Regulatory architecture of the RCA gene cluster captures an intragenic TAD boundary and enhancer elements in B cells

Jessica Cheng[1,*], Joshua S. Clayton[2,3,*], Rafael D. Acemel[4], José L. Gómez-Skarmeta[4], Rhonda L. Taylor[1,2,3], John B. Harley[5,6,7], Elizabeth Quail[1,8] and Daniela Ulgiati[1]

[1] School of Biomedical Sciences, The University of Western Australia, Crawley, WA, 6009, Australia

[2] Harry Perkins Institute of Medical Research, QEII Medical Centre, Nedlands, WA, 6009, Australia

[3] Centre for Medical Research, The University of Western Australia, Crawley, WA, 6009, Australia

[4] Centro Andaluz de Biología del Desarrollo, Universidad Pablo de Olavide, Sevilla, Andalucía 41013, Spain

[5] Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229, United States

[6] Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, 45267, United States

[7] US Department of Veterans Affairs Medical Centre, US Department of Veterans Affairs, Cincinnati, OH, 45220, United States

[8] School of Molecular Sciences, The University of Western Australia, Crawley, WA, 6009, Australia

*These authors contributed equally to this work.

Correspondence: daniela.ulgiati@uwa.edu.au

ABSTRACT

The Regulators of Complement Activation (RCA) gene cluster comprises several tandemly arranged genes which share functions in the innate immune system. RCA members, such as complement receptor 2 (*CR2*), are well-established susceptibility genes in complex autoimmune diseases. Altered expression of RCA genes has been demonstrated at both the functional and genetic level, but the mechanism underlying their regulation are not fully characterised. We aimed to investigate the structural organisation of the RCA gene cluster to identify key regulatory elements that influence the expression of *CR2* and other genes in this immunomodulatory region. Using 4C, we captured extensive CTCF-mediated chromatin looping across the RCA gene cluster in B cells and showed these were organised into two topological associated domains (TADs). Interestingly, the inter-TAD boundary was located within the *CR1* gene at a well-characterised segmental duplication. Additionally, we mapped numerous gene-gene and gene-enhancer interactions across the region, revealing extensive co-regulation. Importantly, we identified an intergenic enhancer and functionally demonstrated this element upregulates two RCA members (*CR2* and *CD55*) in B cells. We have uncovered novel, long-range mechanisms whereby SLE susceptibility may be influenced by genetic variants, highlighting the important contribution of chromatin topology to gene regulation and complex genetic disease.

INTRODUCTION

The complement system is a major immune network of soluble proteins and membrane receptors which elicit potent, innate responses against pathogens, immune complexes and apoptotic cells (*1*). The complement system is activated by one of three major pathways (classical, alternative or lectin), triggering a series of proteolytic cleavage events which ultimately converge to form the C3 convertase. The C3 convertase enzyme catalyses, in part, the formation of complement effector peptides (C3a, C5a, C3b and C5b) which mediate local inflammation, cell lysis and cell clearance (*1*). Additionally, complement components are capable of binding to numerous immune cell types and activating other immune pathways, including adaptive B cell and T cell responses (*2*, *3*). Complement therefore represents an important bridge between the innate and adaptive immune systems and allows for effective co-ordination of immune responses (*4*).

The complement cascade is intricately controlled to ensure a sufficient immune response is generated while preventing damage to self (*1*). In humans, a number of these regulatory proteins are located in a gene cluster known as the Regulators of Complement Activation (RCA) at chromosome 1q32.2 (*5*). This includes the plasma protein C4 binding protein (encoded by alpha (*C4BPA*) and beta (*C4BPB*) subunits), and several membrane receptors: decay-accelerating factor (DAF, *CD55*), complement receptors 2 and 1 (*CR2* and *CR1*), and membrane co-factor protein (MCP, *CD46*) (*5*). Several duplicated pseudogenes within the RCA cluster have also been identified (*6*, *7*) of which CR1-like (*CR1L)* has been best characterised (*8*). All members of the RCA gene cluster are composed of tandem 60 – 70 amino acid motifs known as short consensus repeats (SCRs) which bind complement components and primarily regulate the complement response through inhibition or activation of C3 convertase (*1*). As such, this gene cluster is believed to have been derived from complex duplications of a common ancestral gene, followed by the diversification of function (*5*). In addition to their important role in innate immune response, members of RCA gene cluster are also involved in the process of tissue injury, inflammation and apoptosis. Accordingly, they have been implicated in a range of inflammatory and autoimmune disorders (*9*–*12*).

A role for CR2 has been well established in the autoimmune disease, Systemic Lupus Erythematosus (SLE). SLE is characterised by the presence of antibodies directed against nuclear antigens and has a complex aetiology with a strong genetic component (*13*, *14*). CR2 regulates B cell responses by modulating B cell activation and antibody production upon binding of complement-tagged antigens (*15*). Aberrant expression of CR2 on the surface of B cells has been demonstrated in both mouse models of the disease (*16*, *17*) and SLE patients (*18*, *19*), which functionally contributes to B cell autoreactivity and autoimmune disease susceptibility (*20*–*22*). Additionally, the *CR2* gene has also been implicated in SLE at the genetic level through linkage analyses (*23*–*25*) and association studies (*25*–*27*). Most recently, a SNP within the first intron of *CR2* (rs1876453) was shown to alter the expression of the neighbouring gene (*CR1*) without influencing CR2 expression (*27*), indicating that expression of these genes in the RCA cluster may be co-regulated. Functionally, rs1876453 was

shown to influence the binding affinity of CCCTC-binding factor (CTCF) to *CR2* (*27*). Whether RCA gene members are co-regulated and the role of CTCF in the RCA cluster are not yet understood.

CTCF is an important transcription factor which was first identified as an insulator of gene expression, and is now known to have several roles in gene regulation (*28*). More recently, CTCF has been shown to play an critical role in forming chromatin loops and mediating interactions between distal loci (*29*). Importantly, it has been recognised that chromatin loops are organised into genomic compartments known as topologically associated domains (TADs) (*30*). The current model proposed to explain TAD formation involves CTCF and the cohesin complex, whereby loops are dynamically formed through 'loop extrusion' between distal CTCF sites in convergent or 'forward-facing' orientation (*31*). TADs are recognised to be constitutively maintained in different cell types but may alternate between active ("A") and inactive ("B") compartment types depending on the cellular context (*32*, *33*). While genes within the same TAD tend to be co-expressed, not all genes within a TAD are necessarily expressed simultaneously. Rather, in a given context, TADs restrict chromatin interactions between genes and distal regulatory elements, such as enhancers, to ensure that gene expression is properly controlled (*32*, *34*).

Enhancers represent an important class of distal-regulatory elements which are largely responsible for governing cell-type specific gene expression patterns (*35*). Enhancers bind transcription factors to upregulate expression of genes and are located distal to gene promoters in the linear genome but are positioned in close proximity by chromatin looping (*35*). Importantly, the majority of disease-associated SNPs from genome-wide association studies (GWAS) fall within enhancer regions (*36*). Enhancer elements have been predicted in the genome by the presence of epigenetic marks such as enrichment of H3K27ac and expression of short, bi-directional transcripts termed enhancer RNA (eRNA) (*37*, *38*). However, enhancers can simultaneously regulate expression of multiple genes, regulate genes large distances away and skip their neighbouring gene/s, which has hindered the identification of their target gene/s (*35*). The mapping of chromatin interactions through high-throughput chromatin conformation capture technologies, such as Hi-C and capture Hi-C (CHi-C), has aided in this, but the physical chromatin interactions detected using these methods may not necessarily be functional (*39*). As such, experimental validation of enhancers and physically associating enhancer-gene pairs is imperative for understanding their influence on gene expression (*35*, *39*–*41*).

The aim of this investigation was to explore the structural organisation of the RCA gene cluster in order to identity regulatory elements which may co-regulate the expression of genes in this important immunomodulatory cluster. In this study, we examined genomic interactions across the RCA gene cluster using chromosome conformation capture and showed that long-range chromatin interactions are involved in the co-regulation and co-expression of several RCA members in the B cell lineage. Further, we identified an intragenic TAD boundary which organises chromatin interactions in the RCA gene cluster into two discrete domains. Importantly, we functionally interrogated a putative long-range

enhancer and demonstrated that it regulates two genes simultaneously within a TAD in B cells. Collectively, we have revealed how three-dimensional chromatin organisation plays an important role in regulating the RCA gene cluster and have uncovered novel regulatory loci which govern the expression of these genes.

MATERIAL AND METHODS

Cell culture

Cell lines Reh (CRL-8286), Raji (CCL-86), SKW (TIB-215), K562 (CCL-243) and HepG2 (HB-8065), were obtained from the American Type Culture Collection. B lymphoblastoid cell lines (B-0028 and B-0056) were derived from healthy individuals and immortalised by Epstein-Barr virus infection (*27*). All suspension cells were cultured in RPMI-1640 with L-glutamine (Life Technologies), supplemented with 10% FBS, 100 µg/mL penicillin and 100 ng/µL streptomycin. The adherent cell line (HepG2) was cultured in high glucose DMEM (Life Technologies) with 10% FBS, 100 µg/mL penicillin and 100 ng/µL streptomycin.

Circular chromosome conformation capture (4C-seq)

B-lymphoblastoid cell lines (5 x 10^6 cells) were harvested by centrifugation and resuspended in 5 mL PBS with 10% FBS. To cross-link cells, 5 mL 4% formaldehyde was added, and samples were incubated for 10 min. Cross-linking was quenched by adding 1 M glycine to 125 mM final concentration and cells collected by centrifugation at 300 x *g* for 10 min at 4°C. 4C-seq assays and data processing were performed as previously reported (*42*, *43*). Sequences of primers used as 4C viewpoints are listed in Supplementary Table 1.

Bioinformatic datasets

Hi-C data and TAD co-ordinates for GM12878 from Rao et al. (*44*) and Dixon et al. (*30*) were visualised using the 3D Genome Browser (*45*). CTCF orientation calls from GM12878 were retrieved from Rao et al. (*44*) or assessed using CTCFBSDB 2.0 (*46*). Enhancer predictions were retrieved from the GeneHancer database (Version J), which leverages data from multiple sources, including ENCODE, FANTOM5 and Ensembl. Histone modifications and transcription factor enrichment was assessed using ENCODE data and visualised on the UCSC Genome Browser on hg19.

Luciferase reporter-gene assays

Candidate enhancers were amplified from human genomic DNA using Q5 Hot-Start High-Fidelity DNA polymerase (New England Biolabs) and directionally cloned into the pGL3-Promoter plasmid (pGL3-P) (Promega) upstream of the SV40 promoter using restriction enzymes (Supplementary Table 1). Plasmid DNA was prepared using the EndoFree Plasmid Maxi Kit (QIAGEN) for transfection. Each enhancer construct (1 µg) was transiently transfected with the pRL-TK *Renilla* internal control vector (50 ng) using 4 µL Viafect™ transfection reagent (Promega) into suspension cell lines or adherent cell lines. Cell lysates were harvested after 24 h of incubation. Firefly and *Renilla* luciferase activity of cell lysates were sequentially assayed using the Dual-Luciferase Reporter Assay System (Promega) on a GloMax Explorer luminometer (Promega). Firefly luciferase readings were normalised to a co-transfected internal *Renilla* luciferase control, and the activity of each enhancer construct was normalised to a pGL3-P control. Sequences of primers used in this paper are listed in Supplementary Table 1.

