



OPEN ACCESS



## EXTENDED REPORT

## Preferential association of a functional variant in complement receptor 2 with antibodies to double-stranded DNA

Jian Zhao,<sup>1</sup> Brendan M Giles,<sup>2</sup> Rhonda L Taylor,<sup>3</sup> Gabriel A Yette,<sup>2</sup> Kara M Lough,<sup>2</sup> Han Leng Ng,<sup>3</sup> Lawrence J Abraham,<sup>3</sup> Hui Wu,<sup>1</sup> Jennifer A Kelly,<sup>4</sup> Stuart B Glenn,<sup>4</sup> Adam J Adler,<sup>4</sup> Adrienne H Williams,<sup>5</sup> Mary E Comeau,<sup>5</sup> Julie T Ziegler,<sup>5</sup> Miranda Marion,<sup>5</sup> Marta E Alarcón-Riquelme<sup>4,6</sup> for the BIOLUPUS and GENLES Networks, Graciela S Alarcón,<sup>7</sup> Juan-Manuel Anaya,<sup>8</sup> Sang-Cheol Bae,<sup>9</sup> Dam Kim,<sup>9</sup> Hye-Soon Lee,<sup>9</sup> Lindsey A Criswell,<sup>10</sup> Barry I Freedman,<sup>11</sup> Gary S Gilkeson,<sup>12</sup> Joel M Guthridge,<sup>4</sup> Chaim O Jacob,<sup>13</sup> Judith A James,<sup>4,14,15</sup> Diane L Kamen,<sup>12</sup> Joan T Merrill,<sup>16</sup> Kathy Moser Sivils,<sup>4,14</sup> Timothy B Niewold,<sup>17</sup> Michelle A Petri,<sup>18</sup> Rosalind Ramsey-Goldman,<sup>19</sup> John D Reveille,<sup>20</sup> R Hal Scofield,<sup>4,15,21</sup> Anne M Stevens,<sup>22,23</sup> Luis M Vilá,<sup>24</sup> Timothy J Vyse,<sup>25</sup> Kenneth M Kaufman,<sup>26,27</sup> John B Harley,<sup>26,27</sup> Carl D Langefeld,<sup>5</sup> Patrick M Gaffney,<sup>4</sup> Elizabeth E Brown,<sup>7,28</sup> Jeffrey C Edberg,<sup>7</sup> Robert P Kimberly,<sup>7</sup> Daniela Ulgiati,<sup>3</sup> Betty P Tsao,<sup>1</sup> Susan A Boackle<sup>2,29</sup>

**Handling editor** Tore K Kvien

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/annrheumdis-2014-205584>).

For numbered affiliations see end of article.

**Correspondence to**

Dr Susan A Boackle, University of Colorado School of Medicine, 1775 Aurora Court, Room 3102 B, Mail Stop B115, Aurora, CO 80045, USA; [susan.boackle@ucdenver.edu](mailto:susan.boackle@ucdenver.edu)

JZ, BMG and RLT contributed equally.

Memberships for the BIOLUPUS and GENLES Networks are available in the Acknowledgements section.

Received 18 March 2014  
Revised 30 July 2014  
Accepted 2 August 2014  
Published Online First  
1 September 2014



CrossMark

**To cite:** Zhao J, Giles BM, Taylor RL, et al. *Ann Rheum Dis* 2016;**75**:242–252.

**ABSTRACT**

**Objectives** Systemic lupus erythematosus (SLE; OMIM 152700) is characterised by the production of antibodies to nuclear antigens. We previously identified variants in complement receptor 2 (*CR2/CD21*) that were associated with decreased risk of SLE. This study aimed to identify the causal variant for this association.

**Methods** Genotyped and imputed genetic variants spanning *CR2* were assessed for association with SLE in 15 750 case-control subjects from four ancestral groups. Allele-specific functional effects of associated variants were determined using quantitative real-time PCR, quantitative flow cytometry, electrophoretic mobility shift assay (EMSA) and chromatin immunoprecipitation (ChIP)-PCR.

**Results** The strongest association signal was detected at rs1876453 in intron 1 of *CR2* ( $p_{\text{meta}}=4.2 \times 10^{-4}$ , OR 0.85), specifically when subjects were stratified based on the presence of dsDNA autoantibodies (case-control  $p_{\text{meta}}=7.6 \times 10^{-7}$ , OR 0.71; case-only  $p_{\text{meta}}=1.9 \times 10^{-4}$ , OR 0.75). Although allele-specific effects on B cell *CR2* mRNA or protein levels were not identified, levels of complement receptor 1 (*CR1/CD35*) mRNA and protein were significantly higher on B cells of subjects harbouring the minor allele ( $p=0.0248$  and  $p=0.0006$ , respectively). The minor allele altered the formation of several DNA protein complexes by EMSA, including one containing CCCTC-binding factor (CTCF), an effect that was confirmed by ChIP-PCR.

**Conclusions** These data suggest that rs1876453 in *CR2* has long-range effects on gene regulation that decrease susceptibility to lupus. Since the minor allele at rs1876453 is preferentially associated with reduced risk of the highly specific dsDNA autoantibodies that are present in preclinical, active and severe lupus,

understanding its mechanisms will have important therapeutic implications.

**INTRODUCTION**

Systemic lupus erythematosus (SLE (OMIM 152700)) is a heterogeneous autoimmune disease with a strong genetic component modified by environmental exposures. The complement system has been linked to its pathogenesis since low serum complement levels were first demonstrated in patients with active disease.<sup>1 2</sup> Although complement activation resulting in tissue damage is an important feature of lupus, deficiencies of early classical complement pathway components are paradoxically strongly associated with lupus susceptibility. Deficiencies or altered function of complement receptors have also been associated with lupus in humans and murine models of disease.<sup>3–5</sup> These effects have been attributed to alterations in antigen clearance, antigen processing, tolerance induction and cell activation, but the exact mechanisms remain poorly understood.

SLE is characterised by the production of class-switched autoantibodies against nuclear antigens that have undergone affinity maturation, suggesting their generation in germinal centre reactions. Complement receptor 2 (*CR2/CD21*) is primarily expressed on mature B cells and follicular dendritic cells, two major components of germinal centres. We first showed *Cr2*, which encodes both CR2 and complement receptor 1 (*CR1/CD35*) in the mouse, to be a candidate gene for lupus susceptibility in the NZM2410 model of lupus based on structural and functional alterations in its gene

products.<sup>6</sup> We subsequently demonstrated strong association of a common three single-nucleotide polymorphism (SNP) *CR2* haplotype (rs3813946 in the 5'UTR, rs1048971 and rs17615 in exon 10) with increased risk of lupus susceptibility ( $p=1.0\times 10^{-5}$ ) in Caucasian and Chinese lupus simplex families with a 1.54-fold increased risk for disease development.<sup>7</sup> We confirmed this in a case-control analysis of an independent European-derived population ( $p=2.3\times 10^{-2}$ , OR 1.1 (95% CI 1.02 to 1.2))<sup>8</sup> and also showed that a haplotype formed by the minor alleles of three SNPs in exons 10 and 11 (rs1048971, rs17615, rs4308977) was associated with decreased risk of lupus ( $p=1.6\times 10^{-2}$ , OR 0.90 (95% CI 0.82 to 0.98)).<sup>8</sup>

In this study, we fine-mapped the region spanning *CR2* in 15 750 subjects from four ancestral groups to identify potential causal variant(s) for these associations with lupus. Additionally, we explored the association of *CR2* polymorphisms with clinical manifestations of lupus in order to generate new hypotheses regarding how *CR2* contributes to disease development.

## METHODS

### Subjects

DNA from individuals recruited from multiple sites was processed at the Oklahoma Medical Research Foundation (OMRF; Large Lupus Association Study 2) with institutional review board approval. All patients with SLE met the 1997 American College of Rheumatology revised classification criteria.<sup>9</sup> Clinical data were collected by chart review or testing in the OMRF Clinical Immunology laboratory. Samples for functional analyses were from healthy non-smoking 18-year-old to 60-year-old adults without family history of autoimmune disease at the University of Colorado School of Medicine.

### Genotyping

Genotyping was performed on the OMRF Illumina iSelect platform.<sup>10–11</sup> Subjects with missing genotype rate >10%, shared identical by descent >0.4 or gender mismatch were removed. Global ancestry was estimated based on the genotype of ancestry informative markers (AIMs), using principal components analysis<sup>12</sup> and ADMIXMAP<sup>13</sup> as described<sup>14</sup> and genetic outliers removed. Final clean data were from European Americans (EA), African Americans (AA; 7.5% Gullahs), Asians (AS; 74.6% Koreans, 16.1% Chinese, 9.3% Japanese and Singaporeans) and Hispanics (HS) enriched for Amerindian-European admixture. 2001 EA cases and 2153 EA controls were previously analysed.<sup>8</sup> Subjects for functional studies were genotyped using a Taqman SNP Genotyping Assay.

### Imputation

SNP and insertion-deletion (INDEL) genotypes of 379 Europeans, 246 Africans, 286 Asians and 181 Americans from the 1000 Genomes Project (V3, Phase 1 integrated data, March 2012 release) were references in imputation for EA, AA, AS and HS subjects, respectively. Imputation was performed using IMPUTE 2.1.2<sup>15</sup>; genotypes with information scores >0.9 and minor allele frequency (MAF)>0.01 were further analysed.

