

ARTICLE

# Haplotype structure of *TNFRSF5-TNFSF5* (*CD40-CD40L*) and association analysis in systemic lupus erythematosus

Sapna Chadha<sup>1</sup>, Katie Miller<sup>2</sup>, Lisa Farwell<sup>2</sup>, Liz B Lightstone<sup>1</sup>, Mark J Daly<sup>2</sup>, John D Rioux<sup>2</sup> and Timothy J Vyse<sup>\*1</sup>

<sup>1</sup>Faculty of Medicine, Imperial College London, Hammersmith Hospital, London, UK; <sup>2</sup>The Broad Institute of the Massachusetts Institute of Technology and Harvard, Cambridge, MA, USA

Systemic lupus erythematosus (SLE) is a prototypic autoimmune disease that is caused by genetic and environmental factors. The tumour necrosis factor (TNF) superfamily of genes play a central role in immune regulation and have been proposed to be involved in the development of SLE. *TNFRSF5* (*CD40*) falls on 20q11–13, a region linked with SLE in three independent genome-wide studies. *TNFSF5* (*CD40L*) falls on Xq26 and is the ligand for *TNFRSF5*. Seven single-nucleotide polymorphisms (SNPs) in *CD40* and eight SNPs in *CD40L* were looked at for linkage disequilibrium (LD) and haplotype analysis in European-Caucasians. Limited haplotype diversity was observed across *CD40* and *CD40L*, and >97% of the diversity was captured. We also examined the association of SNPs and haplotypes in *CD40* and *CD40L* with SLE in European-Caucasians. There was no evidence of association for *CD40* or *CD40L* in 408 European-Caucasian families with SLE from UK. Haplotype tagging SNPs (htSNPs) are made known, which will facilitate analysis for susceptibility in other autoimmune diseases and risk for infectious disease.

*European Journal of Human Genetics* (2005) 13, 669–676. doi:10.1038/sj.ejhg.5201367

Published online 12 January 2005

**Keywords:** SLE; *TNFRSF5*; *TNFSF5*; *CD40*; *CD40L*; SNP; linkage disequilibrium; haplotype; association

## Introduction

Systemic lupus erythematosus (SLE) is a complex autoimmune disease of unknown aetiology and is primarily characterised by the production of autoantibodies directed against cell surface and nuclear components. Autoantibodies contribute to end-organ damage by a variety of mechanisms, and formation of immune complexes can result in glomerulonephritis, arthritis, rashes, serositis, and vasculitis.

SLE is estimated to affect about one in 2000 people in some populations, and the clinical presentation is diverse and can sometimes be fatal.<sup>1</sup> There is a strong gender bias,

with a female:male ratio of about 9:1 seen between the ages of 15 and 50 years.<sup>1,2</sup> Ethnicity influences this disease, and African-Americans/African-Caribbeans, and Hispanics are three times as likely to contract the disease compared to European-Caucasians.<sup>3,1</sup>

The origin of SLE involves contributions both from the environment and from the genetic composition of the individual. Considerable evidence supports a genetic basis for susceptibility to SLE. Concordance rates in monozygotic twins range between 25 and 69%, while the rate is only 1–2% in dizygotic twins. Familial recurrence risks have been ascertained in SLE, and the recurrence risk of siblings of probands to the risk in the population as a whole ( $\lambda_s$ ) is 20 for SLE.<sup>4</sup>

Given our current understanding of the pathophysiological mechanisms at play in SLE, we were interested in examining a potential role for the CD40 receptor (*CD40*) and CD40 ligand (*CD40L*) in disease susceptibility. This

\*Correspondence: Dr TJ Vyse, Faculty of Medicine, Imperial College London, Hammersmith Hospital, Du Cane Road, London, UK. Tel: +44 20 83832339; fax: +44 20 87433109; E-mail: t.vyse@imperial.ac.uk  
Received 17 August 2004; revised 30 November 2004; accepted 1 December 2004

pair of interacting genes was selected based on genetic and functional data suggesting these as strong candidate genes.

*CD40* (also known as *TNFRSF5*, *p50*, and *Bp50*) gene falls in the region 20q11–13, which has been linked with SLE in three independent investigations in European-Caucasians, Mexican-Americans, and African-Americans.<sup>5–7</sup> *CD40L* (also known as *TNFSF5*, *CD154*, *TRAP*, and *gp39*) is located on Xq26.