Quantitative PCR

Total RNA was extracted from cells using the RNeasy Mini Kit (QIAGEN) with on-column DNase I treatment. RNA quantity and purity were determined by spectrophotometry. RNA was reverse-transcribed into cDNA using SuperScript III VILO reverse transcriptase (Life Technologies) and diluted with UltraPure dH_2O (Life Technologies). qPCR reactions comprised 1X SYBR Green No-Rox (Bioline), 250 nM forward and reverse primers, and 2 µL diluted cDNA up to a final volume of 10 µL. Cycling and analysis were conducted using a Mic qPCR Cycler (BioMolecular Systems) with the following conditions: 95°C for 10 min, and 35 cycles of 95°C for 15 s, 60°C for 15 s, and 72°C for 15 s. Melt curve analysis was used to confirm specific amplification of targets. Relative mRNA expression levels were calculated using the comparative Ct method, normalised to β-actin (*ACTB*).

Chromatin immunoprecipitation

Briefly, 4 x 10^7 cells were fixed using 1% formaldehyde (Sigma-Aldrich) for 10 min. Cells were washed in PBS and lysed using NP-40 lysis buffer. Cell nuclei were resuspended in 2 mL 0.4% SDS shearing buffer for sonication with a Covaris S220X sonicator (Covaris) for 7 min. For each immunoprecipitation, 25 µg chromatin was diluted with IP dilution buffer and pre-cleared with Protein A agarose beads (Merck-Millipore) for 1 h at 4°C. Chromatin was incubated with 5 µL anti-CTCF (Merck-Millipore), 5 µg anti-H3K27ac (abcam), or 5 µg rabbit IgG isotype control antibody (Merck-Millipore) for 16 h at 4°C with rotation. Immune complexes were collected by centrifugation and cleared using Protein A agarose beads (Millipore) and incubated for 1.5 h at 4°C. Complexes were washed and eluted in 500 µL ChIP elution buffer. Crosslinks were reversed by adding 25 µL 4M NaCl and incubation for 16 h at 65°C with shaking (600 rpm). Samples were treated with RNase A and Proteinase K, and DNA was purified using the QIAquick PCR Purification kit (QIAGEN) according to the manufacturer's specifications using 50 µL Buffer EB. For analysis, 2 µL of purified DNA was used for qPCR reactions with a Mic qPCR cycler as described above. Enrichment was determined using the percent input method.

Chromatin accessibility by real-time PCR

Chromatin accessibility by real-time PCR (ChART-PCR) was performed as previously described (*47*) using 20 U DNase I (Promega). To assess nucleosome occupancy, 1000 Gel Units MNase (New England Biolabs) was used. Digested and undigested samples were purified using the QIAquick PCR Purification kit (QIAGEN). For analysis, qPCR reactions consisting of 50 ng DNA, 1X SYBR Green (Bioline), 250 nM primers up to a final volume of 10 µL were cycled using a ViiA7 real-time thermocycler and QuantStudio V1.3 (Applied Biosystems). Cycling conditions were as follows: 95°C for 10 min, 40 cycles of: 95°C for 15 s, 60°C for 15 s, 72°C for 30 s, followed by melt curve analysis. Accessibility levels were determined using the comparative Ct method for undigested vs digested samples, normalised to the lung-specific *SFTPA2* promoter (SPA2-P) control locus. For MNase nucleosome occupancy assays, normalised data were transformed such that a value of 1.0

represents completely compacted nucleosomes, and lower values indicate reduced nucleosome occupancy.

CRISPR plasmids

All CRISPR plasmid constructs were modified from pSpCas9(BB)-2A-GFP (PX458), a gift from Feng Zhang (*48*) (Addgene plasmid #48138). To generate a large genomic deletion, PX458 was modified to express two guide RNAs (gRNAs) to cut the 5' and 3' ends of the target region. Guide RNAs (gRNAs) were designed using CRISPRscan (*49*) to select highest scoring sequences with minimal off-target effects (Supplementary Table 2). gRNAs were cloned into the *Bbs*I restriction sites of PX458 using T4 DNA ligase (New England Biolabs). gRNA expression cassette inserts (U6 RNA polymerase III, gRNA sequence and gRNA scaffold) were amplified using PCR with primers containing oligonucleotides with Acc65I and XbaI restriction ends for sub-cloning (Supplementary Table 1). For the negative control construct, the gRNA expression cassette of PX458 was removed using *Pci*I/*Xba*I digestion and purified using the QIAquick Gel Extraction kit (QIAGEN). The linearised plasmid was blunted using T4 DNA polymerase (New England Biolabs) and re-ligated.

CRISPR plasmid constructs (2 µg) were electroporated into 2 x 10^6 Raji cells using the Amaxa Cell Line Nucleofector Kit V (Lonza Bioscience) (Program M-013). Cells were incubated at 37°C with 5% CO_2 for 24 h and then sorted for GFP+ using fluorescent activated cell sorting (FACS) on a FACSAria II (BD Bioscience). The GFP+ pool was expanded for up to 3 weeks before extractions or flow cytometry. Genomic DNA and total RNA extraction using the QIAamp DNA Blood Mini kit (QIAGEN) and RNeasy Mini kit (QIAGEN), respectively. RNA was reverse-transcribed and transcript abundance was measured by qPCR as previously described.

CRISPR deletion screening

gDNA was qualitatively screened for genomic deletion using PCR with oligonucleotides amplifying across the targeted region (deletion; D) and within the target region (non-deletion; ND) (Supplementary Table 1). DNA (50 ng) was amplified using 1X GoTaq Green Master Mix (Promega), 0.8 µM oligonucleotides and 5% DMSO up to a volume of 20µL, and cycled as follows: 95°C for 5 min, followed by 30 cycles of 95°C for 30 s, 61°C for 18 s, 72°C for 5 s, and a final extension at 72°C for 5 min.

Flow cytometry

Cells (1 x 10^6) were harvested and washed with cold staining buffer (PBS with 5% FBS) at 300 x *g* for 5 min at 4°C. Cells were resuspended in 80 µL staining buffer and incubated with 20 µL of anti-human CD21-PE (BD Biosciences, Cat #555422) or IgG1κ-PE isotype control (BD Bioscience, Cat #555749) for 20 min. After incubation, cells were washed and resuspended in 0.5 mL staining buffer and analysed using a BD Accuri C6 flow cytometer (BD Bioscience) and FlowJo software V10.5.2 (Tree Star). Samples were run alongside unstained cells. Single-colour anti-mouse Simply Cellular® Compensation standard control (Bangs Laboratories Inc.) with IgG1-Alexa488 antibody (Invitrogen,

Lot 1208522A) was used to calculate a compensation matrix using FlowJo software V10.5.2 (Tree Star) and correct for spectral overlap in transfected GFP+ cells with PE-conjugated antibody staining.

Statistical analysis

Differences in transcriptional activity and mRNA expression were assessed using Student's unpaired t-test with a confidence interval of 95% ($p < 0.05$). Statistics and graphs were generated using GraphPad Prism version 7.0 (GraphPad). Graphed values represent the mean ± SEM of at least three independent experiments.

RESULTS

Chromatin interactions within the RCA gene cluster are organised into two TADs

To investigate the structural arrangement of the RCA gene cluster in B cells, we examined raw Hi-C data in the GM12878 B lymphoblastoid cell line from Rao *et al.* (*44*) at 10 kb resolution. The intergenic region between *CD55* and *CR2,* and loci across the complement receptor genes (*CR2* and *CR1*) engaged in highly frequent interactions with loci more than 400 kb upstream near *C4BPB* and *C4BPA* (Figure 1A). No notable interactions between these regions with downstream genes *CR1L* and *CD46* were observed, indicating that chromatin interactions in this region may be directionally constrained (Figure 1A). This pattern of interaction was consistent in 7 cell lines also examined at 10 kb resolution in this Hi-C dataset, including several non-B cell types (Supplementary Figure 1A). Interaction data was unavailable in all cell lines between exon 5 – 20 of *CR1* (Figure 1A, Supplementary Figure 1A) due to poor sequence mappability. Corresponding TAD co-ordinates from this dataset (*30*, *44*) varied between these cell lines, but RCA members *C4BPB, C4BPA, CD55, CR2* and *CR1* (exon 1 to exon 5) were consistently placed within the same predicted TAD (Supplementary Figure 1B). As TADs are recognised to be constitutive between cell-types (*30*), these data are consistent with the RCA gene cluster being organised into more than one domain.

To further interrogate the TAD organisation of the RCA gene cluster and precisely map long-range chromatin interactions in B cells, we performed 4C-seq at 30 bp resolution in an analogous B lymphoblastoid cell line (B-0028). As CTCF is known to play an important role in chromatin looping, we selected 4C viewpoints (VP) from CTCF sites utilising B cell ChIP-seq data (GM12878) which co-localised with regions of highly frequent interactions observed in the Hi-C data (Figure 1A and Figure 1B). These viewpoints included the intergenic region between *CD55* and *CR2* (VP1), intron 1 of *CR2* (VP2) and intron 1 of *CR1* (VP3) (Figure 1B). We also selected a 4C viewpoint from a CTCF binding site within intron 29 of *CR1* (VP4) which did not markedly engage in chromatin interactions with this upstream region (Figure 1A and Figure 1B). We confirmed enrichment of CTCF at these VPs in the B-0028 cell line using ChIP-qPCR (Supplementary Figure 2).

4C maps from VP1, VP2 and VP3 were highly similar and yielded consistent 4C signal peaks at upstream CTCF sites near RCA member *C4BPB* and within non-RCA member *YOD1* (Figure 1B and 1C, asterisks), corresponding to Hi-C data (Figure 1A). Common interactions from these CTCF viewpoints were also identified within intron 6 of *CD55* (Figure 1C, asterisks), and the intragenic CTCF viewpoints in *CR2* (VP2) and *CR1* (VP3) interacted with the intergenic CTCF viewpoint (VP1) (Figure 1C). Chromatin interactions from VP1 – 3 did not extend to CTCF sites upstream of *YOD1* or downstream of *CR1* exon 7 (Figure 1C). In contrast, VP4 produced a unique 4C map whereby interactions were constrained to approximately 60 kb downstream of this viewpoint within the *CR1* gene body and did not extend upstream (Figure 1C). We replicated 4C-seq from these CTCF viewpoints in another B cell line (B-0056) (Supplementary Figure 3), which produced a highly similar

interaction map whereby CTCF interactions were also organised to two discrete regions, potentially representing two TADs in the RCA gene cluster.