### Association tests

SNPs and INDELs showing biased Hardy–Weinberg equilibrium ( $p<0.01$  in controls,  $p<0.0001$  in cases), missing genotype rate >5% or different missing genotype rates between cases and controls (rate >2% and  $p<0.05$ ) were excluded. Variants were assessed for association with SLE under a logistic regression model, and haplotypic and haplotype-based conditional association tests were performed, adjusting for gender and the first

three principal components estimated using AIMs. For trans-ancestral meta-analysis, a fixed effect model was applied if Cochran's Q statistic showed no evidence of genetic heterogeneity among ORs ( $p>0.05$ ); otherwise, a random effect model was used. Analyses were performed using PLINK V1.07.<sup>16</sup>

### Sample preparation

Peripheral blood mononuclear cells (PBMC) were isolated over Ficoll–Paque (Sigma–Aldrich). DNA was purified using the QIAamp DNA Mini Kit (QIAGEN). B cells were purified using the Easy Sep Human B Cell Enrichment Kit (StemCell Technologies). RNA was processed using the RNeasy Plus Mini Kit (QIAGEN). Quality check and quantification of RNA was performed using the Agilent 2100 Bioanalyzer. RNA and DNA were stored at  $-70^{\circ}\text{C}$ . Epstein–Barr virus (EBV)-transformed B cell lines were generated by incubating PBMCs with supernatant from cell line GM7404A and cyclosporine A.

### qPCR and quantitative flow cytometry

qPCR of primary B cell transcripts was performed using cDNA transcribed using random primers and MultiScribe reverse transcriptase (Applied Biosystems), customised *CR2* primers (5'-CGAGAAGTATATTCTGTTGATCCATACAA-3', 5'-CTAATC AATATTCCGCTGAATTCCA-3') and probe (6FAM-AACTGG TGTGTGCCTCA-MGBNFQ), Taqman assays for *CR1* (4331182) and  $\beta$ -actin (4352935E), and the Applied Biosystems 7500 Real-Time PCR System. Relative expression levels of *CR2* and *CR1*, normalised to  $\beta$ -actin, were calculated using the comparative  $C_T$  method.<sup>17</sup>

Quantitative flow cytometry was performed using Quantum Simply Cellular microbeads (Bangs Laboratories). Purified B cells were incubated with ethidium monoazide to exclude dead cells and mouse IgG to block Fc receptors and stained with biotinylated anti-human CD19 (clone HD37) followed by AF405-labelled streptavidin (Molecular Devices) and either AF488-labelled anti-human *CR1* (clone 6B1.H12) or AF488-labelled anti-human *CR2* (clone THB5, Harlan). Data acquired on fixed cells within 24 h using the LSR II and CellQuest software (BD Biosciences) were analysed with FlowJo (Tree Star).

### Electrophoretic mobility shift assay

39-mer oligonucleotides (5'-GGGCCAAAAGCGAGACGGT[G/A]GGGGCAGTGCTCGACG-3', 5'-CGTCGAGCACTGCCCC [C/T]ACCGTCTCGCTTTTGGCCC-3') were synthesised, HPLC-purified and biotin-labelled; 25 fmol labelled oligonucleotides were incubated with 2–4  $\mu\text{g}$  nuclear extracts (NE) or 50 mM potassium chloride (for unbound controls) in 4% Ficoll, 20 mM N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid (HEPES), 1 mM EDTA, 0.5 mM dithiothreitol and 1  $\mu\text{g}$  poly dI:dC for 30 min on ice. For competition and blocking experiments, NE were pre-incubated with unlabelled oligonucleotide or anti-CCCTC-binding factor (CTCF) (sc-15914; Santa Cruz Biotechnology) for 10 or 30 min on ice. Reactions were electrophoresed on 6% DNA retardation gels (Life Technologies), transferred and cross-linked using ultraviolet light to nylon membranes and visualised using the Chemiluminescent Nucleic Acid Detection Module (Thermo Fisher Scientific).

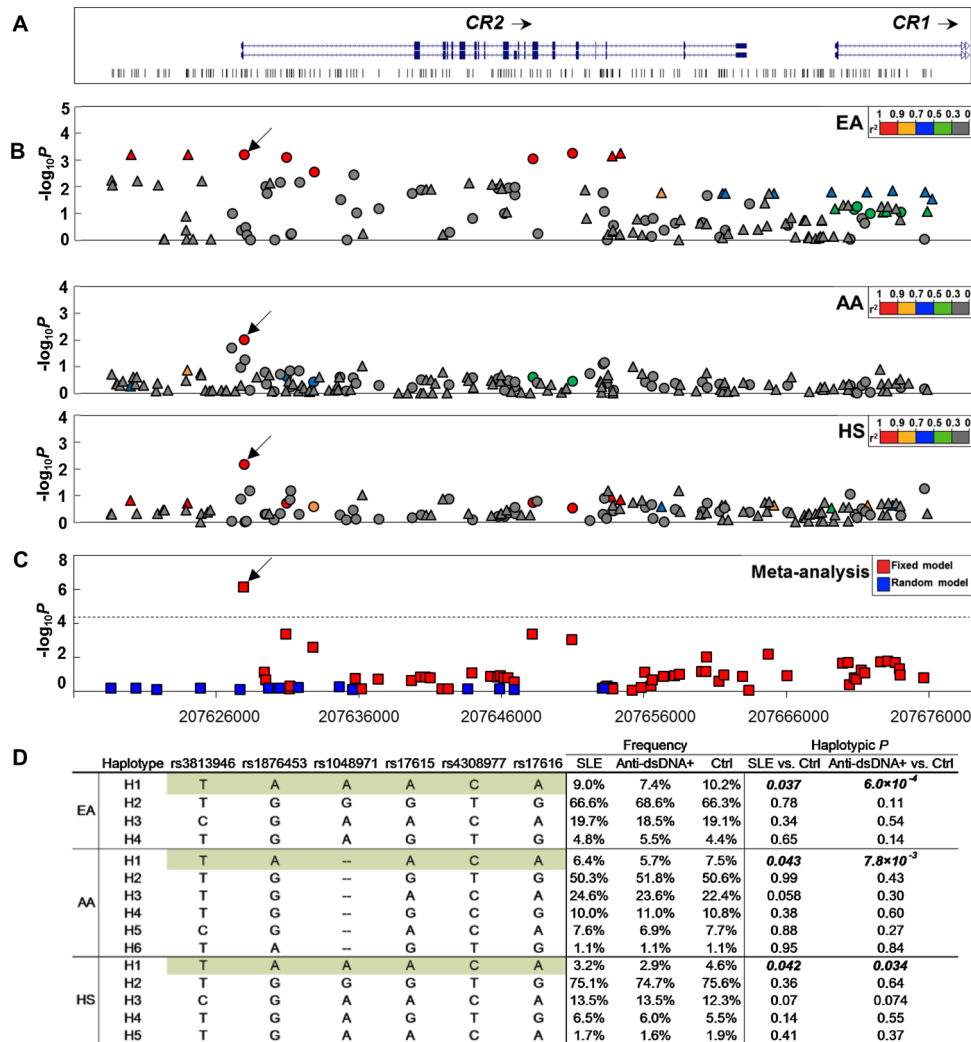
### Chromatin immunoprecipitation-PCR

Chromatin immunoprecipitation (ChIP) was performed using the Upstate Biotechnology ChIP assay kit with minor modifications.<sup>18</sup>  $1.1\times 10^8$  cells were incubated with 1% formaldehyde for 15 min and 0.125 M glycine added for 5 min. Washed cells