The X chromosome was not examined in most of the SLE genome-wide scans. Only one study investigated the X chromosome and although no linkage was found this single study does not exclude a role for an X-linked gene effect in SLE.<sup>8</sup> The gender bias in SLE prevalence and the increased frequency of disease in patients with Klinefelter's syndrome would be consistent with some genetic effects arising from the X chromosome.<sup>9</sup>

The *CD40*–*CD40L* pair has been well defined and regulates multiple phases of the humoral and cellular immune response. *CD40* is expressed constitutively by B-lymphocytes, dendritic cells, endothelial cells and macrophages, whereas *CD40L* is upregulated on *CD4* + T cells, platelets, mast cells, and basophils upon activation. Engagement of *CD40* with *CD40L* induces the formation of memory B-lymphocytes, promotes immunoglobulin isotype switching, and is involved in thrombotic events.<sup>10,11</sup> Several groups have reported hyperexpression of *CD40L* by T cells and elevated soluble *CD40L* concentrations in human SLE.<sup>12–14</sup>

A rare inherited immune deficiency disorder, X-linked hyper IgM syndrome, is caused by mutations in *CD40L*. An autosomal recessive form of hyper IgM is caused by mutations in *CD40*.<sup>15</sup> The hyper IgM disorders are characterised by defective IgG antibody formation, lack of immunoglobulin isotype switching, and impaired B differentiation. *CD40L*-deficient and *CD40*-deficient mice generated by gene targeting show defective class-switching, but do not exhibit the spontaneous hyper IgM.<sup>16</sup>

The critical role of *CD40*–*CD40L* in B-cell activation is such that *CD40* has been considered as a therapeutic target in SLE. Administration of anti-*CD40* ligand antibody has been found to limit lupus nephritis.<sup>17,18</sup> Furthermore, anti-*CD40L* antibody in clinical trials in patients with SLE have shown promising results and that the drug has immunomodulatory actions.<sup>19,20</sup>

We have defined for the first time a high-resolution haplotype structure for *CD40* and *CD40L* and examined association of these haplotypes with SLE in European-Caucasians from the UK.

## Materials and methods

### Family collection

A large collection of SLE nuclear families has been collected from the UK with predominantly one affected offspring per family. 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 (MREC). The clinical manifestations of SLE are variable and diagnosis criteria have been established by the American College of Rheumatology (ACR).<sup>21,22</sup> 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.

The demographic details of the families studied are summarised in Table 2. The 623 families were randomly allocated into cohorts. Cohort 1 was used as an initial screen and cohort 2 was utilised for rare (low-frequency) Single-nucleotide polymorphisms (SNPs) and also to follow-up on SNPs showing hints of association from the initial screen. *CD40L* SNPs were genotyped on both cohorts because it is X-linked and there were less informative transmissions from only one parent.

DNA samples were prepared from 40 ml blood by phenol–chloroform extraction.<sup>23</sup> DNA concentration was obtained by pico-green quantification. DNA samples were stored at 4°C.

### 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 are 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. Our study focused on common SNPs (minor allele frequency (MAF) > 5%).

### Genotyping

The SNPs were genotyped by matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry (MS), using the Sequenom MassARRAY™ (Sequenom Inc., San Diego, CA, USA) methodology. This system uses samples in chip-based, high-density arrays, whereby hundreds of samples can be run in parallel. This system was used to evaluate more than one SNP (multiplexing) per sample. This method has been

previously described in detail.<sup>24</sup> In brief, PCR primer pairs were designed to amplify the region around the SNP, along with extension primers using the Sequenom SpectroDE-SIGNER™ software. The PCR primers (Metabion, Hamburg, Germany) were used to amplify 5 ng of genomic DNA in a multiplex reaction. Unincorporated nucleotides were removed from the PCR products, using shrimp alkaline phosphatase treatment. The homogeneous mass extend reaction was then performed, adding one or two bases to the extension primer, to include the alternate variant plus one base beyond the SNP. Adding a suspension of cation exchange beads desalted the extension reaction, and the resulting supernatant was spotted onto a 384 SpectroCHIP™. The extension products on the chip are then analysed on the Bruker Autoflex TOF-MS and the composition of the SNP can be determined from the mass. The resulting spectra were analysed using the Sequenom SpectroTYPER RT™ software.

### Statistical analysis

Hardy–Weinberg equilibrium (HWE) testing was carried out using the exact test courtesy of G Abecasis and J Wigginton (University of Michigan Center for Statistical Genetics). SNPs were chosen for further analysis if they did not deviate from HWE using the observed and expected heterozygosity results. Pedigree checking was performed using PedCheck 1.1.<sup>25</sup> All genotyping data were run through to check for Mendelian inconsistencies in pedigree data. Markers with >5 errors were removed from the analysis.