To further delineate the TAD arrangement of the RCA gene cluster, we examined recognised characteristics of TAD boundaries, including CTCF enrichment and binding site orientation, as well as sequence conservation (*50*). Regions flanking the RCA gene cluster (upstream of *PFKFB2* and downstream of *CD46*) and the region within *CR1* between 4C viewpoints VP3 and VP4 were enriched for CTCF binding (Figure 1B), and these all bound CTCF across multiple cell types (Supplementary Figure 4) thus representing potential constitutive TAD boundaries. The domains formed by these boundaries were flanked by CTCF sites underlined by convergent CTCF sequence motifs (Figure 1D). The CTCF sites flanking the RCA cluster (within *YOD1* and downstream of *CD46*) were contained at highly conserved regions in vertebrates (Figure 1D). Taken together, these data indicate that the RCA gene cluster is divided into two TADs and that a TAD boundary is located within *CR1* (Figure 1E).

Putative B cell enhancers in the RCA gene cluster were predicted to regulate multiple RCA genes

CTCF plays an important role in establishing long-range contacts within TADs and mediating enhancer-gene interactions (*29*). As CTCF-mediated chromatin looping was identified in the RCA gene cluster in B cells (Figure 1), we asked whether enhancer elements were present in this region and if they regulate expression of genes in the RCA cluster in this cell lineage. In the B cell lineage, only membrane-bound RCA members (*CD55, CR2, CR1, CD46)* are expressed in this cell lineage, not soluble protein members (*C4BPB* and *C4BPA*) (*5*, *11*). We confirmed this expression pattern using qPCR in the B-0028 cell line (Figure 1F). To note, *CR2* was the most highly expressed gene member in this B cell model, while *CR1* was the lowest (Figure 1F).

To identify putative enhancers in the RCA, we leveraged a novel enhancer database, GeneHancer, which integrates enhancer datasets from multiple consortium-based projects and other functional datasets to generate enhancer predictions and identify their potential gene-targets (*51*). Confidence scores for each enhancer prediction (GeneHancer score) and enhancer-gene prediction (gene-association score) were computationally assigned to each based on the evidence retrieved. Predicted enhancers were classified as “double elite” if both their GeneHancer and gene-association scores were derived from more than one source of data, representing a prediction which is more likely to be functional (*51*). Numerous predicted enhancers on GeneHancer were identified across TAD 1 and TAD 2, but only a subset of these were classified as “double elite” (Figure 2A, 2B, Supplementary Table 3).

Enhancers are important in cell-type specific regulation of gene expression and act by looping to their target gene promoters (*40*). To identify active enhancers that were most likely to be functional in B cells, we examined candidate enhancer for enrichment of epigenetic marks characteristic of

enhancers, such as H3K27ac and DNase I hypersensitivity (DHS). We located four strong candidate B cell enhancers (BENs) in TAD 1 that showed strong H3K27ac enrichment and DHS consistently in both B cell lines and primary B cells were identified (Figure 2C). These candidate enhancers were located within *CD55* (BEN-1) or the intergenic region between *CD55* and *CR2* (BEN-2, BEN-3 and BEN-4) (Table 1). Furthermore, each candidate BEN was bound by numerous transcription factors (based on ENCODE ChIP-seq data) including those important in B cell development, such as early B cell factor 1 (EBF1) (*52*) and PAX5 (*53*), and general regulatory factors (eg. EP300, CTCF and RNA polymerase II) (Table 1). To note, BEN-1 and BEN-3 were located at CTCF binding sites which were involved in chromatin looping as was identified in 4C-seq maps (Figure 1). The four BENs identified were supported by multiple lines of evidence to be active enhancer elements in B cells and were prioritised for further investigation.

Predicted enhancers on GeneHancer were assigned putative gene targets using multiple methods and datasets, including expression quantitative trait loci (eQTL) analysis, enhancer-promoter interactions generated by capture Hi-C in the GM12878 cell line (CHi-C) and eRNA-mRNA co-expression from the FANTOM5 Enhancer Atlas (*38*, *51*, *54*). Each candidate BEN was predicted to regulate multiple genes, including RCA genes (*C4BPA, CD55, CR2*, *CR1* and *CD46*) and non-RCA genes (*PIGR, FCAMR, C1orf116*) (Figure 2D, Supplementary Table 4). However, only interactions between BEN-1 and *CD55*, BEN-2 and *CR2,* and BEN-3 and *CR2*, represented high-confidence ("elite") associations; identified by more than one contrasting method (Figure 2E, Supplementary Table 4). Of these, only BEN-1 was predicted to regulate a gene (*CD46*) located downstream of the intragenic TAD boundary in *CR1* (Figure 2D). However, this predicted interaction had the lowest score among all gene-enhancer predictions for these BENs (Figure 2D, Supplementary Table 4). While these gene-enhancer interactions were based on bioinformatic predictions, this highlighted the potential for the RCA genes to be co-regulated in B cells.

Strong candidate B cell enhancers in the RCA gene cluster were functional *in vitro*

To test the functionality of each BEN, we performed luciferase reporter gene assays using a constitutive minimal promoter (SV40) to drive luciferase expression. Each BEN was cloned upstream of the SV40 promoter in both forward and reverse orientation and the transcriptional effects were assayed in a panel of B cell lines (Reh, Raji, B-0028, SKW) and a non-B cell control (HepG2, liver cell-type) (Figure 3). Interestingly, activity patterns of BEN-1 and BEN-3 were not consistent with that of an active enhancer, such that activity was unchanged or reduced relative to the control (pGL3-P, no enhancer) across the B cell lines, the latter indicative of silencer activity (Figure 3). BEN-4 displayed some enhancer activity in the B cell lines but the relative increase in transcriptional activity was only significant in the SKW cell line in the reverse orientation (p = 0.0368, n = 3). (Figure 3). In contrast, BEN-2 significantly increased luciferase activity by approximately 3-fold relative to the control in SKW, in both forward (p = 0.0219, n = 3) and reverse orientation (p = 0.0436, n = 3), and by 1.5-fold in Raji in the forward orientation (p = 0.0003, n = 4) (Figure 3). Notably, transcriptional activity was significantly decreased by 50% in the non-B cell line control (HepG2) in both enhancer

orientations (forward $p = 0.0321$, $n = 4$; reverse $p = 0.0255$, $n = 3$) (Figure 3). These data, taken together with bioinformatics, indicated that BEN-2 was the most likely candidate BEN to be active in the B cell lineage.

To support the functional role of BEN-2 in this cell type, we quantified H3K27ac enrichment and chromatin accessibility in the panel of cell lines used in the luciferase assays. Nucleosome occupancy was consistently lower in the B cell lines than in the non-B cell control, HepG2 (Figure 4A). Conversely, chromatin accessibility was consistently high across all B cell lines, but low (inaccessible) in HepG2 (Figure 4A and Figure 4B), indicating that this region is transcriptionally active in the B cell lineage. Accordingly, H3K27ac enrichment at BEN-2 was not observed in HepG2 but enriched in all B cell lines (Figure 4C). Altogether, these data are in support of BEN-2 acting as a functional B cell enhancer *in vitro.*

CRISPR deletion of an intergenic B cell enhancer (BEN-2) decreased *CR2* and *CD55* expression

As reporter gene assays remove regulatory elements from their genomic context, which is an important aspect of enhancer function, we sought to assess the functional activity of BEN-2 *in vivo*. We also wished to confirm the predicted gene targets of BEN-2 identified on GeneHancer, including *CD55* and *CR2* which directly flank the enhancer (Figure 2). CRISPR deletion machinery was delivered using a plasmid-based method into the Raji mature B cell line, which expresses *CD55, CR2* and *CD46,* but *CR1* is not expressed at levels detectable by qPCR (Figure 5A). This pattern of gene expression is in accordance with other B cell lines, such as B-0028 (Figure 1F). We modified the PX458 CRISPR plasmid to express two guide RNA sequences (gRNA) that cut either side of BEN-2 (g1g2 or g3g4) to delete a 4 kb region encompassing the enhancer (Figure 5B). The CRISPR plasmids, containing a GFP marker, was delivered into Raji cells and successfully transfected GFP-positive cells were enriched by fluorescence activated cell sorting (FACS). Successful enhancer deletion within the enriched cell populations was confirmed by PCR (Figure 5B and Figure 5C).

CRISPR deletion of BEN-2 with g1g2 significantly decreased *CR2* transcript abundance to approximately 30% of WT levels ($p = 0.0184$, $n = 3$) and *CD55* to approximately 50% of WT levels ($p = 0.0167$, $n = 3$) (Figure 5C). A no-guide construct expressing only Cas9 and GFP did not result in enhancer deletion (Figure 5B) or significantly alter mRNA expression of these genes in wild-type Raji cells (WT) (Figure 5C). To account for potential off-target effects with g1g2, we repeated the deletion using an additional pair of guide sequences (g3g4). In agreement with altered transcript abundance from deletion of BEN-2 with g1g2, expression of *CR2* and *CD55* were both decreased in the enriched g3g4 population in comparison to WT. The decrease in *CR2* expression was less extensive with g3g4 (approximately 60% of WT levels), but this was still significant compared to the NG control ($p = 0.0213$, $n = 3$). Expression of *CD55* was significantly reduced to similar levels in both g1g2 and g3g4 enriched population (g3g4 $p = 0.0281$, $n = 3$). We also measured transcript abundance of *CD46,* which was not predicted to be targeted by BEN-2 on GeneHancer and located to the neighbouring

373 TAD (Figure 1). Enhancer deletion with either guide combination did not alter *CD46* transcript
374 abundance (Figure 5C). These data confirmed that BEN-2 is a functional enhancer in B cells and
375 functionally demonstrate that it regulates *CD55* and *CR2* within this cellular context.
376
377 CR2 (CD21) is a well-established marker for B cell maturity (*55*) and surface expression of CR2 is
378 highly robust in the Raji B cell line (*56*); representative of a mature B cell (*57*). As *CR2* transcript
379 levels were significantly decreased with BEN-2 CRISPR deletion in the Raji cell line, we next
380 determined if CR2 surface protein expression were concomitantly affected. We used flow cytometry to
381 assess CR2 surface expression in the polyclonal FACS-enriched populations used and observed
382 decreases that correlated well with corresponding transcript decreases (Figure 6). A decrease in CR2
383 expression was observed in g1g2 populations relative to WT across three biological replicates (Figure
384 6A). To quantify the decrease in CR2 surface expression more definitively, we gated the WT
385 population based on side scatter (SSC-A) *versus* CR2-PE into four quadrants such that the CR2-
386 positive cells were equally divided into Q3 and Q4, and then applied this gate to the CRISPR FACS-
387 enriched populations. We observed a reduction in CR2-positive cells in Q3 from 53% to 39.7%, with a
388 concomitant increase in cells in Q4 (WT Q4 = 47.0%, g1g2 Q4 = 60.3%) (Figure 6B). In line with *CR2*
389 transcript abundance data (Figure 5C), CR2 surface expression was unchanged in the NG
390 populations and decreased in the g3g4 population relative to WT (Supplementary Figure 5). This
391 confirmed that the reduction in *CR2* transcript expression with CRISPR deletion of BEN-2 also
392 reduced surface protein levels, and thus further validates that BEN-2 regulates CR2 both at the level
393 of mRNA and subsequent protein receptor expression in a B cell context.
394

DISCUSSION

Here, we have explored for the first time how chromatin architecture governs the RCA gene cluster and demonstrated that its members are transcriptionally co-regulated by common long-range mechanisms. Using high-resolution 4C-seq maps in B cell models, we showed that several distal CTCF sites in the RCA gene cluster interact with one another, representing physical chromatin looping and defining a functional role for these CTCF sites in this cell type. This includes the CTCF site modulated by SLE-associated SNP (rs1876453) (*27*), revealing an extensive and complex mechanism by which this genetic variant may influence gene expression and autoimmunity. CTCF-mediated chromatin looping was constrained to one of two distinct regions despite the broad distribution of CTCF sites across the entire gene cluster, illustrating that the RCA spans two adjacent TADs in B cells and separating the members of the human RCA gene cluster into TAD 1 (*C4BPB, C4BPA, CD55, CR2* and *CR1*) or TAD 2 (*CR1L* and *CD46*).