**Table 1** Association of rs1876453 with 14 SLE subphenotypes

Subphenotypes	N			MAF			Pos vs Control		Pos vs Neg	
	Group	Pos	Neg	Pos	Neg	Control	p Value	OR (95% CI)	p Value	OR (95% CI)
Malar rash	EA	2108	1171	0.088	0.087	0.099	0.11	0.89 (0.78 to 1.03)	0.92	1.01 (0.85 to 1.21)
	AA	685	714	0.08	0.066	0.086	0.24	0.87 (0.69 to 1.10)	0.15	1.24 (0.92 to 1.65)
	HS	836	572	0.03	0.032	0.051	<b><i>3.0×10<sup>-3</sup></i></b>	0.57 (0.39 to 0.83)	0.62	0.89 (0.57 to 1.40)
	Meta						<b><i>5.7×10<sup>-3</sup></i></b>	0.85	0.52	1.05
Discoid rash	EA	573	2486	0.088	0.088	0.099	0.38	0.91 (0.73 to 1.13)	0.98	1.00 (0.80 to 1.25)
	AA	485	915	0.088	0.065	0.086	0.81	0.97 (0.75 to 1.25)	<b><i>0.029</i></b>	1.39 (1.04 to 1.86)
	HS	180	1227	0.026	0.031	0.051	<b><i>0.032</i></b>	0.46 (0.22 to 0.93)	0.43	0.75 (0.37 to 1.53)
	Meta						0.2	0.9	0.26	1.1
Photosensitivity	EA	2345	1142	0.091	0.08	0.099	0.29	0.93 (0.82 to 1.06)	0.16	1.14 (0.95 to 1.36)
	AA	685	712	0.078	0.069	0.086	0.14	0.84 (0.66 to 1.06)	0.39	1.14 (0.85 to 1.52)
	HS	840	564	0.037	0.022	0.051	<b><i>0.047</i></b>	0.70 (0.49 to 0.99)	<b><i>0.05</i></b>	1.64 (1.00 to 2.67)
	Meta						<b><i>0.029</i></b>	0.89	<b><i>0.031</i></b>	1.17
Oral ulcers	EA	1574	1642	0.09	0.09	0.099	0.22	0.91 (0.78 to 1.06)	0.9	0.99 (0.83 to 1.17)
	AA	495	903	0.059	0.081	0.086	<b><i>1.7×10<sup>-3</sup></i></b>	0.63 (0.47 to 0.84)	0.06	0.74 (0.54 to 1.01)
	HS	562	844	0.036	0.027	0.051	0.094	0.71 (0.48 to 1.06)	0.2	1.33 (0.85 to 2.08)
	Meta						<b><i>3.5×10<sup>-3</sup></i></b>	0.83	0.59	0.96
Arthritis	EA	2987	619	0.088	0.091	0.099	0.075	0.89 (0.79 to 1.01)	0.76	0.97 (0.78 to 1.20)
	AA	1166	232	0.076	0.059	0.086	<b><i>0.048</i></b>	0.82 (0.67 to 1.01)	0.19	1.33 (0.87 to 2.02)
	HS	1007	402	0.031	0.031	0.051	<b><i>3.7×10<sup>-3</sup></i></b>	0.60 (0.42 to 0.85)	0.96	1.01 (0.62 to 1.66)
	Meta						<b><i>9.9×10<sup>-4</sup></i></b>	0.84	0.76	1.03
Serositis	EA	1341	2060	0.084	0.092	0.099	<b><i>0.042</i></b>	0.85 (0.72 to 0.99)	0.26	0.91 (0.76 to 1.08)
	AA	653	742	0.07	0.076	0.086	<b><i>0.025</i></b>	0.75 (0.59 to 0.96)	0.55	0.92 (0.69 to 1.22)
	HS	384	877	0.029	0.03	0.051	<b><i>0.022</i></b>	0.56 (0.34 to 0.92)	0.94	0.98 (0.58 to 1.64)
	Meta						<b><i>6.0×10<sup>-4</sup></i></b>	0.8	0.21	0.91
Renal disorder	EA	1102	2143	0.079	0.098	0.099	<b><i>0.016</i></b>	0.80 (0.67 to 0.96)	<b><i>0.018</i></b>	0.80 (0.67 to 0.96)
	AA	697	701	0.065	0.081	0.086	<b><i>6.1×10<sup>-3</sup></i></b>	0.71 (0.55 to 0.91)	0.092	0.78 (0.58 to 1.04)
	HS	644	746	0.032	0.028	0.051	<b><i>0.029</i></b>	0.64 (0.43 to 0.96)	0.41	1.21 (0.77 to 1.91)
	Meta						<b><i>4.4×10<sup>-5</sup></i></b>	0.75	<b><i>0.013</i></b>	0.83
Neurological disorder	EA	575	2524	0.099	0.092	0.099	0.89	1.02 (0.82 to 1.26)	0.51	1.07 (0.87 to 1.33)
	AA	372	1025	0.088	0.068	0.086	0.75	0.95 (0.72 to 1.27)	0.069	1.34 (0.98 to 1.82)
	HS	194	1213	0.024	0.032	0.051	<b><i>0.036</i></b>	0.47 (0.23 to 0.95)	0.42	0.75 (0.37 to 1.52)
	Meta						0.58	0.95	0.18	1.12
Haematological disorder	EA	2173	1081	0.085	0.089	0.099	<b><i>0.038</i></b>	0.87 (0.76 to 0.99)	0.67	0.96 (0.80 to 1.15)
	AA	1011	384	0.07	0.079	0.086	<b><i>0.012</i></b>	0.76 (0.62 to 0.94)	0.45	0.89 (0.65 to 1.22)
	HS	811	459	0.031	0.028	0.051	<b><i>9.6×10<sup>-3</sup></i></b>	0.61 (0.42 to 0.89)	0.79	1.07 (0.65 to 1.77)
	Meta						<b><i>1.9×10<sup>-4</sup></i></b>	0.81	0.53	0.95
Anti-dsDNA	EA	1179	1500	0.071	0.104	0.099	<b><i>6.6×10<sup>-4</sup></i></b>	0.73 (0.61 to 0.88)	<b><i>1.2×10<sup>-4</sup></i></b>	0.68 (0.56 to 0.83)
	AA	675	746	0.067	0.083	0.086	<b><i>9.7×10<sup>-3</sup></i></b>	0.72 (0.56 to 0.92)	0.097	0.79 (0.59 to 1.04)
	HS	726	564	0.03	0.028	0.051	<b><i>7.1×10<sup>-3</sup></i></b>	0.58 (0.39 to 0.86)	0.61	1.13 (0.70 to 1.84)
	Meta						<b><i>7.6×10<sup>-7</sup></i></b>	0.71	<b><i>1.9×10<sup>-4</sup></i></b>	0.75
Anti-Sm	EA	225	2064	0.109	0.093	0.099	0.37	1.15 (0.85 to 1.56)	0.24	1.20 (0.88 to 1.64)
	AA	430	755	0.087	0.076	0.086	0.78	0.96 (0.74 to 1.26)	0.41	1.14 (0.84 to 1.54)
	HS	294	905	0.03	0.029	0.051	0.069	0.60 (0.35 to 1.04)	0.71	1.11 (0.63 to 1.98)
	Meta						0.78	0.97	0.15	1.16
Anti-RNP	EA	255	1382	0.104	0.095	0.099	0.46	1.12 (0.83 to 1.51)	0.49	1.12 (0.82 to 1.53)
	AA	378	253	0.078	0.077	0.086	0.34	0.87 (0.65 to 1.16)	0.91	1.03 (0.67 to 1.57)
	HS	171	443	0.024	0.04	0.051	<b><i>0.033</i></b>	0.44 (0.21 to 0.94)	0.16	0.57 (0.26 to 1.26)
	Meta						0.45	0.93	0.86	1.02
Anti-SSA/Ro	EA	450	1331	0.086	0.093	0.099	0.28	0.87 (0.68 to 1.12)	0.51	0.92 (0.70 to 1.19)
	AA	292	542	0.062	0.08	0.086	<b><i>0.023</i></b>	0.66 (0.46 to 0.94)	0.21	0.77 (0.51 to 1.15)
	HS	238	491	0.028	0.039	0.051	<b><i>0.027</i></b>	0.51 (0.28 to 0.93)	0.28	0.70 (0.36 to 1.34)
	Meta						<b><i>5.4×10<sup>-3</sup></i></b>	0.76	0.13	0.85
Anti-SSB/La	EA	148	1506	0.091	0.095	0.099	0.76	0.94 (0.62 to 1.41)	0.87	0.97 (0.64 to 1.46)
	AA	81	692	0.037	0.08	0.086	<b><i>0.025</i></b>	0.39 (0.17 to 0.89)	0.057	0.44 (0.19 to 1.03)
	HS	71	558	0.042	0.033	0.051	0.63	0.81 (0.34 to 1.92)	0.53	1.33 (0.54 to 3.30)
	Meta						0.18	0.79	0.5	0.89

N, number in sample; MAF, minor allele frequency; Pos, subphenotype positive; Neg, subphenotype negative; p value, p value and OR were calculated using a logistic regression model adjusted for gender and global ancestry; EA, European-American; AA, African-American; HS, Hispanic; Meta, Meta-analysis; SLE, systemic lupus erythematosus. p Values ≤0.05 are marked with bold italics; significant p values after Bonferroni correction were ≤3.6×10<sup>-3</sup>.



**Figure 1** Association of single-nucleotide polymorphisms (SNPs) in the *CR2* region with dsDNA autoantibodies. (A) The genomic structure of the *CR2* region and positions of genetic variants are indicated. (B) The allelic p value ( $-\log_{10}P$  value) of each genetic variant with dsDNA autoantibodies is plotted against its position as a circle (genotyped) or a triangle (imputed) for European American (EA), African American (AA) and Hispanic (HS), respectively. Genetic variants are highlighted using different colours according to their strength of linkage disequilibrium (LD) ( $r^2$ ) with rs1876453. An arrow is used to indicate the position of rs1876453. (C) Transancestral meta-analysis p value generated using fixed and random model are highlighted as red and blue, respectively. The dashed line represents the significance level after Bonferroni correction. (D) Frequencies, ORs and p values of haplotypes formed by lupus-associated *CR2* SNPs in various ancestral groups. Haplotype H1 corresponds to the previously reported systemic lupus erythematosus (SLE)-associated haplotype and is highlighted in green.

were incubated in NP-40 lysis buffer and nuclei pelleted, resuspended in sodium dodecyl sulfate lysis buffer and sonicated five times on ice. Preleared soluble chromatin was incubated with anti-CTCF, an isotype control, or no antibody followed by Protein G agarose/salmon sperm DNA. qPCR was performed on immune complex-associated DNA using primers spanning rs1876453 (5'-GGAAAGTTTCTGTGCCGGA-3', 5'-GACAA TCAGGACCAGGCGGT-3'), SYBR Green-based detection, and the Illumina Eco Real-Time PCR System. A standard curve was constructed using a chromatin input control.

qPCR of EBV-transformed cell transcripts was performed using cDNA transcribed using the Superscript VILO cDNA synthesis kit (Life Technologies), primers specific for  $\beta$ -actin (5'-GATGACCCAGATCATGTTTGAG-3', 5'-GACTCCATG CCCAGGAAGGAA-3'), *CTCF* (5'-CAACCAGCCCCAACAGAA CCAG-3', 5'-TCCTCTTCTCCTCCTCCTGTC-3'), *CR2* (5'-TG CCTGTAAACCAACTTCTC-3', 5'-AGCAAGTAACCAGATT CACAG-3') and *CR1* (5'-TGCTAAGGACAGGTGCAGAC-3',

5'-GGCAGACGAGGAACCAATGA-3'), SYBR Green-based detection, and the Eco Real-Time PCR System (Illumina). Transcript abundance was determined by comparison with a standard curve and normalised to  $\beta$ -actin.

