman A, Worthington J: Association between rheumatoid arthritis and polymorphism of tumor necrosis factor receptor II, but not tumor necrosis factor receptor I, in Caucasians. *Arthritis Rheum* 2001; **44**: 61–65.
50 Tsuchiya N, Kawasaki A, Tsao BP, Komata T, Grossman JM, Tokunaga K: Analysis of the association of HLA-DRB1, TNFalpha promoter and TNFR2 (TNFRSF1B) polymorphisms with SLE using transmission disequilibrium test. *Genes Immun* 2001; **2**: 317–322.
51 Bridges Jr SL, Jenq G, Moran M, Kuffner T, Whitworth WC, McNicholl J: Single-nucleotide polymorphisms in tumor necrosis factor receptor genes: definition of novel haplotypes and racial/ethnic differences. *Arthritis Rheum* 2002; **46**: 2045–2050.
52 Gemund C, Ramu C, tenberg-Greulich B, Gibson TJ: Gene2EST: a BLAST2 server for searching expressed sequence tag (EST) databases with eukaryotic gene-sized queries. *Nucl Acids Res* 2001; **29**: 1272–1277.
53 Rutherford K, Parkhill J, Crook J *et al*: Artemis: sequence visualization and annotation. *Bioinformatics* 2000; **16**: 944–945.
54 Akesaka T, Lee SG, Ohashi J *et al*: Comparative study of the haplotype structure and linkage disequilibrium of chromosome 1p36.2 region in the Korean and Japanese populations. *J Hum Genet* 2004; **49**: 603–609, [E-pub October 6, 2004].
55 Risch N, Merikangas K: The future of genetic studies of complex human diseases. *Science* 1996; **273**: 1516–1517.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)