Unexpectedly, the interactions of the two CTCF sites within *CR1* diverged to in TAD 1 or TAD 2, partitioning the *CR1* promoter to TAD 1 and indicating that the boundary between the two TADs in the RCA gene cluster is positioned within the body of the *CR1* gene. We also identified convergent CTCF sites flanking each domain, supporting the positioning of this TAD boundary in line with loop extrusion model of TAD formation (*31*). To our knowledge, a TAD boundary distinctly located within the body of an expressed protein-coding gene has not been characterised before and the question remains as to if this novel TAD boundary directly influences *CR1* expression itself.

This boundary element was not investigated further in this study as Hi-C and 4C interaction data were missing within *CR1* in the vicinity of this boundary. This is due to a well-characterised tandem segmental duplication within *CR1* known as ‘low copy repeat 1’ (LCR1) which results in different structural alleles of *CR1* (*58*, *59*); CR1-A (two copies of LCR1) and CR1-B (one copy of LCR1) alleles are most common (*60*). As LCR1 is 18 kb in size and possesses more than 99% sequence identity, it is not amenable to sequencing-based technologies, including 4C and Hi-C. Missing Hi-C interaction data has been shown to influence computational TAD calls (*61*). Thus the missing data at LCR1 likely explains why the predicted TADs from Dixon *et al.* (*30*), differed to the TAD organisation we uncovered. Nonetheless, both these data are complimentary. Interestingly, LCR1 has had a long-standing genetic association with SLE (*62*), but the impact of this repeat on the biological function of CR1 and how it contributes to disease susceptibility has remained unknown. Our findings highlight the importance of validating TAD predictions from Hi-C using complimentary methods (*39*), and propose that the effect of different copy numbers of LCR1 may influence the topological architecture of RCA gene cluster in the B cell lineage.

The *CR1* boundary element and TAD organisation of the may be pertinent from an evolutionary perspective. In mice, the RCA gene cluster is conserved but separated to two chromosomal positions located more than 6 Mb apart (*63*), matching closely to the TAD organisation of the gene cluster we

identified. Notably, there are some key differences between RCA members in humans and mice. In humans, CR2 and CR1 are encoded by separate genes, while mouse CR2 and CR1 are expressed from a single gene, *Cr2*, through alternative splicing (*64*). It has been observed that breaks in synteny between species commonly occur at TAD boundaries (*65*), thus the TAD boundary we identified in *CR1* may represent the breakpoint region for the genomic rearrangement of the RCA or complement receptor genes in humans and mice.

Importantly, the chromatin organisation of the RCA gene cluster we uncovered indicated that there are two distinct regulatory landscapes which control its members in B cells. We observed direct physical interactions between several expressed gene members in TAD 1, including *CD55* and *CR2*, forming a strong basis for these genes to be regulated by enhancers and transcriptionally co-regulated in this cell type. Indeed, we able to establish this through the identification of a functional intergenic enhancer, BEN-2, in TAD 1. We showed that BEN-2 is an active enhancer in B cells by characterising its transcriptional activity and chromatin marks in several B cell models. Genomic deletion of BEN-2 in the Raji B cell line resulted in significant decreases in *CR2* and *CD55* mRNA expression, which confirmed that BEN-2 is functional enhancer in B cells and that it regulates the expression of both these genes. Additionally, deletion of this enhancer did not influence the RCA gene in TAD 2 (*CD46*), indicating that the TADs in the RCA cluster function to restrict enhancers to target genes within the same TAD. It is possible that BEN-2 also regulates *CR1* in TAD 1, as predicted by CHi-C in the GM12878 B cell line (*54*). Altogether, we showed that chromatin looping and TADs in the RCA gene cluster functions, in part, to bring long-range elements to gene promoters and established for the first time that the RCA cluster is transcriptionally co-regulated.

This finding has a number of potential implications in B cell biology and autoimmunity. In addition to improving our understanding of *CR2* and *CD55* transcriptional regulation, we showed that deletion of BEN-2 reduced and sustained *CR2* mRNA expression to levels which also diminished CR2 protein levels at the cell surface. As receptor expression of CR2 on B cells is directly involved in modulating antibody production (*15*), this finding highlights an important role for BEN-2 in the control of CR2 expression and the potential for altered BEN-2 enhancer activity to influence B cells at the functional level. By revealing that BEN-2 regulates the transcription of these genes in B cells simultaneously, we have established a clear gene-gene interaction between *CR2* and *CD55*. The expression of CD55 on B cells has been also shown to be altered in SLE patients (*66*, *67*), as has been similarly demonstrated for CR2 (*18*, *19*). As these regulators share key functions in complement and both have known roles in SLE, alterations to the expression of both genes through BEN-2 or chromatin looping could considerably impair immune response and B cell autoimmunity.

Our findings also define an important non-coding region in the RCA gene cluster situated in TAD 1 between *CD55* and *CR2*. This region encompasses BEN-2, as well as also several strong candidate regulatory elements. The transcriptional activity of all candidate enhancers we assessed in this study deviated from the control in at least one B cell line in luciferase assays, indicative of regulatory

potential in this cell type. Intriguingly, two of the candidate enhancers (BEN-1 and BEN-3) demonstrated a pattern of regulatory activity in B cells consistent with that of a silencer. To note, these candidates were the only two indicated to bind CTCF; a known insulator of gene expression (*29*). Given that the majority of disease-associated variants map to non-coding regulatory regions (*36*), it will be of interest to investigate genetic variants within this region to identify potential associations with SLE susceptibility.

Among BEN-2 and the candidate enhancers in TAD 1, only one predicted enhancer-gene interaction out of 16 crossed into TAD 2 (BEN-1 and *CD46*), emphasising the partition between the two domains of the RCA gene cluster. While these results are only based on bioinformatic data, this reveals the potential extent to which transcriptional co-regulation may occurs in the RCA. This is of particular interest as the RCA members are expressed on cell types other than B cells, and have unique expression patterns individually. For example, *CD55* and *CD46* are expressed across many cell types (*5*, *11*), whereas *CR2* and *CR1* are predominantly expressed on B cells (*68*) and erythrocytes (*69*), respectively. Investigation of the chromatin architecture and long-range regulation of the RCA gene cluster in other immune cell types will uncover the extent to which these genes are transcriptional co-regulated, the long-range mechanisms which control RCA genes and how this is dysregulated in the context of autoimmunity.

In conclusion, we have shown that the RCA gene cluster comprises of a complex network of enhancer-gene and gene-gene interactions (epistasis) with a distinct three-dimensional genome structure in the B cell lineage. Indeed, epistasis has been a long-recognised contributor to complex disease susceptibility but is often unaddressed, particularly due to the difficulty in identifying such interactions (*70*). We used chromatin maps and functional analyses to define such interactions at the molecular level in the RCA gene cluster and identify novel co-regulatory mechanisms which collectively govern these key immunomodulatory genes. In sum, we have revealed novel mechanisms by which the RCA gene cluster is controlled, expanding the scope for future investigations in the context of evolution, immunity and complex genetic disease.

ACCESSION NUMBERS

4C-seq data were deposited in the Gene Expression Omnibus (GEO) database under accession number GSE140127.

ACKNOWLEDGEMENTS

We would like to thank Kevin Li and the FACS Facility at the Harry Perkins Institute of Medical Research for technical assistance with cell sorting. Further, we thank Kathy Fuller and Henry Hui for their technical expertise and reagents for flow cytometry, and Susan Boackle for the lymphoblastoid cell lines (B-0028 and B-0056).

FUNDING

This work was supported by the National Institutes of Health [R01 AI24717 to J.B.H.], the Australian Government Research Training Program Scholarship at the University of Western Australia [to J.C. and J.S.C.], the Spanish Government [BFU2016-74961-P to J.S.G-.S] and an institutional grant Unidad de Excelencia María de Maeztu [MDM-206-0687 to the Department of Gene Regulation and Morphogenesis, Centro Andaluz de Biología del Desarrollo].

CONTRIBUTIONS

J.C., J.S.C. and R.D.A. performed the experiments; J.C., J.S.C., J.L.G.-S., R.L.T. and D.U. designed the experiments; J.C., J.S.C., E.Q. and D.U. analysed results; J.S.C. and D.U. conceptualised the project; J.L.G.-S., R.L.T. and J.B.H. provided intellectual input and resources; E.Q. and D.U. supervised the project; J.C. wrote the manuscript draft with input from J.S.C., R.L.T., E.Q. and D.U.. All authors reviewed the final manuscript.

CONFLICT OF INTEREST

The authors declare no competing interests.