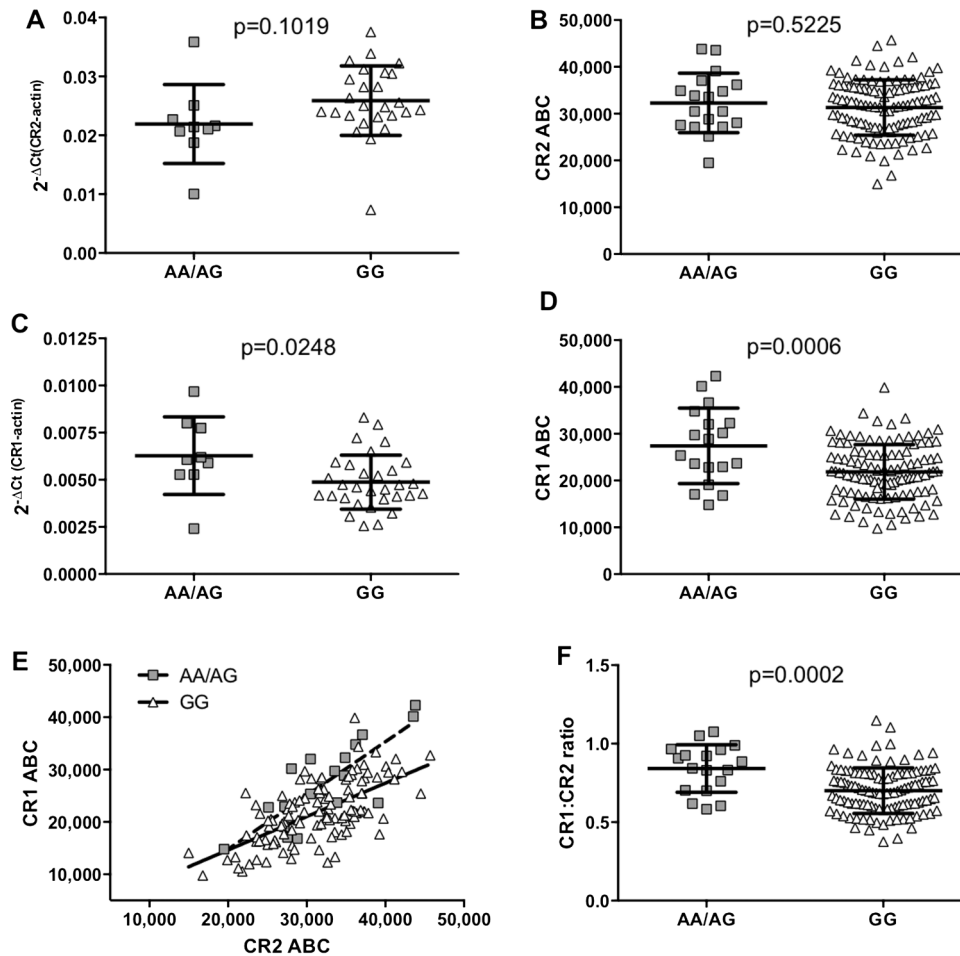
### Statistical analysis

A two-tailed unpaired Student t test detected differences between groups. Spearman rank correlation determined correlation within a group. A p value of <0.05 was considered significant. Statistics and graphs were generated using GraphPad Prism software.

## RESULTS

### Association of rs1876453 with decreased risk of lupus

To localise causal variants of the *CR2* gene, we genotyped 56 SNPs in *CR2* and *CR1* and 347 AIMs in 15 750 unrelated case-control subjects from four ancestral groups: EA (3872 cases vs



**Figure 2** Allele-specific effects of rs1876453 on *CR1* and *CR2* expression. Relative amounts of *CR2* (A) or *CR1* (C) RNA transcripts in primary B cells from 35 or 40 healthy donors respectively were measured by qPCR using the comparative Ct method.<sup>17</sup> Levels of surface *CR2* (B) and *CR1* (D) on primary B cells from 131 healthy donors were determined by quantitative flow cytometry (ABC, antibody binding capacity). Each point represents a unique subject, and the line and error bars represent the mean  $\pm$  SD for each group; p values were determined using a two-tailed Student t test, and a p value of  $<0.05$  was considered significant. (E) Levels of surface *CR1* and *CR2* were plotted and subjected to correlation analysis and linear regression. Each point represents a unique subject (minor allele, squares; major allele, triangles) and the lines represent the line of best fit for each allele (minor allele, dashed line; major allele, solid line). (F) Levels of surface *CR1* and *CR2* were also used to calculate the ratio of *CR1* to *CR2*. Each point represents a unique subject, and the line and error bars represent the mean  $\pm$  SD for each group; p values were determined using a two-tailed Student t test and a p value of  $<0.05$  was considered significant.

3449 controls), AA (1676 vs 1929), AS (1265 vs 1260) and HS (1492 vs 807). Genotypes for additional variants (SNPs and INDELS) were imputed using reference data from the 1000 Genomes Project. After applying quality control measures, 138 EA, 167 AA, 102 AS and 133 HS genetic variants deeply covering a 57.6 kB region 5' upstream of *CR2* to intron 1 of *CR1* were assessed for association with SLE (see online supplementary figure S1 and table S1). Only rs1876453, in intron 1 of *CR2*, was consistently associated with SLE in EA (MAF 0.086 in cases vs 0.099 in controls,  $p=0.045$ , OR 0.89 (95% CI 0.79 to 0.99)), AA (0.075 vs 0.086,  $p=0.045$ , OR 0.83 (95% CI 0.70 to 0.99)) and HS (0.032 vs 0.051,  $p=3.5 \times 10^{-3}$ , OR 0.62 (95% CI 0.46 to 0.86)). rs1876453 was not associated with SLE in AS given the low MAF in this group (0.0004 vs 0.0008,  $p=0.68$ , OR 0.60 (95% CI 0.05 to 6.74)). In transancestral meta-analysis of 75 genetic variants assessed in EA, AA, AS and HS (see online supplementary figure S1 and table S1), the strongest association signal was detected at rs1876453 ( $p=4.2 \times 10^{-4}$ , OR 0.85) and only this signal reached the Bonferroni-corrected significance level ( $p < 6.7 \times 10^{-4}$ ).

### Increased signal associated with rs1876453 after subphenotype stratification

We next assessed the association of rs1876453 with 14 lupus subphenotypes. In case controls, the minor allele was consistently associated with decreased risk of serositis, renal disorder, haematological disorder and dsDNA autoantibodies in EA, AA and HS, of which the strongest association was detected with dsDNA autoantibodies ( $p_{EA}=6.6 \times 10^{-4}$ , OR 0.73 (95% CI 0.61 to 0.88);  $p_{AA}=9.7 \times 10^{-3}$ , OR 0.72 (95% CI 0.56 to 0.92);  $p_{HS}=7.1 \times 10^{-3}$ , OR 0.58 (95% CI 0.39 to 0.86);  $p_{meta}=7.6 \times 10^{-7}$ , OR 0.71) (table 1). In cases only, the minor allele of rs1876453 was associated with decreased risk of dsDNA autoantibodies ( $p_{meta}=1.9 \times 10^{-4}$ , OR 0.75) and renal disorder ( $p_{meta}=1.3 \times 10^{-2}$ , OR 0.83) (table 1), but only the association with dsDNA autoantibodies remained significant after correction for multiple comparisons.

### rs1876453 tags the protective EA haplotype

No other SNP in this region exhibited stronger association with dsDNA autoantibodies in EA, AA and HS than rs1876453

(figure 1A–C, see online supplementary table S2). When conditioning on rs1876453, the association signals at all other SNPs were eliminated (see online supplementary table S2). The H1 haplotype, constructed using rs1876453 and previously associated or tightly linked CR2 SNPs, was equivalent to the protective EA SLE haplotype<sup>8</sup> and was perfectly tagged in EA by the minor allele of rs1876453 (figure 1D). In all three ancestral groups, it was associated with decreased risk of SLE ( $p_{EA}=0.037$ ,  $p_{AA}=0.043$  and  $p_{HS}=0.042$ ) and dsDNA autoantibodies ( $p_{EA}=6.0 \times 10^{-4}$ ,  $p_{AA}=7.8 \times 10^{-3}$  and  $p_{HS}=0.034$ ).

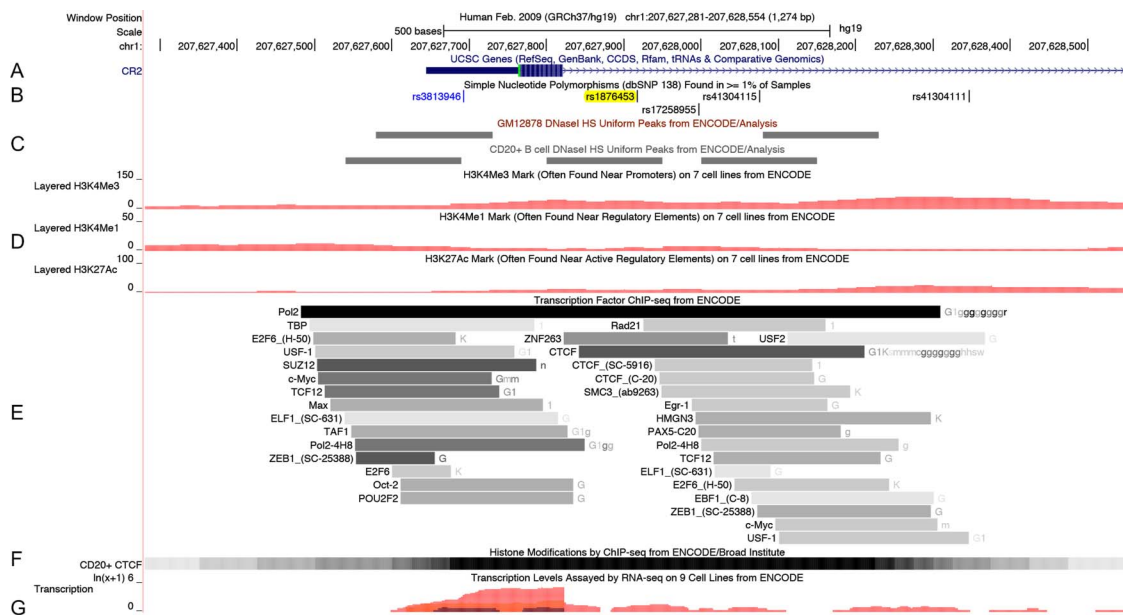
### The minor allele at rs1876453 is associated with altered expression of B cell CR1

We did not detect allele-specific differences in CR2 mRNA or protein levels in primary B cells from healthy controls (figure 2A, B). However, CR1 mRNA and protein levels were significantly higher in subjects with the minor A allele (figure 2C, D). Although B cell CR1 and CR2 levels were positively correlated in both groups (major allele Spearman  $r=0.6342$ ,  $p<0.0001$ ; minor allele Spearman  $r=0.7998$ ,  $p<0.0001$ , figure 2E), the ratio of CR1 to CR2 on B cells was higher in subjects with the minor allele ( $p=0.0002$ ; figure 2F). Allele-specific

differences in CR1 expression were not detected on other peripheral blood cells (see online supplementary figure S2). These data suggest that rs1876453 has long-range effects on the regulation of expression of CR1, which lies directly 3' of CR2, that are either B cell-specific or dependent on coexpression of CR2.