Two measures of linkage disequilibrium (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 Haploview 2.05 software (<http://www.broad.mit.edu/personal/jcbarret/haploview/index.php>). 95% confidence interval boundaries for pairs of SNPs were used to estimate recombination.<sup>26</sup> Multilocus  $D'$  between haplotype blocks is calculated by computing the  $2 \times 2 D'$  score of each allele at first locus with allele at second locus and then taking a weighted average of these values.<sup>27</sup>

The transmission disequilibrium test (TDT) method evaluates whether the frequency of transmission of alleles from heterozygous parents to their affected children deviates from 50%, the expected Mendelian frequency when there is no linkage. TDT analysis on the X chromosome was performed using ASPEX v.2.2.<sup>28</sup> ASPEX enables analysis of the X chromosome by using all affected siblings within a family and separates maternal and paternal transmissions. ASPEX calculates the probability of association occurring independently of linkage within families by permuting parental alleles while fixing the identity by descent (IBD) status of siblings. GENEHUNTER (version 3.0) was used for TDT analysis for nuclear simplex families.<sup>29</sup> TRANSMIT (version 2.5.4) was used for TDT analysis of

single-parent families.<sup>30</sup> TRANSMIT can deal with transmission of multilocus haplotypes, even if phase is unknown, and parental genotypes may be unknown. Many of the SLE families only have one parent and this program 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.<sup>31</sup> The most discordant siblings were selected based on being female and having negative antinuclear antibody (ANA) 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.

Haplotype analysis was conducted using Haploview 2.05 software. Haplotypes with <1% frequency were not included in the analysis. Haplotypes were estimated using an accelerated expectation-maximisation (EM)-based algorithm that 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 (MLE) and has highly accurate population frequency estimates for phased haplotypes.<sup>32</sup>

### Results

Using SNPs markers, the *CD40* and *CD40L* genes were investigated for association with SLE. In all, 14 SNPs were selected for *CD40* genotyping (as described in Materials and methods). Two assays failed (rs1535044 and rs1883832) and one marker (rs1801293) was monomorphic when genotyped in cohort 1. These three SNPs were removed from analysis. Mendelian error checking found SNP rs1569723 to have six pedigree errors and this marker was excluded from analysis. Three SNPs (rs7273698, rs11086998 and rs1004731) had rare allele frequency <0.1% in cohort 1 and were further genotyped on cohort 2. There was no change in allele frequency in increased families for these rare polymorphisms and these SNPs were excluded from LD and haplotype analysis. Seven SNPs across *CD40* genotyped well (greater than 75% individuals successfully genotyped) and fell within HWE (Table 1a, Figure 1a).

The *CD40* SNPs were genotyped on cohort 1 (Table 2). Allele frequencies for the individual SNPs showed significant differences between the different ethnic groups (Table 1a). For this study, all association analyses were performed on European-Caucasian samples for power due to the small numbers of other ethnicities in our cohorts (Table 2).

**Table 1** *CD40* SNP summary on (a) cohort 1 and (b) cohorts 1 and 2

SNP	Marker ID	MAF (%) <sup>a</sup>			Alleles <sup>b</sup>	Position <sup>c</sup> (bp)
		E	A	I		
<b>(a)</b>						
1	rs1009373	22.5	5.9	36.2 <sup>d</sup>	G/T	–4983
2	rs1800686	23.3	6.9 <sup>d</sup>	35.2 <sup>d</sup>	G/A	–581
3	rs752118	20.7	4.8	31.8	C/T	–246
4	rs1535045	22.8	6.0 <sup>d</sup>	35.6 <sup>d</sup>	C/T	+1115
5	rs3746821	11.1	15.3	25.4 <sup>d</sup>	T/G	+8127
6	rs3765459	22.8	<0.1 <sup>d</sup>	32.4 <sup>d</sup>	G/A	+10423
7	rs1883838	39.9	28.8	26.7 <sup>d</sup>	C/T	+14912
<b>(b)</b>						
1	rs975379	18.7	47.3 <sup>d</sup>	20.5	T/C	–6825
2	rs3092952	18.2	49.2 <sup>d</sup>	18.1	A/G	–3459
3	rs1126535	21.5	12.7	19.7	T/C	+146
4	rs715762	10.4	1.7 <sup>d</sup>	8.9	G/A	+3004
5	rs3092933	7.7	10.5	13.0	G/A	+6808
6	rs3092927	8.4	42.9 <sup>d</sup>	14.0	G/A	+8853
7	rs3092923	8.2	29.4 <sup>d</sup>	16.7 <sup>d</sup>	T/C	+10776
8	rs3092920	9.2	47.1 <sup>d</sup>	19.6 <sup>d</sup>	G/T	+13582

<sup>a</sup>Minor allele frequency (MAF) is representative of all genotypes in the major ethnic groups (European-Caucasian (E), Afro-Caribbean (A) and Indo-Asian (I)).

<sup>b</sup>Common/rare alleles.

<sup>c</sup>Positions are given from the methionine start codon; with A of methionine equivalent to zero base-pair position.