REFERENCES

1. V. M. Holers, Complement and Its Receptors: New Insights into Human Disease. *Annu. Rev. Immunol.* **32**, 433–459 (2014).
2. M. C. Carroll, D. E. Isenman, Regulation of Humoral Immunity by Complement. *Immunity*. **37**, 199–207 (2012).
3. E. E. West, M. Kolev, C. Kemper, Complement and the Regulation of T Cell Responses. *Annu. Rev. Immunol.* **36**, 309–338 (2018).
4. D. Ricklin, E. S. Reis, J. D. Lambris, Complement in disease: a defence system turning offensive. *Nat. Rev. Nephrol.* **12**, 383–401 (2016).
5. D. Hourcade, V. M. Holers, J. P. Atkinson, The regulators of complement activation (RCA) gene cluster. *Adv. Immunol.* **45**, 381–416 (1989).
6. D. Hourcade *et al.*, Analysis of the human regulators of complement activation (RCA) gene cluster with yeast artificial chromosomes (YACs). *Genomics*. **12**, 289–300 (1992).
7. F. Pardo-Manuel de Villena, S. Rodríguez de Córdoba, C4BPAL2: a second duplication of the C4BPA gene in the human RCA gene cluster. *Immunogenetics*. **41**, 139–143 (1995).
8. C. M. Logar, W. Chen, H. Schmitt, C. Yung, D. J. Birmingham, A human CR1-like transcript containing sequence for a binding protein for iC4 is expressed in hematopoietic and fetal lymphoid tissue. **40**, 831–840 (2004).
9. D. Ermert, A. M. Blom, C4b-binding protein: The good, the bad and the deadly. Novel functions of an old friend. *Immunol. Lett.* **169**, 82–92 (2016).
10. S. H. Dho, J. C. Lim, L. K. Kim, Beyond the Role of CD55 as a Complement Component. *Immune Netw.* **18**, 1–13 (2018).
11. P. F. Zipfel, C. Skerka, Complement regulators and inhibitory proteins. *Nat. Rev. Immunol.* **9**, 729–740 (2009).
12. M. K. Liszewski, J. P. Atkinson, Complement regulator CD46: genetic variants and disease associations. *Hum. Genomics*. **9**, 7 (2015).
13. A. Erdei *et al.*, Expression and role of CR1 and CR2 on B and T lymphocytes under physiological and autoimmune conditions. *Mol. Immunol.* **46**, 2767–2773 (2009).
14. Y. Deng, B. P. Tsao, Genetic susceptibility to systemic lupus erythematosus in the genomic era. *Nat. Rev. Rheumatol.* **6**, 683–92 (2010).
15. P. W. Dempsey *et al.*, C3d of complement as a molecular adjuvant: bridging innate and acquired immunity. *Science*. **271**, 348–350 (1996).
16. B. S. Andrews *et al.*, Spontaneous murine lupus-like syndromes. Clinical and immunopathological manifestations in several strains. *J. Exp. Med.* **148**, 1198–1215 (1978).
17. A. N. Theofilopoulos, F. J. Dixon, Murine models of systemic lupus erythematosus. *Adv. Immunol.* **37**, 269–390 (1985).
18. J. G. Wilson, W. D. Ratnoff, P. H. Schur, D. T. Fearon, 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. *Arthritis Rheum.* **29**, 739–747 (1986).
19. H. V Marquart *et al.*, Complement receptor expression and activation of the complement cascade on B lymphocytes from patients with systemic lupus erythematosus (SLE). *Clin. Exp. Immunol.* **101**, 60–65 (1995).
20. J. M. Ahearn *et al.*, Disruption of the Cr2 locus results in a reduction in B-1a cells and in an impaired B cell response to T-dependent antigen. *Immunity*. **4**, 251–262 (1996).
21. H. Molina *et al.*, Markedly impaired humoral immune response in mice deficient in complement receptors 1 and 2. *Proc. Natl. Acad. Sci.* **93**, 3357–3361 (1996).
22. K. J. Marchbank, L. Kulik, M. G. Gipson, B. P. Morgan, V. M. Holers, Expression of human complement receptor type 2 (CD21) in mice during early B cell development results in a reduction in mature B cells and hypogammaglobulinemia. *J. Immunol.* **169**, 3526–3535 (2002).
23. L. Morel *et al.*, Genetic reconstitution of systemic lupus erythematosus immunopathology with polygenic murine strains. *Natl. Acad. Sci.* **97**, 6670–6675 (2000).
24. S. A. Boackle *et al.*, Cr2, a candidate gene in the murine Sle1c lupus susceptibility locus, encodes a dysfunctional protein. *Immunity*. **15**, 775–785 (2001).
25. H. Wu *et al.*, Association of a common complement receptor 2 haplotype with increased risk of systemic lupus erythematosus. *Natl. Acad. Sci.* **104**, 3961–3966 (2007).
26. K. B. Douglas *et al.*, Complement receptor 2 polymorphisms associated with systemic lupus erythematosus modulate alternative splicing. *Genes Immun.* **10**, 457–469 (2009).
27. J. Zhao *et al.*, Preferential association of a functional variant in complement receptor 2 with antibodies to double-stranded DNA. *Ann. Rheum. Dis.* **75**, 242–252 (2016).

28. M. Herold, M. Bartkuhn, R. Renkawitz, CTCF: insights into insulator function during development. *Development*. **139**, 1045–1057 (2012).
29. C.-T. Ong, V. G. Corces, CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* **15**, 234–246 (2014).
30. J. R. Dixon *et al.*, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. **485**, 376–380 (2012).
31. G. Fudenberg *et al.*, Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
32. J. R. Dixon, D. U. Gorkin, B. Ren, Chromatin Domains: The Unit of Chromosome Organization. *Mol. Cell*. **62**, 668–680 (2016).
33. E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. **326**, 289–293 (2009).
34. G. Andrey, S. Mundlos, The three-dimensional genome: regulating gene expression during pluripotency and development. *Development*. **144**, 3646–3658 (2017).
35. D. Shlyueva, G. Stampfel, A. Stark, Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
36. M. T. Maurano *et al.*, Systematic localization of common disease-associated variation in regulatory DNA. *Science*. **337**, 1190–1195 (2012).
37. M. P. Creyghton *et al.*, Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–21936 (2010).
38. R. Andersson *et al.*, An atlas of active enhancers across human cell types and tissues. *Nature*. **507**, 455–461 (2014).
39. M. A. Marti-Renom *et al.*, Challenges and guidelines toward 4D nucleome data and model standards. *Nat. Genet.* **50**, 1352–1358 (2018).
40. S. Heinz, C. E. Romanoski, C. Benner, C. K. Glass, The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* **16**, 144–154 (2015).
41. D. Kleftogiannis, P. Kalnis, V. B. Bajic, Progress and challenges in bioinformatics approaches for enhancer identification. *Brief. Bioinform.* **17**, 967–979 (2016).
42. A. Fernández-Miñán, J. Bessa, J. J. Tena, J. L. Gómez-Skarmeta, in *Methods in Cell Biology*, H. William Detrich, M. Westerfield, L. I. B. T.-M. in C. B. Zon, Eds. (Academic Press, 2016; http://www.sciencedirect.com/science/article/pii/S0091679X16000200), vol. 135, pp. 413–430.
43. E. Splinter, E. de Wit, H. J. G. van de Werken, P. Klous, W. de Laat, Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: From fixation to computation. *Methods*. **58**, 221–230 (2012).
44. S. S. P. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. **159**, 1665–1680 (2014).
45. Y. Wang *et al.*, The 3D Genome Browser: A web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.* **19**, 1–12 (2018).
46. J. D. Ziebarth, A. Bhattacharya, Y. Cui, CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. *Nucleic Acids Res.* **41**, 188–194 (2013).
47. M. Cruickshank, E. Fenwick, L. J. Abraham, D. Ulgiati, Quantitative differences in chromatin accessibility across regulatory regions can be directly compared in distinct cell-types. *Biochem. Biophys. Res. Commun.* **367**, 349–355 (2008).
48. F. A. Ran *et al.*, Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
49. M. A. Moreno-Mateos *et al.*, CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods*. **12**, 982–8 (2015).
50. M. Franke, J. L. Gómez-Skarmeta, An evolutionary perspective of regulatory landscape dynamics in development and disease. *Curr. Opin. Cell Biol.* **55**, 24–29 (2018).
51. S. Fishilevich *et al.*, GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database*, 1–17 (2017).
52. M. Busslinger, Transcriptional Control of Early B Cell Development. *Annu. Rev. Immunol.* **22**, 55–79 (2004).
53. C. Cobaleda, A. Schebesta, A. Delogu, M. Busslinger, Pax5: The guardian of B cell identity and function. *Nat. Immunol.* **8**, 463–470 (2007).
54. B. Mifsud *et al.*, Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
55. T. F. Tedder, L. T. Clement, M. D. Cooper, Expression of C3d receptors during human B cell differentiation: immunofluorescence analysis with the HB-5 monoclonal antibody. *J. Immunol.* **133**, 678–683 (1984).

56. R. L. Taylor *et al.*, Focused transcription from the human CR2 / CD21 core promoter is regulated by synergistic activity of TATA and Initiator elements in mature B cells. *Cell. Mol. Immunol.* **13**, 119–131 (2016).
57. D. Benjamin *et al.*, Immunoglobulin secretion by cell lines derived from African and American undifferentiated lymphomas of Burkitt's and non-Burkitt's type. *J. Immunol.* **129**, 1336–1342 (1982).
58. H. Crehan *et al.*, Complement receptor 1 (CR1) and Alzheimer's disease. *Immunobiology*. **217**, 244–250 (2012).
59. W. W. Wong, Structural and functional correlation of the human complement receptor type 1. *J. Invest. Dermatol.* **94**, 64S-67S (1990).
60. J. M. Moulds, J. D. Reveille, F. C. A. Division, C. Immunogenetics, Structural polymorphisms of complement receptor 1 (CR1) in systemic lupus erythematosus (SLE) patients and normal controls of three ethnic groups. *Clin. Exp. Immunol.* **1**, 302–305 (1996).
61. Y. Zheng, F. Ay, S. Keles, Generative modeling of multi-mapping reads with mHi-C advances analysis of Hi-C studies. *Elife*. **8** (2019), doi:10.7554/elife.38070.
62. S. K. Nath, J. B. Harley, Y. H. Lee, Polymorphisms of complement receptor 1 and interleukin-10 genes and systemic lupus erythematosus: a meta-analysis. *Hum. Genet.* **118**, 225–234 (2005).
63. B. Y. S. F. Kingsmore *et al.*, Receptor-Related Genes in the Mouse. *J. Exp. Med.* **169**, 1479–1484 (1989).
64. H. Molina, T. Kinoshita, K. Inoue, J. C. Carel, V. M. Holers, A molecular and immunochemical characterization of mouse CR2. Evidence for a single gene model of mouse complement receptors 1 and 2. *J. Immunol.* **145**, 2974–83 (1990).
65. Q. Szabo, F. Bantignies, G. Cavalli, Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668 (2019).
66. A. P. Alegretti *et al.*, Expression of CD55 and CD59 on peripheral blood cells from systemic lupus erythematosus (SLE) patients. *Cell. Immunol.* **265**, 127–132 (2010).
67. I. García-Valladares *et al.*, Diminished expression of complement regulatory proteins (CD55 and CD59) in lymphocytes from systemic lupus erythematosus patients with lymphopenia. *Lupus*. **15**, 600–5 (2006).
68. J. J. Weis, T. F. Tedder, D. T. Fearon, Identification of a 145,000 Mr membrane protein as the C3d receptor (CR2) of human B lymphocytes. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 881–885 (1984).
69. M. Pascual, J. A. Schifferli, The binding of immune complexes by the erythrocyte complement receptor 1 (CR1). *Immunopharmacology*. **24**, 101–106 (1992).
70. T. A. Manolio *et al.*, Finding the missing heritability of complex diseases. *Nature*. **461**, 747–53 (2009).