### The minor allele at rs1876453 alters transcription factor binding

rs1876453 is 97 nucleotides from the 5' end of the first intron of CR2. This ~12 kb intron contains a conserved silencing domain that controls CR2 expression in a cell type-specific and developmentally regulated manner.<sup>19–22</sup> The ENCYCLOPEDIA OF DNA ELEMENTS (ENCODE) database<sup>23</sup> reports the in vivo interaction of the region surrounding rs1876453 with multiple transcription factors in B cells, including Pax5, Oct2 and CTCF (figure 3). We confirmed the formation of multiple protein-DNA complexes containing the sequence surrounding rs1876453 by electrophoretic mobility shift assay (EMSA) using NE from several cell lines (K562, erythroid, CR1<sup>+</sup>CR2<sup>-</sup>; Reh, pre-B, CR1<sup>+</sup>CR2<sup>-</sup>; Ramos, mature B, CR1<sup>+</sup>CR2<sup>+</sup>, see online supplementary figure S3). In the presence of the minor A allele at rs1876453, transcription factor binding was reduced or



**Figure 3** The ENCYCLOPEDIA OF DNA ELEMENTS (ENCODE) Project data surrounding rs1876453. (A) The first exon and 5' end of the first intron of the CR2 gene. The 5' untranslated region (5' UTR) is shown before the methionine start codon, in green. These data are derived from the University of California Santa Cruz (UCSC) Genes Track. (B) The location of rs1876453 (highlighted in yellow) and previously reported systemic lupus erythematosus-associated rs3813946 (in blue font). These data are derived from the Common Single-Nucleotide Polymorphisms (SNP) (138) Track ([ft.ncbi.nlm.nih.gov/snp](http://ft.ncbi.nlm.nih.gov/snp)).<sup>39</sup> (C) DNaseI hypersensitive sites in the GM12878 Epstein–Barr virus (EBV)-transformed B cell line and in primary CD20+ B cells derived by DNase-seq. These data are derived from the UCSC Uniform DNaseI HS Track. Signal values are shown as grayscale-coloured items where higher signal values correspond to darker-coloured blocks. Primary B cells contain an additional hypersensitivity site that overlies rs1876453. (D) Histone marks surrounding rs1876453, as determined by chromatin immunoprecipitation (ChIP)-seq. These data were derived from the Layered H3K4Me3, H3K4Me1 and H3K27Ac Tracks. The H3K4Me3 histone mark is associated with poised or active promoters, the H3K4me1 histone mark is associated with enhancers and with DNA regions downstream of transcription sites and the H3K27Ac histone mark may enhance transcription by blocking the spread of the repressive histone mark H3K27Me3. Data shown are for the GM12878 EBV-transformed B cell line. (E) Transcription factor binding sites determined by ChIP-seq are shown as grey boxes that encompass the peaks of transcription factor occupancy, with the darkness of the box proportional to the maximal signal strength observed in any cell line. The name to the left of the box is the transcription factor, and includes in parentheses the antibody used for ChIP. The letters to the right of the box indicate the cell lines in which a signal is detected, with the darkness of the letter proportional to the signal strength in the cell line. Data are derived from the Transcription Factor ChIP Track. CCCTC-binding factor (CTCF) binding was seen in multiple EBV-transformed B cell lines (G, g) as well as a variety of other cell lines. (F) CTCF binding to primary CD20+ B cells by ChIP-seq. Peak occupancy lies over exon 1 and the 5' UTR. Data are derived from the Broad Histone Track. (G) Transcription levels for several cell types assayed by high-throughput sequencing of polyadenylated RNA (RNA-seq). Each cell line is associated with a particular colour; the GM12878 cell line is shown in pink. This figure was obtained from the UCSC Genome Browser (Human Feb 2009 (GRCh37/hg19) Assembly; <http://genome.ucsc.edu>).<sup>40</sup>

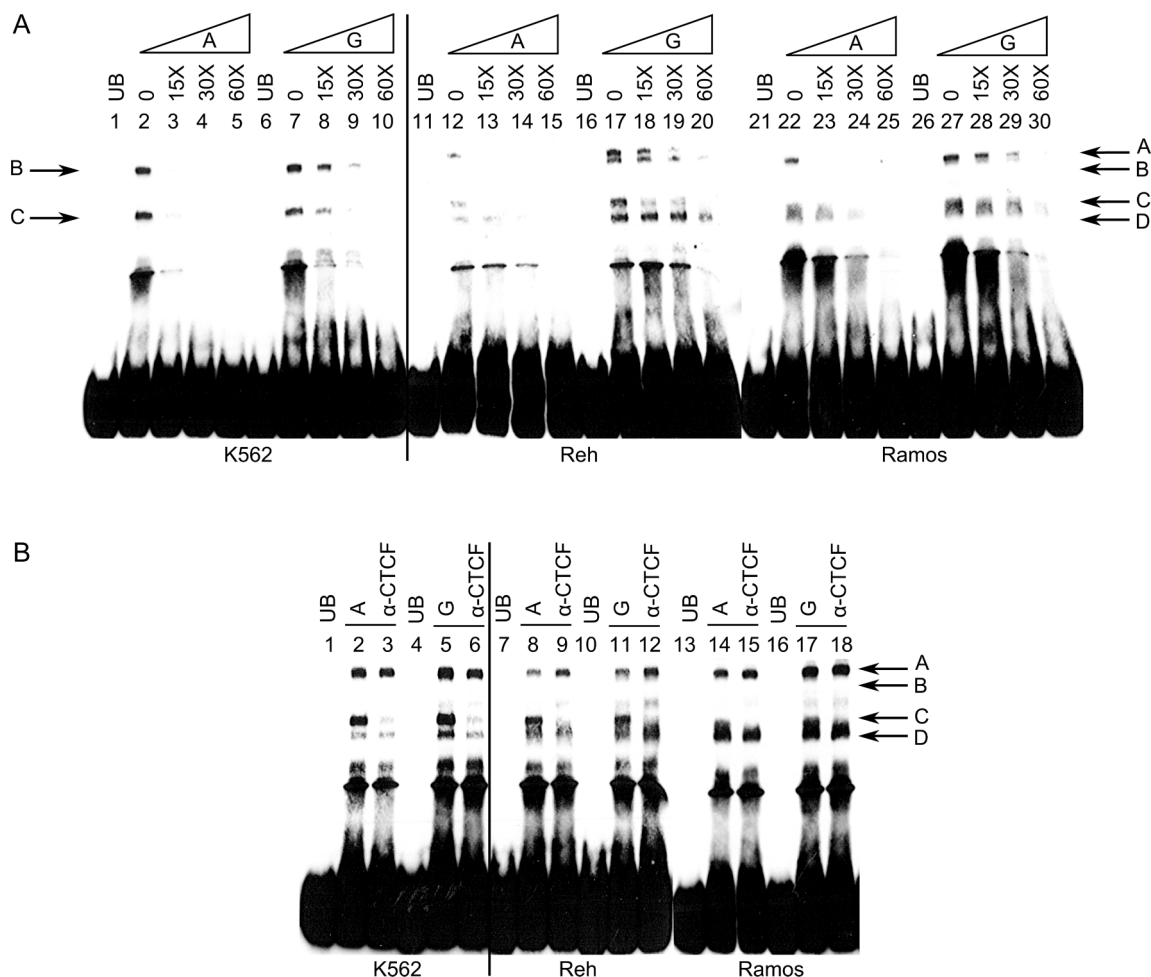
ablated and less cold competitor was required to out-compete protein-DNA complexes (figure 4A). Addition of an antibody to CTCF blocked the formation of complex C (figure 4B), and specific enrichment of the region surrounding rs1876453 was three-fold higher in the presence of the major allele by ChIP-PCR ( $p=0.0178$ ; figure 5A–D), suggesting that CTCF binds the region surrounding rs1876453 in an allele-specific manner. CTCF abundance was comparable between the two cell lines (figure 5E), and transcript abundance of *CR1* relative to *CR2* was higher in the cell line homozygous for the minor allele ( $p=0.023$ ; figure 5F), consistent with the data from primary B cells.