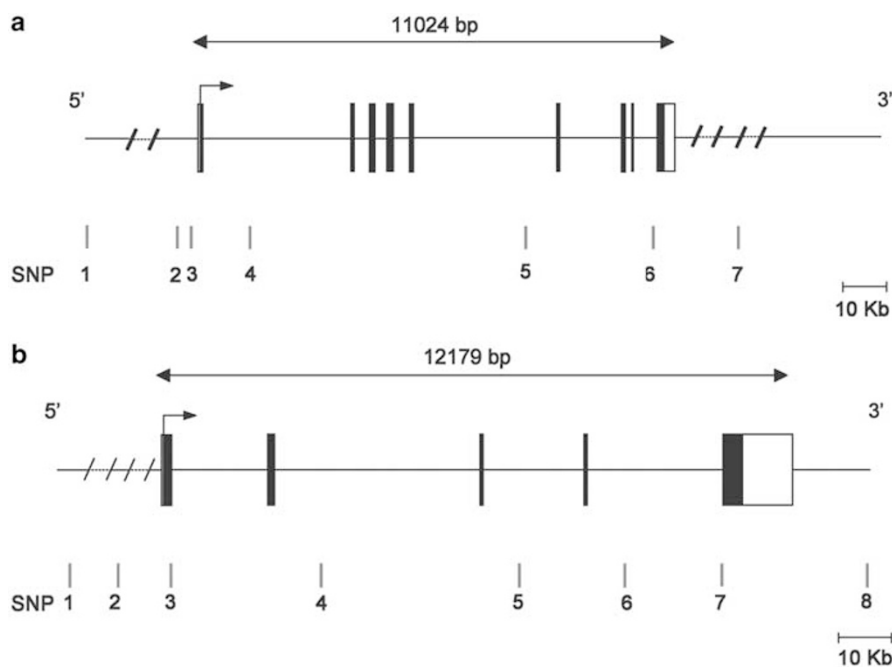
<sup>d</sup>SNPs where allele frequency differs significantly ( $P < 0.003$ ) from European-Caucasians.

*CD40* LD analysis showed strong LD between SNP 1 and SNP 5 and strong LD between SNP 6 and SNP 7 (Figure 2a). Values of  $D'$  were close to 1.0 for most of the SNP pairs.

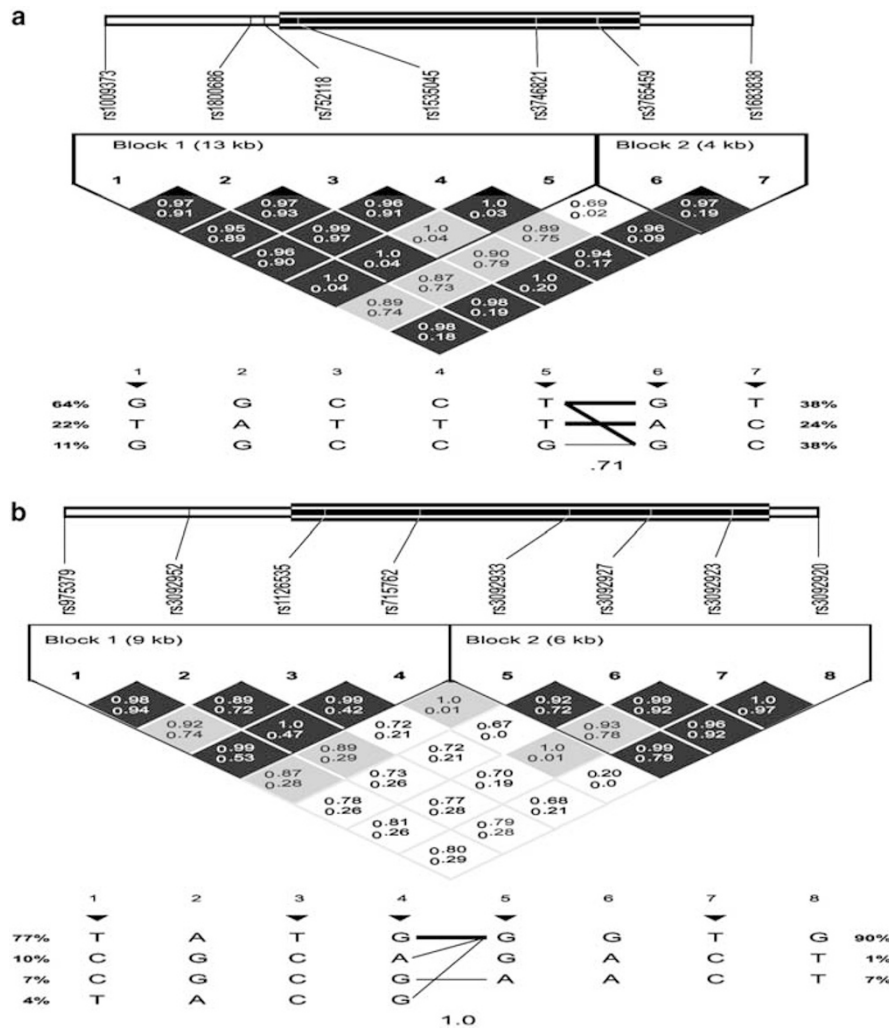
**Table 2** Pedigree breakdown and demographic features of the two SLE cohorts examined in this study