FIGURES AND TABLES

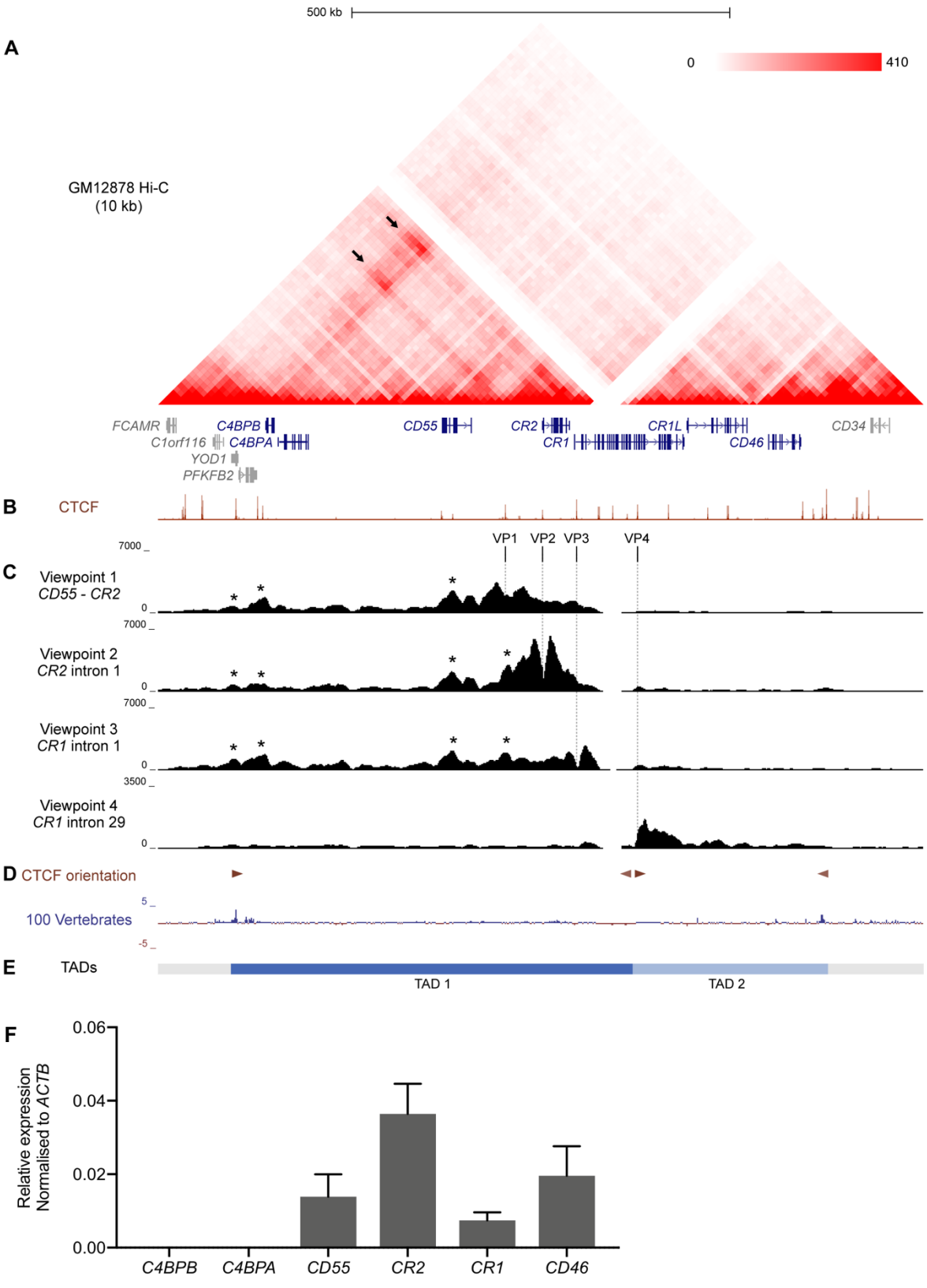


Figure 1: Chromatin conformation and TAD organisation of the Regulators of Complement Activation (RCA) gene cluster in B cells.

A. Hi-C heatmap matrix (10 kb resolution) for the GM12878 B cell line from Rao *et al*. (*44*) for the 1 Mb region across the RCA genes (dark blue) on hg19 (chr1:207,120,000-208,130,000). Relative interaction frequencies between two loci are indicated by colour intensity (range 0-410). High frequency long-range interactions (> 350 kb) were observed between distal RCA genes *C4BPB* and the complement receptor genes (*CR2* and *CR1*) (arrows), as well as between intervening loci, indicating that these genes reside in the same TAD.
B. GM12878 ChIP-seq signal for CTCF from ENCODE shows CTCF enrichment at multiple sites across the RCA gene cluster which may engage in long-range chromatin looping.
C. Chromatin conformation of the RCA gene cluster was fine-mapped using 4C-seq in the B-0028 cell line. Maps were generated from four viewpoints on CTCF binding sites in the intergenic region between *CR2* and *CD55* (viewpoint 1), intron 1 of *CR2* (viewpoint 2), the intron 1 of *CR1* (viewpoint 3) and intron 29 of CR1 (viewpoint 4). Viewpoints are represented by vertical dotted lines. Several 4C-seq peaks were common between VP1 – 3 and aligned with CTCF binding sites within *YOD1,* upstream of *C4BPB* and within *CD55* (asterisks). VP4 showed a distinct interaction profile to all other viewpoints.
D. The RCA TADs were bordered by convergent CTCF sites (triangles, reverse and forward orientation). Sequence flanking the RCA gene cluster was indicated to be conserved across vertebrates, which was examined using the 100 Vertebrates Basewise Conservation by phyloP track on UCSC with the default settings. Conserved loci are assigned positive scores (blue), while non-conserved loci are assigned negative scores (red).
E. TAD boundaries were determined collectively from the above data (Hi-C, 4C-seq, CTCF binding sites and evolutionary conservation). We propose that the RCA gene cluster is divided into two TADs in B cells. An intragenic TAD boundary is located at *CR1* and separates RCA genes into TAD 1 (*C4BPB, C4BPA, CD55, CR2* and *CR1*) and TAD 2 (*CR1L* and *CD46*).
F. Transcript abundance of RCA genes was measured by qPCR. Values were normalised to the β-actin gene (*ACTB*) using the ΔΔCt method. Bars represent mean relative expression ± SEM from 3 biological replicates.

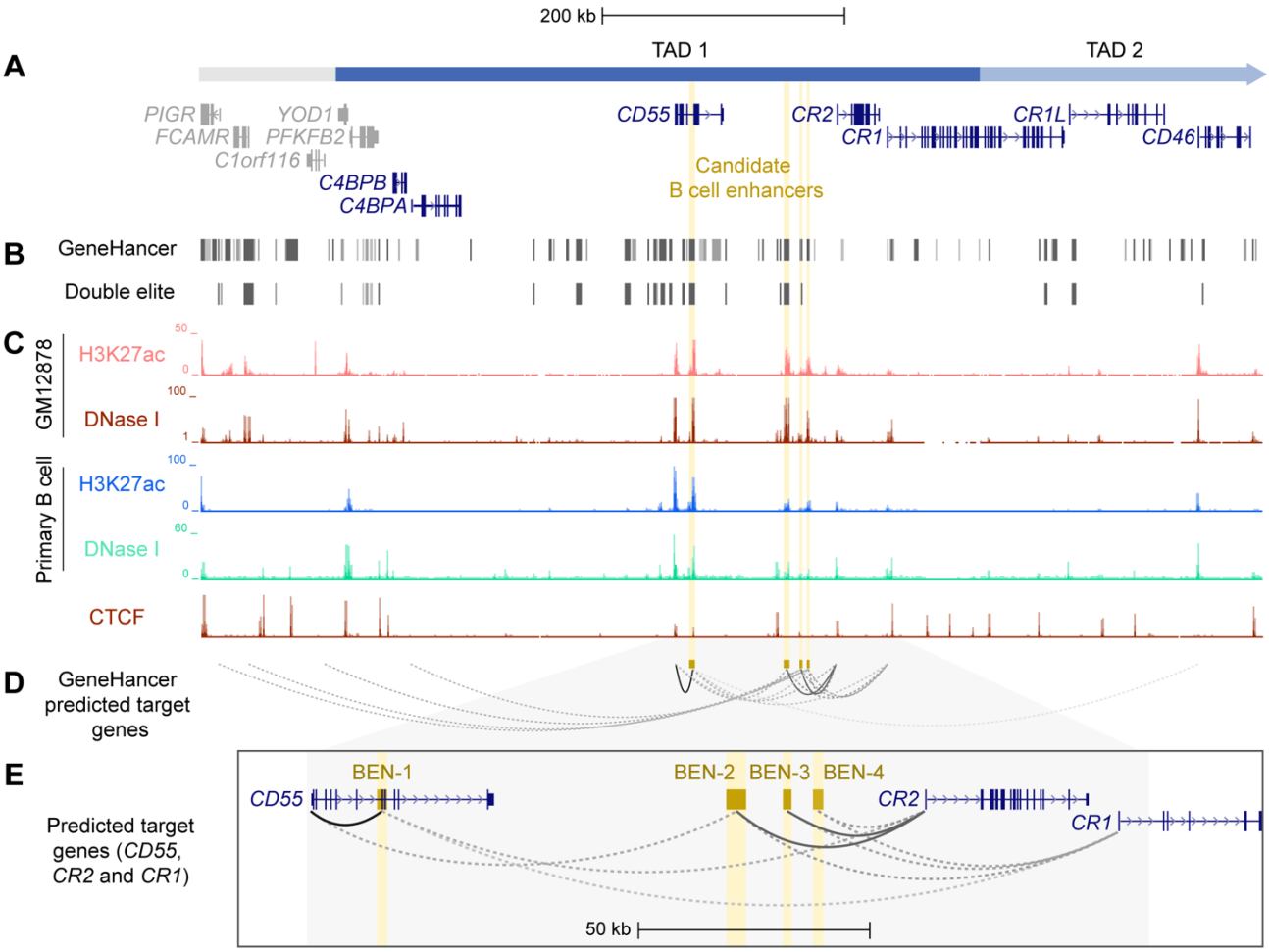


Figure 2: Identification and prioritisation of candidate B cell enhancers in TAD 1.

A. The structural organisation of the RCA genes (dark blue) and upstream genes (grey) *PIGR, FCAMR, C1orf116* and *YOD1* on hg19 (chr1:207,104,491-207,978,031).

B. Putative enhancers were identified by GeneHancer from multiple datasets from different consortia, such as ENCODE, Ensembl and FANTOM5 (GeneHancer). Each putative enhancer was also assigned predicted gene targets based on one or more methods. However, only a subset of putative enhancers were classified as 'double elite' on GeneHancer (Double elite).

C. Four candidate B cell enhancers (yellow) were identified using ENCODE data for H3K27ac enrichment and DNase I hypersensitivity in different B cell samples (GM12878 B cell line and primary B cells from peripheral blood). CTCF ChIP-seq signal for the GM12878 B cell line from ENCODE is also shown.

D. Candidate B cell enhancers were predicted to regulate multiple genes. Target gene predictions that were identified by more than one method in GeneHancer are represented by a solid line. Predictions that were identified by just one method are represented by a dotted line. The opacity of each line represents the relative score/confidence for each gene-enhancer prediction as determined by GeneHancer whereby higher confidence predictions are darker.

E. Region across *CD55, CR2* and *CR1* (exon 1 – 6) on hg19 (chr1:207,484,047-207,700,935). Candidate B cell enhancers (BEN) were named based in order of chromosomal position (BEN-1, BEN-2, BEN-3 and BEN-4). Evidence for BENs to regulate *CD55, CR2* and *CR1* was strongest among all gene-enhancer predictions.