## DISCUSSION

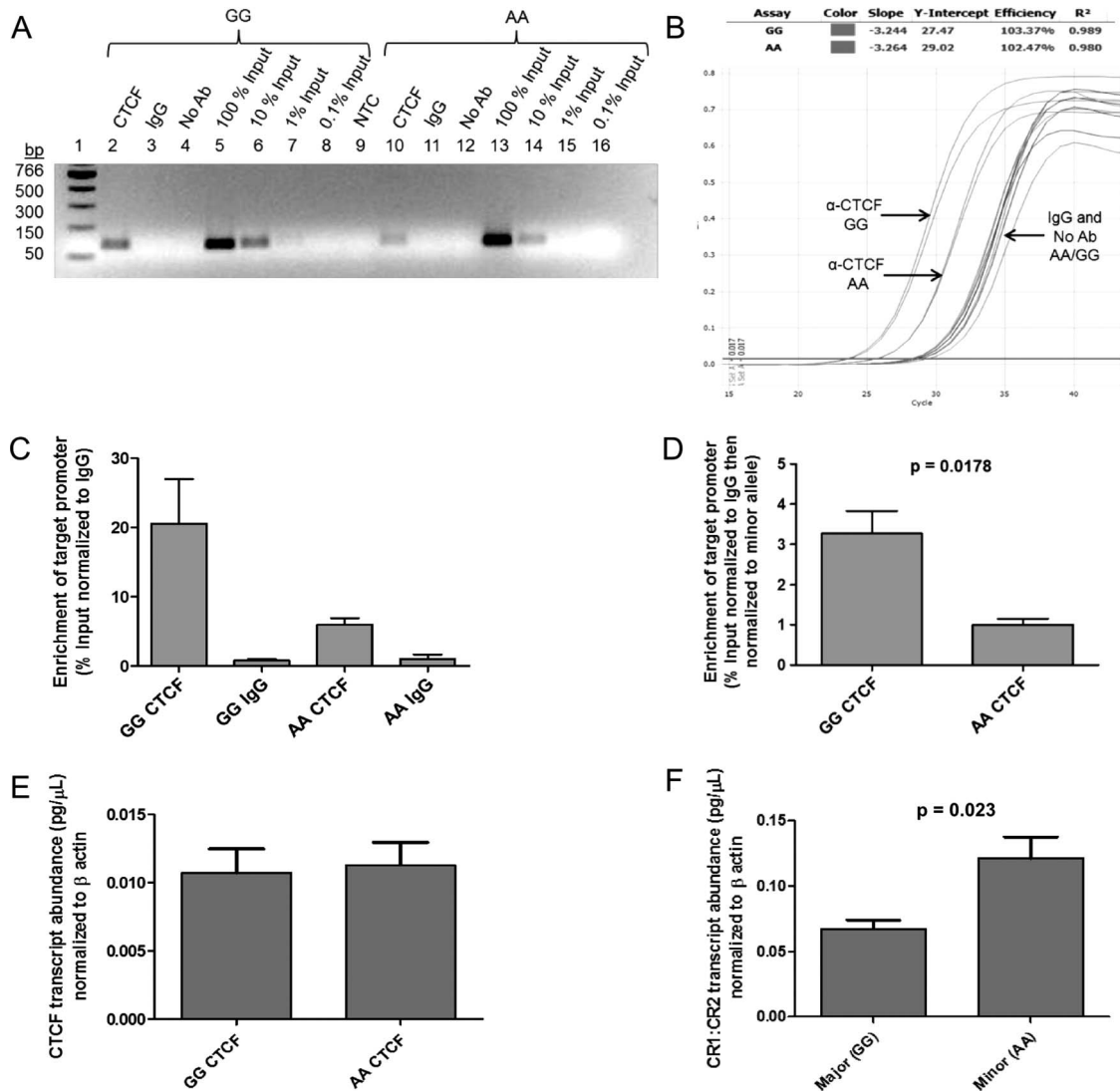
The minor allele of rs1876453, located in intron 1 of *CR2*, was associated with decreased risk of lupus in three of four ancestral groups studied, with the strongest association when cases were stratified based on the presence of dsDNA autoantibodies. This allele altered the formation of multiple DNA-protein complexes, including one containing CTCF, which has been termed the master regulator of chromatin organisation. Furthermore, it was

associated with increased B cell expression of *CR1*, suggesting long-range effects of rs1876453 on gene regulation and providing a plausible mechanism by which it may alter lupus susceptibility.

We previously identified the SLE-associated haplotype tagged by the minor allele of rs1876453 (H1 as shown in figure 1D) in EA,<sup>8</sup> and show here that it is found also in AA and HS (figure 1D). Although it is difficult to prove that a SNP is causal based on association testing, our transancestral study, which analysed a high density of SNP markers and capitalised on the weak linkage disequilibrium around rs1876453 in AA (figure 1B), provides compelling evidence that rs1876453 rather than the previously identified SNPs (rs3813946, rs1048971, rs17615 and rs4308977) or SNPs tightly linked to rs1876453 in EA (rs17258955 and rs61821130 in intron 1, rs61735651 in exon 14 and rs7549152 in intron 15) best explains the association signal. Modest association signals of rs1876453 increased after stratification of subjects based on specific clinical manifestations of lupus, with the strongest signals in the case-control analysis when subjects were stratified based on the presence of dsDNA autoantibodies. Although increased signals were seen with other



**Figure 4** Allelic differences in complex formation at rs1876453. (A) Protein-DNA complexes (indicated by arrows; A–D) were formed with oligonucleotides including either the minor A or major G allele in the absence or presence of K562 (Lanes 1–10), Reh (Lanes 11–20) and Ramos (Lanes 21–30) nuclear extracts. Specificity and binding affinity of the protein-DNA complexes were demonstrated by the addition of 15-molar to 60-molar excess of unlabelled oligonucleotides. UB represents unbound control. (B) Anti-CTCF (CCCTC-binding factor) antibody was included during the formation of protein-DNA complexes to determine whether CTCF was involved in forming these complexes. Data shown are representative of at least three independent experiments.



**Figure 5** CCCTC-binding factor (CTCF) interacts with *CR2* intron 1 in vivo and demonstrates differential affinity for rs1876453 alleles. (A) Chromatin immunoprecipitation performed using an antibody specific for CTCF yielded allele-specific enrichment of the region surrounding rs1876453 from Epstein–Barr virus (EBV)-transformed B cells homozygous for the major or minor allele at rs1876453. The qPCR products were visualised by ethidium bromide staining of a 1.8% agarose gel and sized using a PCR marker (New England Biolabs). A non-specific IgG control (IgG) and a control without antibody (No Ab) were included to measure background enrichment. Decreasing amounts of enrichment were observed with serially diluted input samples (Lanes 5–8; 13–16). NTC, no template control. (B) A representative qPCR amplification plot. (C) The percentage enrichment at the *CR2* promoter was determined by quantification against the standard curve. CTCF enrichment was normalised to the background enrichment generated by a non-specific IgG. (D) CTCF enrichment normalised to the minor allele at rs1876453. (E) CTCF transcript abundance for each homozygous cell line after normalisation to  $\beta$ -actin. (F) Transcript abundance of *CR1* relative to *CR2* in each homozygous cell line, with each transcript normalised to  $\beta$ -actin. Data shown are the mean  $\pm$  SEM for three independent experiments.

clinical manifestations, only dsDNA autoantibodies were significantly associated in both case-control and case-only analyses, suggesting that rs1876453 has a preferential effect on the production or presence of dsDNA autoantibodies. The mechanism underlying this dsDNA specificity is unclear, and characterising it will lead to a greater understanding of the etiopathogenesis of lupus, since dsDNA autoantibodies are highly specific for SLE (reviewed in refs<sup>24–25</sup>), and can predict onset and exacerbation of disease.<sup>26–28</sup> Anti-dsDNA antibody levels in the serum of genotyped patients with SLE did not correlate with rs1876453 genotype (data not shown), but therapy and other disease-related factors may obscure an association. Future studies with longitudinal analyses of patients will address this more stringently.

rs1876453 is located in a transcriptional regulatory region defined by DNase hypersensitivity and containing histone marks associated with active enhancers (H3K27ac, H3K4me1 and H3K4me3) (figure 3). We demonstrated that the minor allele at rs1876453 alters the formation of multiple protein-DNA complexes (figure 4A) and that CTCF is a component of one of these complexes (figures 4B and 5). CTCF, a highly conserved and ubiquitously expressed zinc finger protein, is a master regulator of genome spatial organisation that mediates chromatin loops within the genome.<sup>29</sup> By binding to insulators and boundary elements, it demarcates chromatin into regulatory regions and blocks communication between promoters and enhancers to regulate gene expression. CTCF interacted with the region surrounding rs1876453 in both lymphoblastoid B cell lines and



primary B cells,<sup>23</sup> with peak occupancy directly over rs1876453 (figure 3), and our data demonstrate an allele-dependent effect on its binding in vivo (figure 5) with less binding of CTCF to the minor A allele, which is associated with higher CR1 expression and is protective, and more binding to the major G risk allele, which is associated with lower CR1 expression. These data suggest that CTCF may have a repressive effect on CR1 transcription when bound to this locus. Efforts to identify the other transcription factors differentially binding this region as a result of the allele at rs1876453 are currently underway.

CR1 expression was increased in B cells of subjects harbouring the minor allele at rs1876453 at both the mRNA and protein level (figure 2). Eukaryotic transcription is tightly controlled by regulatory elements that, though perhaps distant from their target genes, can be brought into close proximity with the general transcriptional machinery at transcription start sites through the formation of chromatin loops.<sup>30</sup> Therefore, altered CTCF binding at intron 1 of CR2 may directly affect CR1 transcription via alterations in the chromatin looping that regulates these genetic regions. Alternatively, CR1 expression may be modified indirectly by regulatory RNA generated from intronic CR2 sequences. The minor allele at rs1876453 introduces a cryptic splice acceptor site (consensus value (CV) 71.05, CV variation of mutant from wild type 68.75%) and is predicted to break splice enhancer and silencer motifs (Human Splicing Finder<sup>31</sup>). Although there are no known microRNA sequences in the first intron of CR2, several expressed sequence tags have been aligned to this region in B cells<sup>32</sup> and the ENCODE project has determined that it is actively transcribed (figure 3).

CR1 binds C3b and C4b activation fragments and is physically associated with CR2 on B cells.<sup>33</sup> Although it has an independent negative regulatory role in B cell receptor (BCR)-mediated B cell activation,<sup>34</sup> it also acts as a cofactor for the Factor I-mediated cleavage of C4b to iC4b and of C3b to iC3b and C3dg. iC3b and C3dg are specific ligands for CR2<sup>35</sup> that either augment or inhibit signalling through the BCR in mature B cells,<sup>36–37</sup> depending on ligand valency. Strength of BCR signalling may also affect tolerance induction in transitional B cells.<sup>38</sup> Therefore, we hypothesise that increased expression of CR1 associated with rs1876453, in individuals susceptible to lupus, may tolerate transitional or arrest mature dsDNA-specific B cells that encounter complement-coated apoptotic debris and, as a consequence, modify the initiation or course of lupus.