	Cohort 1	Cohort 2
<b>Family structure</b>		
Nuclear simplex	229	179
Single parent simplex	127	88
Discordant siblings <sup>a</sup>	110	108
<b>Ethnicity of families</b>		
European-Caucasian	312	227
Afro-Caribbean	15	20
Indo-Asian	21	12
Oriental	3	3
Mixed ethnicity	5	5
<b>Gender of proband</b>		
Male	33	15
Female	323	252
<b>Proband clinical characteristics</b>		
Lupus nephritis	146	63
Thrombosis	86	61

<sup>a</sup>Defined in Methods and materials.



**Figure 1** Structure of genes encoding *CD40* and *CD40L* with relative positions of the SNPs studied. (a) *CD40* gene with numbered SNPs corresponding to SNPs in Table 1a. (b) *CD40L* gene with numbered SNPs corresponding to SNPs in Table 1b. Orientation and transcription start is marked with an arrow. Exons are represented by boxes, and introns and intergenic regions by lines. Filled boxes denote translated regions; open boxes untranslated regions. Horizontal lines below gene depiction indicate SNP position within gene. Line breaks are equivalent to 3000 bp.



**Figure 2** Results of LD between SNP pairs and haplotype blocks. (a) *CD40* LD and haplotype diversity in Caucasians for cohort 1. (b) *CD40L* LD and haplotype diversity in Caucasians for cohorts 1 and 2. LD between SNP pairs was calculated using the LD coefficient  $D'$ . Values for  $D'$  (upper number) and  $r^2$  (lower number) are presented in each box. Black boxes indicate strong evidence for LD ( $D' > 0.75$  with small  $D'$  confidence intervals (CI)), grey boxes indicate intermediate LD ( $D' > 0.75$  with large  $D'$  CI), and white boxes indicate inconclusive LD or evidence for recombination ( $D' < 0.75$ ,  $r^2 < 0.30$ , with large  $D'$  CI). Horizontal line above the LD diagram representative of chromosome with location of SNPs indicated. Black shaded box on chromosome line represents location of gene. Shown below LD diagrams are haplotype blocks. These were created based on the 95% CI cutoff. Marker numbers are shown across the top. Haplotype tagging SNPs are highlighted with a triangular pointer. Haplotype frequencies are shown next to 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.

There is a 3 kb distance between haplotype block 1 and haplotype block 2. Haplotype block 1 spans the 5'-flanking region to intron 5 of *CD40*. Haplotype block 2 spans intron 8 to 3'-flanking region of *CD40* (Figure 2a). The strength of LD within the blocks is reflected by the restricted haplotype diversity observed. Three haplotypes are observed in block 1, with one high-frequency haplotype (64%). Three haplotypes are seen in block 2 and there are two common haplotypes (38%) created mainly from the common haplotype in block 1 (Figure 2a). The haplotypes in block 1 represent 97% of all haplotypes present and the

haplotypes in block 2 captured all of the diversity (100%) seen. A multilocus  $D'$  of 0.71 between the two haplotype blocks shows intermediate LD between the blocks and minimal recombination.

In total, 11 SNPs were genotyped across *CD40L* on cohorts 1 and 2. Two assays failed (rs3092945 and rs3092936) and one marker (rs3092922) was monomorphic. These three markers were removed from the study. The remaining eight SNPs spanning *CD40L* showed no evidence for deviation from HWE (Table 1b and Figure 1b). *CD40L* allele frequencies of individual SNPs

were found to vary significantly between Afro-Caribbean and European-Caucasian populations for most of the SNPs studied (Table 1b). In comparison, only two SNPs allele frequencies varied significantly between Indo-Asians and European-Caucasian populations (Table 1b).