Table 1: Strong candidate B cell enhancers (BENs) on GeneHancer were identified from multiple enhancer databases, contained multiple bind multiple transcription factor binding sites (TFBS) and were predicted to regulate multiple genes.

	BEN-1	BEN-2	BEN-3	BEN-4
GeneHancer ID	GH01J207333	GH01J207411	GH01J207424	GH01J207429
Location	*CD55* (exon 5 – 8)	Intergenic (*CD55* – *CR2*)	Intergenic (*CD55* – *CR2*)	Intergenic (*CD55* – *CR2*)
Enhancer sources	FANTOM5 ENCODE Ensembl dbSUPER	FANTOM5 ENCODE Ensembl	FANTOM5 ENCODE Ensembl	ENCODE
GeneHancer score	1.6*	1.2*	1.1*	0.7
TFBS ^	64 EBF1 IRF4 PAX5 RELA SPI1 CTCF EP300 POLR2A	35 EBF1 IRF4 PAX5 RELA EP300 POLR2A	13 EBF1 RELA SPI1 CTCF POLR2A	22 EBF1 RELA SPI1
Predicted gene targets	*CD55* CR2 CR1 CD46*	*CR2* CR1 CD55*	*CR2* CR1 C1orf116 C4BPA*	*CR2 CR1 CD55 FCAMR PIGR*

^ Only transcription factors important in B cell development (EBF1, IRF4, PAX5, RELA, SPI1) and gene regulation or chromatin organisation (CTCF, POL2RA and EP300) assayed using ChIP-seq in ENCODE are listed.

* These predicted enhancers and gene-enhancer interactions were identified using more than one method by GeneHancer.

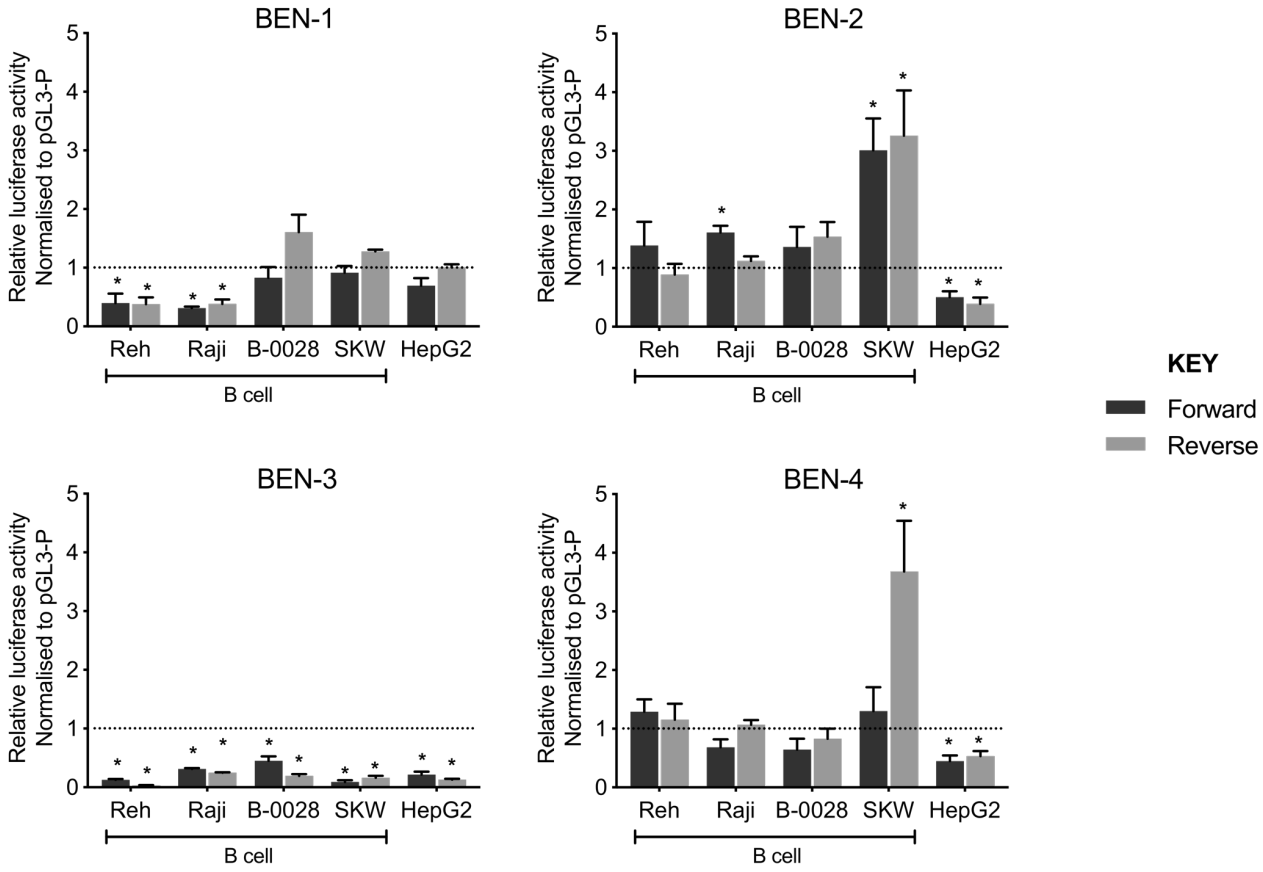


Figure 3: Candidate B cell enhancers demonstrated regulatory potential in luciferase assays, and BEN-2 increased relative transcriptional activity across a panel of B cell lines.
Enhancer constructs for strong candidate B cell enhancers were cloned into the pGL3-P (Promega) luciferase plasmid, upstream of an SV40 minimal promoter in forward (blue) and reverse (orange) orientation. Bars represent mean relative luciferase activity ± SEM after normalisation to an empty pGL3-P (no enhancer) control plasmid (n = 3 to 8). Asterisks represent statistically significant differences between normalised values and the pGL3-P control ($p < 0.05$). Dotted line at $y = 1$ represents normalised pGL3-P control value.

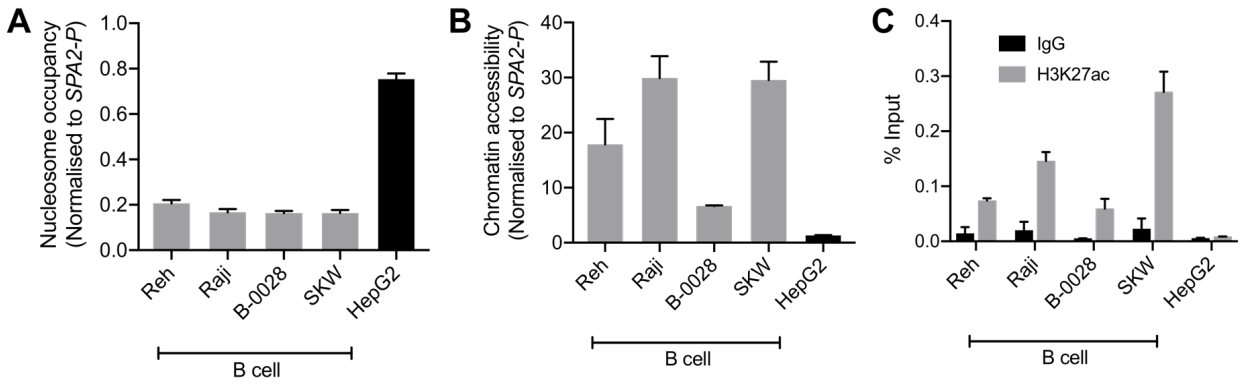


Figure 4: BEN-2 shows B cell-specific nucleosome occupancy, chromatin accessibility and enrichment for the H3K27ac active enhancer histone mark across a panel of B cell lines.

A. Nucleosome occupancy at BEN-2 as measured by ChART-PCR with MNase digestion. Data was normalised to the inaccessible *SFTPA2* gene promoter such that a value of 1.0 represents fully compacted nucleosomes, and lower values indicate less compacted nucleosomes.

B. Chromatin accessibility at BEN-2 as measured by ChART-PCR with DNase I digestion. Data have been normalised to the inaccessible *SFTPA2* gene promoter.

C. H3K27ac enrichment at BEN-2 as determined by ChIP-qPCR using the percent input method. Grey bars indicate H3K27ac enrichment at the target locus, and black bars show enrichment using a non-specific IgG control antibody. All data are presented as mean ± SEM from at least 3 biological replicates.

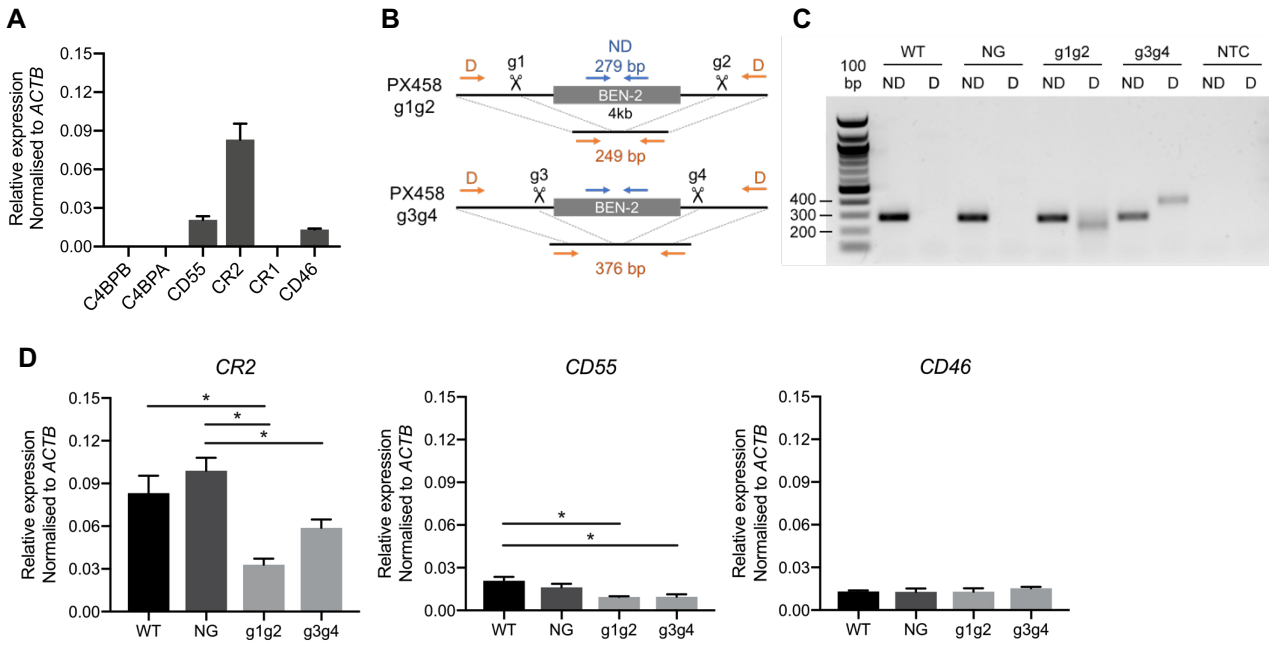


Figure 5: CRISPR deletion of BEN-2 decreased *CR2* and *CD55* mRNA expression in the Raji mature B cell line.