In sum, we show that the minor allele at rs1876453 reduces risk of lupus and that this effect is due, in part, to increased expression of CR1 on B cells. Since this allele alters binding of transcription factors with long-range effects, allelic variation at this SNP may modify expression of additional genes involved in lupus pathogenesis. Given the common association of rs1876453 across ancestral groups, further investigation of its mechanisms and effects has broad implications for the management of patients with this disease.

#### Author affiliations

<sup>1</sup>Division of Rheumatology, Department of Medicine, University of California at Los Angeles, Los Angeles, California, USA

<sup>2</sup>Division of Rheumatology, University of Colorado School of Medicine, Aurora, Colorado, USA

<sup>3</sup>School of Pathology and Laboratory Medicine, Centre for Genetic Origins of Health and Disease, The University of Western Australia, Crawley, Western Australia, Australia

<sup>4</sup>Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA

<sup>5</sup>Department of Biostatistical Sciences and Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA

<sup>6</sup>Pfizer-Universidad de Granada-Junta de Andalucía Center for Genomics and Oncological Research, Granada, Spain

<sup>7</sup>Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA

<sup>8</sup>Center for Autoimmune Diseases Research (CREA), Universidad del Rosario, Bogotá, Colombia

<sup>9</sup>Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, South Korea

<sup>10</sup>Rosalind Russell/Ephraim P. Engleman Rheumatology Research Center, University of California San Francisco, San Francisco, California, USA

<sup>11</sup>Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA

<sup>12</sup>Division of Rheumatology, Medical University of South Carolina, Charleston, South Carolina, USA

<sup>13</sup>Department of Medicine, University of Southern California, Los Angeles, California, USA

<sup>14</sup>Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA

<sup>15</sup>Department of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA

<sup>16</sup>Department of Clinical Pharmacology, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA

<sup>17</sup>Division of Rheumatology and Department of Immunology, Mayo Clinic, Rochester, Minnesota, USA

<sup>18</sup>Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

<sup>19</sup>Division of Rheumatology, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

<sup>20</sup>Department of Rheumatology and Clinical Immunogenetics, University of Texas Health Science Center at Houston, Houston, Texas, USA

<sup>21</sup>US Department of Veterans Affairs Medical Center, Oklahoma City, Oklahoma, USA

<sup>22</sup>Division of Rheumatology, Department of Pediatrics, University of Washington, Seattle, Washington, USA

<sup>23</sup>Center for Immunity and Immunotherapies, Seattle Children's Research Institute, Seattle, Washington, USA

<sup>24</sup>Division of Rheumatology, Department of Medicine, University of Puerto Rico Medical Sciences Campus, San Juan, Puerto Rico

<sup>25</sup>Division of Genetics and Molecular Medicine and Immunology, King's College London, London, UK

<sup>26</sup>Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA

<sup>27</sup>US Department of Veterans Affairs Medical Center, Cincinnati, Ohio, USA

<sup>28</sup>Department of Epidemiology, University of Alabama at Birmingham, Birmingham, Alabama, USA

<sup>29</sup>Denver Veterans Affairs Medical Center, Denver, Colorado, USA

**Acknowledgements** We thank the volunteers who participated in this study; Agnessa Gadiliya, Carissa Homme, Najuk Jain and Lauren Kuhlman (University of Colorado School of Medicine, Aurora, Colorado, USA) for recruiting subjects, processing samples and technical assistance; Krista Bean (OMRF, Oklahoma City, Oklahoma, USA) for assistance with data analysis; Dr Henry Marsh and Celldex Therapeutics, Boston, Massachusetts, USA for providing the anti-CR1 monoclonal antibody 6B1.H12; Dr Ellen Vitetta from University of Texas-Southwestern, Dallas, Texas, USA for providing the anti-CD19 monoclonal antibody HD37; and Drs V. Michael Holers, Tasha E Fingerlin, and Richard A Spritz (University of Colorado School of Medicine, Aurora, Colorado, USA) for helpful discussions. The BIOLUPUS network is composed of Johan Frostegård, MD, PhD (Huddinge, Sweden), Lennart Truedsson, MD, PhD (Lund, Sweden), Enrique de Ramón, MD PhD (Málaga, Spain), José M Sabio, MD, PhD (Granada, Spain), María F González-Escribano, PhD (Sevilla, Spain), Javier Martín, MD, PhD (Granada, Spain), Norberto Ortego-Centeno (Granada, Spain), José Luis Callejas MD (Granada, Spain), Julio Sánchez-Román, MD (Sevilla, Spain), Sandra D'Alfonso, PhD (Novara, Italy), Sergio Migliarese MD (Napoli, Italy), Gian-Domenico Sebastiani MD (Rome, Italy), Mauro Galeazzi MD (Siena, Italy), Torsten Witte, MD, PhD (Hannover, Germany), Bernard R Lauwerys, MD, PhD (Louvain, Belgium), Eموke Endreffy, PhD (Szeged, Hungary), László Kovács, MD, PhD (Szeged, Hungary), Carlos Vasconcelos, MD, PhD (Porto, Portugal) and Berta Martins da Silva, PhD (Porto, Portugal). The members of GENLES Network are Hugo R Scherbarth, Pilar C Marino, Estela L Motta, Susana Gamron, Cristina Drenkard, Emilia Menso, Alberto Allievi, Guillermo A Tate, Jose L Presas, Simon A Palatnik, Marcelo Abdala, Mariela Bearzotti, Alejandro Alvarillos, Francisco Caeiro, Ana Bertoli, Sergio Paire, Susana Roverano, Cesar E Graf, Estela Bertero, Cesar Caprarulo, Griselda Buchanan, Carolina Guillerón, Sebastian Grimaudo, Jorge Manni, Luis J Catoggio, Enrique R Soriano, Carlos D Santos, Cristina Prigione, Fernando A Ramos, Sandra M. Navarro, Guillermo A Berbotto, Marisa Jorfen, Elisa J Romero, Mercedes A Garcia, Juan C Marcos, Ana I Marcos, Carlos E Perandones, Alicia Eimon, Sanatorio Parque and Cristina G. Battagliotti in Argentina; Eduardo Acevedo and Mariano Cucho in Perú; Ignacio García de la Torre, Mario Cardiel Ríos, José Francisco Moctezuma and Marco Maradiaga Ceceña in Mexico.

The UCSC Genes Track from ENCODE was produced at UCSC using a computational pipeline developed by Jim Kent, Chuck Sugnet, Melissa Cline and Mark Diekhans. It is based on data from NCBI RefSeq, UniProt (including TrEMBL and TrEMBL-NEW), CCDS and GenBank, as well as data from Rfam and the Todd Lowe lab. The UCSC Uniform DNaseI HS Track was created by the 'Open Chromatin' (Duke/UNC/UT-A) ENCODE group: Duke University Institute for Genome Sciences & Policy (IGSP): Alan Boyle, Lingyun Song, and Greg Crawford, University of North Carolina at Chapel Hill; Paul Giresi, Jason Lieb, and Terry Furey, University of Texas at Austin; Zheng Liu, Ryan McDaniel, Bum-Kyu Lee and Vishy Iyer. European Bioinformatics Institute: Paul Flicek, Damian Keefe, and Ewan Birney, and University of Cambridge, Department of Oncology and CR-UK Cambridge Research Institute (CRI): Stefan Graf with support from NHGRI and by the University of Washington (UW) ENCODE group (contact: Richard Sandstrom). The Layered H3K4Me3, H3K4Me1 and H3K27Ac Tracks show data from the Bernstein Lab at the Broad Institute. The Transcription Factor ChIP-seq Track shows data from the Myers Lab at the Hudson Alpha Institute for Biotechnology, the labs of Michael Snyder, Mark Gerstein and Sherman Weissman at Yale University, Peggy Farnham at UC Davis, Kevin Struhl at Harvard, Kevin White at The University of Chicago, and Vishy Iyer at The University of Texas Austin. The Broad Histone Track shows data generated at the Broad Institute and in the Bernstein lab at the Massachusetts General Hospital/Harvard Medical School, with data generation and analysis supported by funds from the NHGRI, the Burroughs Wellcome Fund, Massachusetts General Hospital and the Broad Institute (Contact: Noam Shores). The Transcription Track shows data from the Wold Lab at Caltech.

**Contributors** JCE, RPK, EEB, BPT, HW and SAB: selected SNPs for the association study, and PMG, KMS, JAK, KMK, CDL and JBH: were responsible for the design of the large lupus association study. MEA-R, GSA, J-MA, S-CB, SAB, EEB, LAC, JCE, BIF, PMG, GSG, JMG, JBH, COI, JAJ, DK, DLK, RPK, H-SL, JTM, KMS, TBN, MAP, RR-G, JDR, RHS, AMS, BPT, LMV and TJV: assisted in the collection and characterisation of the SLE cases and controls. AJA, KMK and PMG: performed the genotyping. SBG, AHW, MEC, JTZ, MCM, KMK and CDL: performed quality control and ancestry analyses. JZ: performed association analyses and imputation under the guidance of BPT, BMG, RLT, GAY, KML and HLN: performed functional studies under the guidance of DU, LJA, and SAB, JZ, BMG, RLT, DU, BPT and SAB: prepared the manuscript. All authors approved the final draft.