*CD40L* LD analysis and *D'* values indicate two areas of strong LD (Figure 2b). There is evidence of disruption in LD between SNP 4 and SNP 5 ( $r^2 = 0.01$ ), and weak LD between SNP pairs across the two haplotype blocks ( $r^2 < 0.30$ ). There is a 4 kb distance between haplotype block 1 and haplotype block 2. Haplotype block 1 spans 5'-flanking region to intron 2 of *CD40L* and haplotype block 2 spans intron 3 to 3'-flanking region of *CD40L* (Figure 2b). There are four haplotypes in block 1 with one high-frequency haplotype (77%) and three haplotypes in block 2 with one high-frequency haplotype (90%). Three of the haplotypes from block 1 all cross to the common 90% haplotype in block 2 (Figure 2b). In all, 98% of the haplotypes seen above 1% frequency are captured in these two blocks for our population. Multilocus *D'* between the haplotype blocks is 1.0, indicating very strong LD between the two blocks. Pairwise LD analysis of markers across *CD40* and *CD40L* was carried out and did not show any evidence of LD between the genes (data not shown).

Single marker TDT analysis using GENEHUNTER of the seven SNPs across *CD40* showed no evidence of association with SLE (Table 3). Furthermore, TRANSMIT TDT and discordant sib-analysis using PDT confirmed the lack of association for the seven SNPs in *CD40* with SLE in this cohort (data not shown). Stratified TDT analysis was carried out on *CD40* SNPs with patients selected for renal disease or thrombosis as affected status and there was no evidence of association (data not shown). Haplotype TDT analysis using GENEHUNTER and TRANSMIT for *CD40* did not reveal any association of haplotypes (shown in Figure 2a) with SLE.

Single marker ASPEX TDT that separates maternal and paternal transmissions to affected child was performed for the eight SNPs across *CD40L*. There was no evidence of association for *CD40L* SNPs with SLE (Table 4). Stratifica-

**Table 3** Results of *CD40* TDT in Cohort 1

SNP	Marker ID	T <sup>a</sup>	NT <sup>b</sup>	Overtransmitted allele	$\chi^2$	P-value <sup>c</sup>
1	rs1009373	49	63	T	1.75	0.19
2	rs1800686	61	65	A	0.13	0.72
3	rs752118 <sup>d</sup>	24	18	C	0.86	0.35
4	rs1535045	49	47	C	0.04	0.83
5	rs3746821	29	30	T	0.02	0.89
6	rs3765459	54	54	G	0.00	1.00
7	rs1883838	87	73	T	1.23	0.27

<sup>a</sup>Transmitted alleles.

<sup>b</sup>Nontransmitted alleles.

<sup>c</sup>P-value (1 d.f.).

<sup>d</sup>Assay failure in two plates of Cohort 1.

**Table 4** Results of *CD40L* ASPEX TDT in Caucasians of cohorts 1 and 2

SNP	Marker ID	T <sup>a</sup>	NT <sup>b</sup>	$\chi^2$ <sup>b</sup>	P-value <sup>c</sup>
1	rs975379	46	47	0.01	0.92
2	rs3092952	32	35	0.13	0.71
3	rs1126535	26	35	1.33	0.24
4	rs715762	23	33	1.79	0.15
5	rs3092933	17	20	0.24	0.62
6	rs3092927	24	25	0.02	0.89
7	rs3092023	26	31	0.44	0.51
8	rs3092920	15	21	100	0.32

<sup>a</sup>Transmitted alleles.

<sup>b</sup>Nontransmitted alleles.

<sup>c</sup>P-value (1 df).

tion for renal disease or thrombosis was not performed due to the loss of power from the low number of heterozygotic transmissions from mothers.

## Discussion

*CD40*–*CD40L* are important tumour necrosis factor (TNF) superfamily members involved in B-cell interactions. Given the role that *CD40*–*CD40L* plays in humoral immune responses, they presented excellent candidate susceptibility genes for SLE. Our study aimed to characterise the haplotype structure of these genes and investigate association of markers with SLE.

The majority of SNPs genotyped in this study were located in introns and untranslated region (UTR) sequences. One coding SNP was successfully genotyped in *CD40* (rs11086998) but was extremely low frequency (<0.1%) in European-Caucasians and could not be used for analysis. Only one synonymous coding SNP in *CD40L* (rs1126535) was taken through analysis in European-Caucasians.

We did not observe any association with SLE for *CD40* or *CD40L* using family-based tests of association. We used clinical information on our cohorts for lupus nephritis and thrombosis and no evidence for association was found. There have been a limited number of association studies performed for *CD40* and *CD40L* in autoimmune disease. One case–control association study found a SNP in *CD40* to be positively associated with Graves' disease in Caucasians.<sup>33</sup> However, this was not replicated in a substantially larger study.<sup>34</sup> More recently, association of a 3'-UTR microsatellite and SLE in *CD40L* was reported in 80 patients from the Canary Islands in Spain.<sup>35</sup> No single microsatellite allele was associated with disease, but a group of longer alleles. The control frequency of this group of alleles was 0.32. This *CD40L* microsatellite lies between SNP 7 and SNP 8 and falls within haplotype block 2 (Figure 2b). It is unlikely that these alleles are associated with SLE in a UK population since we have defined *CD40L*

haplotypes flanking this marker and our study is considerably larger than the one reported.