**Funding** The quality of the RNA samples prepared from healthy human subjects was evaluated in the University of Colorado Cancer Center Microarray Core, which is supported by the NIH/NCI Cancer Core Support Grant (P30 CA046934). Flow cytometry was carried out in the Barbara Davis Center Flow Cytometry Core Facility, which was supported by National Institutes of Health (NIH) grant P30 DK57516. This work was also supported by the US National Institutes of Health [R01AI070983 (S.A.B., B.P.T., D.U.), K24AI078004 (S.A.B.), T32AR07534 (B.M.G.), K24AR002138 (R.R.G.), LRP0171651 (T.B.N.), K08AI083790 (T.B.N.), N01AR062277 (J.B.H.), P01AI083194 (J.B.H.), P01AR049084 (R.P.K., J.B.H., J.C.E., E.E.B., G.S.A., J.D.R., R.R.G., and M.A.P.), P2ORR020143 (J.B.H.), P30AR048311 (E.E.B.), P30AR053483 (J.A.J. and J.M.G.), P30AR055385 (E.E.B.), P30GM103510 (J.A.J.), P60AR030692 (R.R.G.), P60AR062755 (D.L.K.), P60AR053308 (L.A.C.), R01AI063274 (P.M.G.), R01AR033062 (R.P.K.), R01AR042460 (J.B.H.), R01AR043274 (K.M.S.), R01AR43727 (M.A.P.), R01AR043814 (B.P.T.), R01AR051545 (A.M.S.), R01AR057172 (C.O.J.), R01CA141700 (M.E.A.R.), R21AI070304 (S.A.B.), R37AI024717 (J.B.H.), RC1AR058621 (M.E.A.R.), U01AI101934 (J.A.J. and J.M.G.), U19AI082714 (J.A.J. and J.M.G.), U54RR023417 (J.D.R.), UL1RR024999 (T.B.N.), UL1RR025014 (A.M.S.), UL1RR025741 (R.R.G.), UL1RR025777 (R.P.K. and J.C.E.), UL1RR029882 (D.L.K.), and UL1TR000004 (L.A.C.)], the Alliance for Lupus Research (S.A.B., B.P.T., D.U., K.M.S., T.B.N., L.A.C. and C.O.J.), the Lupus Research Institute (B.P.T., T.B.N.), the US Department of Veterans Affairs (Merit Awards; J.B.H., G.S.G.), the US Department of Defense (PR094002, J.B.H.), the Arthritis National Research Foundation (Eng Tan Scholar Award; J.Z. and T.B.N.), the Arthritis Foundation (A.M.S., and P.M.G.), the Korea Healthcare Technology R&D Project, Ministry for Health and Welfare, Republic of Korea (A121983; S.C.B.), the European Science Foundation RNP (BIOLUPUS Research Network), the Wellcome Trust (T.J.V.), Arthritis Research UK (T.J.V.), a Kirkland Scholar Award (L.A.C.), and the Wake Forest School of Medicine Center for Public Health Genomics (C.D.L.). The funders had no role in study design, data collection, analysis and interpretation, writing of the report or decision to submit the paper for publication.

**Competing interests** None.

**Ethics approval** IRB at University of Colorado School of Medicine and multiple institutions.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open Access** This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>

## REFERENCES

- 1 Wedgwood RJP, Janeway CA. Serum complement in children with "collagen diseases". *Pediatrics* 1953;11:569–81.
- 2 Elliott JA, Mathieson DR. Complement in disseminated (systemic) lupus erythematosus. *AMA Arch Dermat Syphilol* 1953;68:119–28.
- 3 Prodeus AP, Georg S, Shen L-M, *et al.* A critical role for complement in the maintenance of self-tolerance. *Immunity* 1998;9:721–31.
- 4 Wu X, Jiang N, Deppong C, *et al.* A role for the *Cr2* gene in modifying autoantibody production in systemic lupus erythematosus. *J Immunol* 2002;169:1587–92.
- 5 Wilson JG, Ratnoff WD, Schur PH, *et al.* Decreased expression of the C3b/C4b receptor (CR1) and the C3d receptor (CR2) on B lymphocytes and of CR1 on neutrophils of patients with systemic lupus erythematosus. *Arth Rheum* 1986;29:739–47.
- 6 Boackle SA, Holers VM, Chen X, *et al.* *Cr2*, a candidate gene in the murine *Sle1c* lupus susceptibility locus, encodes a dysfunctional protein. *Immunity* 2001;15:775–85.
- 7 Wu H, Boackle SA, Hanvivadhanakul P, *et al.* Association of a common complement receptor 2 haplotype with increased risk of systemic lupus erythematosus. *Proc Natl Acad Sci USA* 2007;104:3961–6.
- 8 Douglas KB, Windels DC, Zhao J, *et al.* Complement receptor 2 polymorphisms associated with systemic lupus erythematosus modulate alternative splicing. *Genes Immun* 2009;10:457–69.
- 9 Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997;40:1725.
- 10 Smith MW, Patterson N, Lautenberger JA, *et al.* A high-density admixture map for disease gene discovery in African Americans. *Am J Hum Genet* 2004;74:1001–13.
- 11 Halder I, Shriver M, Thomas M, *et al.* A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Hum Mutat* 2008;29:648–58.
- 12 Price AL, Patterson NJ, Plenge RM, *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–9.
- 13 Hoggart CJ, Shriver MD, Kittles RA, *et al.* Design and analysis of admixture mapping studies. *Am J Hum Genet* 2004;74:965–78.
- 14 Lessard CJ, Adrianto I, Kelly JA, *et al.* Identification of a systemic lupus erythematosus susceptibility locus at 11p13 between PDHX and CD44 in a multiethnic study. *Am J Hum Genet* 2011;7:83–91.
- 15 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5:e1000529.
- 16 Purcell S, Neale B, Todd-Brown K, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
- 17 Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative  $C_T$  method. *Nat Protoc* 2008;3:1101–8.
- 18 Cruickshank MN, Fenwick E, Karimi M, *et al.* Cell- and stage-specific chromatin structure across the complement receptor 2 (CR2/CD21) promoter coincides with CBF1 and C/EBP- $\beta$  binding in B cells. *Mol Immunol* 2009;46:2613–22.
- 19 Zabel MD, Byrne BL, Weis JJ, *et al.* Cell-specific expression of the murine CD21 gene depends on accessibility of promoter and intronic elements. *J Immunol* 2000;165:4437–45.
- 20 Hu H, Martin BK, Weis JJ, *et al.* Expression of the murine CD21 gene is regulated by promoter and intronic sequences. *J Immunol* 1997;158:4758–68.
- 21 Makar KW, Pham CTN, Dehoff MH, *et al.* An intronic silencer regulates B lymphocyte cell- and stage-specific expression of the human complement receptor type 2 (CR2, CD21) gene. *J Immunol* 1998;160:1268–78.
- 22 Makar KW, Ulgiati D, Hagman J, *et al.* A site in the complement receptor 2 (CR2/CD21) silencer is necessary for lineage specific transcriptional regulation. *Int Immunol* 2001;13:657–64.
- 23 Drezner TR, Karolchik D, Zweig AS, *et al.* The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res* 2012;40:D918–23.
- 24 Hahn B. Antibodies to DNA. *N Engl J Med* 1998;338:1359–68.
- 25 Kavanaugh AF, Solomon DH, Guidelines TACoRAHCoIT. Guidelines for immunologic laboratory testing in the rheumatic diseases: anti-DNA antibody tests. *Arthritis Rheum* 2002;47:546–55.
- 26 ter Borg EJ, Horst G, Hummel EJ, *et al.* Measurement of increases in anti-double-stranded DNA antibody levels as a predictor of disease exacerbation in systemic lupus erythematosus. A long-term, prospective study. *Arthritis Rheum* 1990;33:634–43.
- 27 Arbuckle MR, James JA, Kohlhasse KF, *et al.* Development of anti-dsDNA autoantibodies prior to clinical diagnosis of systemic lupus erythematosus. *Scand J Immunol* 2001;54:211–19.
- 28 Arbuckle MR, McClain MT, Rubertone MV, *et al.* Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N Engl J Med* 2003;349:1526–33.

- 29 Ren L, Wang Y, Shi M, *et al.* CTCF mediates the cell-type specific spatial organization of the Kcnq5 locus and the local gene regulation. *PLoS ONE* 2012;7:e31416.
- 30 Gheldof N, Leleu M, Noordermeer D, *et al.* Detecting long-range chromatin interactions using the chromosome conformation capture sequencing (4C-seq) method. *Methods Mol Biol*; 2012:211–25.
- 31 Desmet F-O, Hamroun D, Lalande M, *et al.* Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009;37:e67.
- 32 Boguski MS, Lowe TMJ, Tolstoshev CM. dbEST—database for “expressed sequence tags”. *Nat Genet* 1993;4:332–3.
- 33 Tuveson DA, Ahearn JM, Matsumoto AK, *et al.* Molecular interactions of complement receptors on B lymphocytes: a CR1/CR2 complex distinct from the CR2/CD19 complex. *J Exp Med* 1991;173:1083–9.
- 34 Józsi M, Prechl J, Bajtay Z, *et al.* Complement receptor type 1 (CD35) mediates inhibitory signals in human B lymphocytes. *J Immunol* 2002;202:2782–8.
- 35 Ahearn JM, Fearon DT. Structure and function of the complement receptors, CR1 (CD35) and CR2 (CD21). *Adv Immunol* 1989;46:183–219.
- 36 Chakravarty L, Zabel MD, Weis JJ, *et al.* Depletion of Lyn kinase from the BCR complex and inhibition of B cell activation by excess CD21 ligation. *Int Immunol* 2002;14:139–46.
- 37 Mongini PKA, Vilensky MA, Highet PF, *et al.* The affinity threshold for human B cell activation via the antigen receptor complex is reduced upon co-ligation of the antigen receptor with CD21 (CR2). *J Immunol* 1997;159:3782–91.
- 38 Habib T, Funk A, Rieck M, *et al.* Altered B cell homeostasis is associated with type I diabetes and carriers of the PTPN22 allelic variant. *J Immunol* 2012;188:487–96.
- 39 Sherry ST, Ward MH, Kholodov M, *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308–11.
- 40 Karolchik D, Barber GP, Casper J, *et al.* The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res* 2014;42:D764–770.