We have shown the LD and haplotype structure for *CD40* and *CD40L* in European-Caucasians. *CD40* markers span the gene from –4983 (from ATG) to +14912 (from ATG). *CD40L* markers span the genomic area enclosing *CD40L* starting at –6825 (from ATG) to +13 582 (from ATG). Variation in the *CD40* and *CD40L* genes in European-Caucasians can be explained by the existence of three/four common haplotypes, which account for >97% frequency (Figure 2). It is clear from the pairwise marker LD and haplotypes in *CD40* that there are high levels of LD across the gene. This is reflected in the three common haplotypes in the two haplotype blocks in *CD40*. In the case of *CD40L*, there is a breakdown in LD between haplotype blocks but still substantial LD between adjacent blocks. All the common haplotypes in *CD40* and *CD40L* can be identified by genotyping of htSNPs as reported in the Results section. The extent of LD across *CD40* and *CD40L* suggests that any putative polymorphism would be in LD with one or more of the SNPs or haplotypes studied.

Haplotype construction across *CD40* has been previously described in Han-Chinese.<sup>36</sup> This group showed strong LD across *CD40* and five haplotypes. Three of the SNPs used by this group were the same as SNPs used in our study (rs1800686, rs752118 and rs3765459). Overall there are similarities in *CD40* haplotype structure between Han-Chinese and European-Caucasians. However, the Han-Chinese had two low-frequency haplotypes (their nomenclature: Hap 4 and Hap 5) not present in European-Caucasians. The Han-Chinese SNPs had similar frequencies to the Indo-Asian samples for *CD40*. There has been one study on haplotypes across *CD40L* in African populations,<sup>37</sup> and this group found similar LD for a section of the *CD40L* gene (rs975379 and rs715762) in Africans.

Allele frequencies of SNPs differed between the different ethnic groups as expected based on demographic histories of different populations. The differences were marked between Africans and European-Caucasians for both *CD40* and *CD40L*. A study on positive selection with malaria resistance in *CD40L* showed similar differences.<sup>37,38</sup> Two of the SNPs (rs975379 and rs715762) we studied were also studied in Sabeti *et al*'s work.<sup>37,38</sup> A more recent study that found no association of *CD40L* with tuberculosis confirmed the allele frequencies we observed for SNP rs1126535 in Afro-Caribbean's.<sup>39</sup>

If there were SLE susceptibility polymorphism/s within *CD40* or *CD40L*, we would have expected to be able to detect an association within this European-Caucasian family study. We cannot exclude a rare variant occurring within *CD40* or *CD40L* and association with SLE. However, we have captured greater than 97% of the haplotype diversity across these genes and all the common haplotypes. We exclude any common polymorphisms in *CD40* and *CD40L* from association with SLE in European-

Caucasians. Furthermore, the *CD40* and *CD40L* htSNPs that we have identified will be useful for genetic association studies on autoimmune and vascular diseases in European-Caucasian populations.

#### Acknowledgements

The Wellcome Trust supported this research (Senior Clinical Research Fellowship to TJV and SC) and Lupus UK gave financial support for collection of samples and advertising this study. We would like 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. We would also like to thank Mr Paul Spencer for coordination and collection of samples and Dr Deborah Cunninghame-Graham for genotyping assistance.

#### References

- 1 Kotzin BL: Systemic lupus erythematosus. *Cell* 1996; **85**: 303–306.
- 2 Hochberg MC: The epidemiology of systemic lupus erythematosus, 1997; 49–65.
- 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, 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.
- 6 Shai R, Quismori FP, Li L *et al*: Genome-wide screen for systemic lupus erythematosus susceptibility genes in multiplex families. *Hum Mol Genet* 1999; **8**: 639–644.
- 7 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.
- 8 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.
- 9 Jimenez-Balderas FJ, Tapia-Serrano R, Fonseca ME *et al*: High frequency of association of rheumatic/autoimmune diseases and untreated male hypogonadism with severe testicular dysfunction. *Arthritis Res* 2001; **3**: 362–367.
- 10 Grammer AC, Lipsky PE: CD154–CD40 interactions mediate differentiation to plasma cells in healthy individuals and persons with systemic lupus erythematosus. *Arthritis Rheum* 2002; **46**: 1417–1429.
- 11 Buchner K, Henn V, Grafe M, De Boer OJ, Becker AE, Kroczeck RA: CD40 ligand is selectively expressed on CD4+ T cells and platelets: implications for CD40–CD40L signalling in atherosclerosis. *J Pathol* 2003; **201**: 288–295.
- 12 Koshy M, Berger D, Crow MK: Increased expression of CD40 ligand on systemic lupus erythematosus lymphocytes. *J Clin Invest* 1996; **98**: 826–837.
- 13 Yellin MJ, D'Agati V, Parkinson G *et al*: Immunohistologic analysis of renal CD40 and CD40L expression in lupus nephritis and other glomerulonephritides. *Arthritis Rheum* 1997; **40**: 124–134.
- 14 Vakkalanka RK, Woo C, Kirou KA, Koshy M, Berger D, Crow MK: Elevated levels and functional capacity of soluble CD40 ligand in systemic lupus erythematosus sera. *Arthritis Rheum* 1999; **42**: 871–881.
- 15 Ferrari S, Giliani S, Insalaco A *et al*: Mutations of CD40 gene cause an autosomal recessive form of immunodeficiency with hyper IgM. *Proc Natl Acad Sci USA* 2001; **98**: 12614–12619.

- 16 Wu H, Shen N, Gu Y *et al*: Association between Bcl-2 gene polymorphism with systemic lupus erythematosus. *Zhonghua Yi Xue Za Zhi* 2002; **82**: 515–518.
- 17 Daikh DI, Finck BK, Linsley PS, Hollenbaugh D, Wofsy D: Long-term inhibition of murine lupus by brief simultaneous blockade of the B7/CD28 and CD40/gp39 costimulation pathways. *J Immunol* 1997; **159**: 3104–3108.
- 18 Kalled SL, Cutler AH, Datta SK, Thomas DW: Anti-CD40 ligand antibody treatment of SNF1 mice with established nephritis: preservation of kidney function. *J Immunol* 1998; **160**: 2158–2165.
- 19 Huang W, Sinha J, Newman J *et al*: The effect of anti-CD40 ligand antibody on B cells in human systemic lupus erythematosus. *Arthritis Rheum* 2002; **46**: 1554–1562.
- 20 Boumpas DT, Furie R, Manzi S *et al*: A short course of BG9588 (anti-CD40 ligand antibody) improves serologic activity and decreases hematuria in patients with proliferative lupus glomerulonephritis. *Arthritis Rheum* 2003; **48**: 719–727.
- 21 Tan EM, Cohen AS, Fries JF *et al*: The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1982; **25**: 1271–1277.
- 22 Hochberg MC: Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997; **40**: 1725.
- 23 Sambrook J, Fritsch E, Maniatis T: *Analysis and Cloning of Eukaryotic Genomic DNA, Molecular Cloning: a Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press, 1989, Vol. 9: pp 9.16–9.19.
- 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 O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63**: 259–266.
- 26 Gabriel SB, Schaffner SF, Nguyen H *et al*: The structure of haplotype blocks in the human genome. *Science* 2002; **296**: 2225–2229.
- 27 Hedrick PW: Gametic disequilibrium measures: proceed with caution. *Genetics* 1987; **117**: 331–341.
- 28 Hinds DA, Risch N: The ASPEX package: affected sib-pair exclusion mapping, 1996, <http://aspex.sourceforge.net/>.
- 29 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES: Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347–1363.
- 30 Clayton D: A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. *Am J Hum Genet* 1999; **65**: 1170–1177.
- 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 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.
- 33 Tomer Y, Concepcion E, Greenberg DA: A C/T Single-nucleotide polymorphism in the region of the CD40 gene is associated with Graves' Disease. *Thyroid* 2002; **12**: 1129–1135.
- 34 Heward JM, Simmonds MJ, Carr-Smith J, Foxall H, Franklyn JA, Gough SC: A single nucleotide polymorphism in the CD40 gene on chromosome 20q (GD-2) provides no evidence for susceptibility to Graves' disease in UK Caucasians. *Clin Endocrinol (Oxf)* 2004; **61**: 269–272.
- 35 Citores MJ, Rua-Figueroa I, Rodriguez-Gallego C *et al*: The dinucleotide repeat polymorphism in the 3'UTR of the CD154 gene has a functional role on protein expression and is associated with systemic lupus erythematosus. *Ann Rheum Dis* 2004; **63**: 310–317.
- 36 Zhou G, Zhai Y, Dong X *et al*: Variants in TNFRSF5 locus and association analysis with hepatitis B virus (HBV) infection. *Hum Mutat* 2004; **23**: 99–100.
- 37 Sabeti PC, Reich DE, Higgins JM *et al*: Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002; **419**: 832–837.
- 38 Sabeti P, Usen S, Farhadian S *et al*: CD40L association with protection from severe malaria. *Genes Immun* 2002; **3**: 286–291.
- 39 Campbell SJ, Sabeti P, Fielding K *et al*: Variants of the CD40 ligand gene are not associated with increased susceptibility to tuberculosis in West Africa. *Immunogenetics* 2003; **55**: 502–507.