A. Transcript abundance of RCA genes in Raji mature B cell line as measured by qPCR. Values were normalised to the β-actin gene (*ACTB*) using the ΔΔCt method. Bars represent mean relative expression ± SEM from 3 biological replicates.

B. Schematic of enhancer deletion and screening strategy using CRISPR-Cas9. Deletion (4 kb) of BEN-2 was mediated two guide RNAs (gRNA) that cut either side of the enhancer. Plasmids were modified from PX458 to express guide pairs guide 1 and guide 2 (PX458 g1g2), or guide 3 and guide 4 (PX458 g3g4), as well as Cas9 and a GFP marker. Screening was performed using PCR primers that flank the enhancer region (deletion; D) which amplify only in cases where a deletion has occured (orange). PCR primers that amplify within the enhancer region (non-deletion; ND) were used as a control (blue).

C. PCR deletion screen of wild-type Raji DNA (WT) or DNA from an enriched Raji B cell population transfected and enriched with PX458 expressing no guides (NG), PX458-g1g2 (g1g2) or PX458-g3g4 (g3g4). Data are representative of three biological replicates.

D. Transcript abundance of RCA genes in the active TAD (*CR2* and *CD55*) and inactive TAD (*CD46*) was measured by qPCR. Values were normalised to the β-actin gene (*ACTB*) using the ΔΔCt method. Bars represent mean relative expression ± SEM from 3 biological replicates. Asterisks represent statistically significant differences between NG and CRISPR-modified samples ($p < 0.05$).

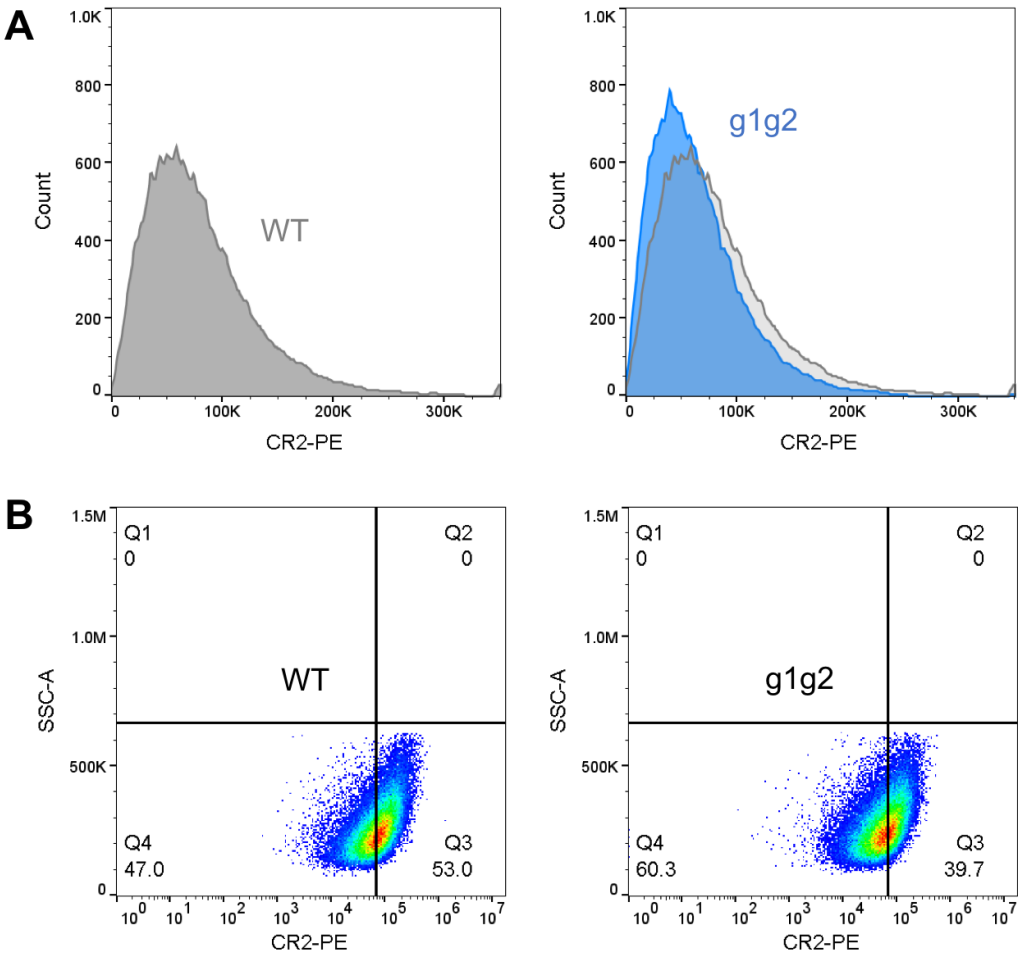


Figure 6: CRISPR deletion of BEN-2 with g1g2 decreases surface expression of CR2 in the Raji B cell line.

A. Cell surface protein expression of CR2 protein was determined using flow cytometry. Cells were labelled with PE-conjugated CR2 antibody or PE-conjugated IgG (isotype control) to confirm CR2-positive expression. Samples were also run alongside unstained and single-colour compensation controls (not shown). For each sample, 50,000 events were collected. Results were consistent across three biological replicates.

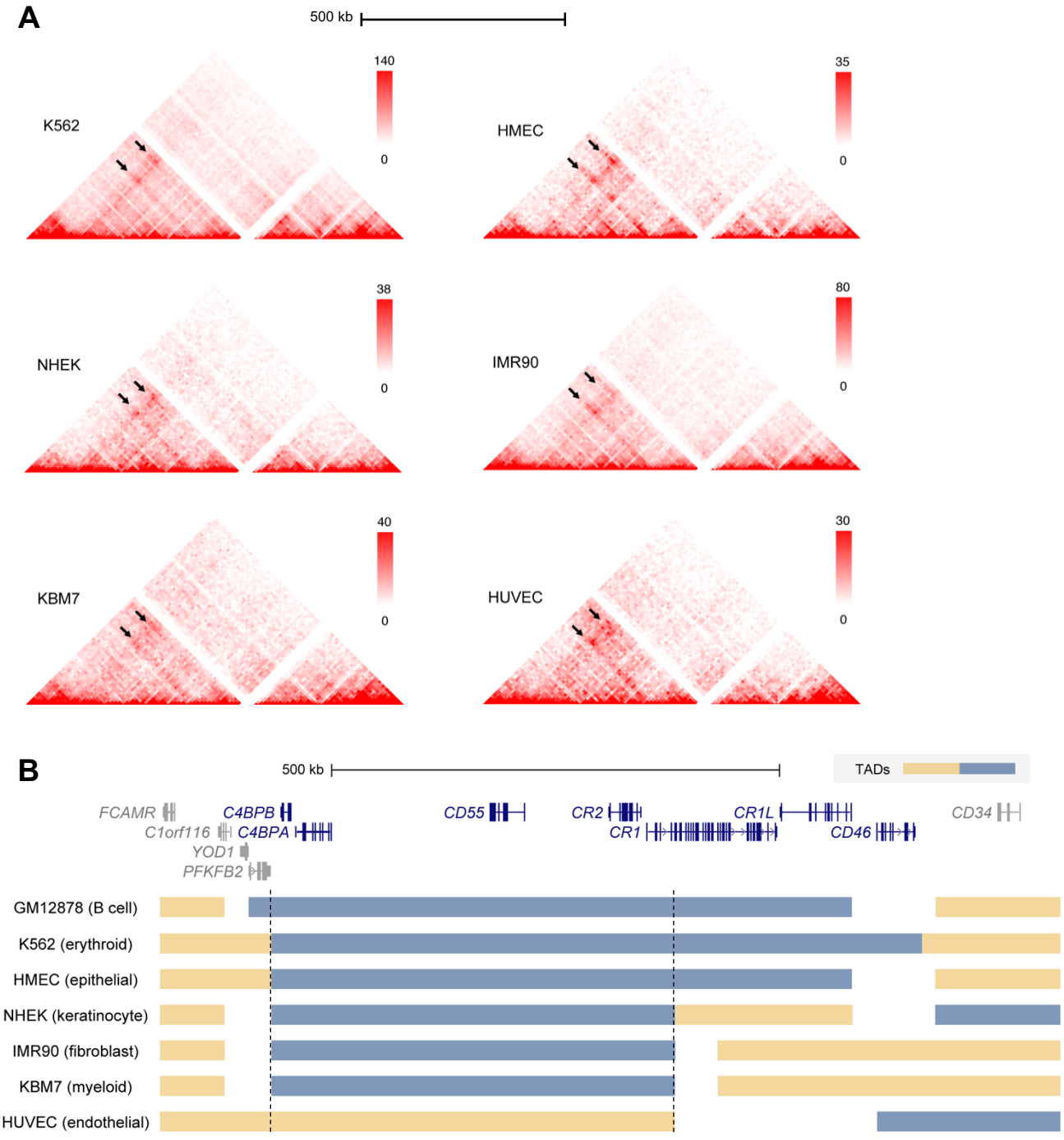
B. WT Raji cells were labelled with CR2-PE were gated based on side-scatter *versus* CR2-PE into four quadrants such that the proportion of cells in Q3 and Q4 were approximately equal. This gate applied to the CR2-PE stained CRISPR enhancer deleted g1g2 population to quantitatively assess shift in CR2-positive cells.

Supplementary Table 1: List of primers used in paper

Primer	Sequence (5' to 3')
4C VP1 read primer	TGCTTTTATGAAGGATTCTCTGATC
4C VP2 read primer	CTACGGATGGTTAATGTTTAGGATC
4C VP3 read primer	TTGGTTTCATCGAGTTGTGATC
4C VP4 read primer	GTGTGACCCAGAATTCAGATC
4C VP1 non-read primer	GTTGGTTGTGTGTATAAGGCG
4C VP2 non-read primer	TCTTGGCTACAGGATGGGTC
4C VP3 non-read primer	GAGGCATATAATTTATGTTCGTGT
4C VP4 non-read primer	ACCATGCAGGATTTCTGGAG
4C read adapter	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTC CGATCT
4C non-read adapter	CAAGCAGAAGACGGCATACGA
BEN-1 forward FWD	CGGGGTACCAGTGCGCATGGGCTATTTACC
BEN-1 forward REV	GCGACGCGTTGAACCAGACCCAGGACTCAG
BEN-1 reverse FWD	GCGACGCGTAGTGCGCATGGGCTATTTACC
BEN-1 reverse REV	CGGGGTACCTGAACCAGACCCAGGACTCAG
BEN-2 forward FWD	CGGGGTACCTGAAGCCATCTCATCCCACAC
BEN-2 forward REV	GCGACGCGTACAGGCATGTGCCAAGTACAC
BEN-2 reverse FWD	GCGACGCGTTGAAGCCATCTCATCCCACAC
BEN-2 reverse REV	CGGGGTACCACAGGCATGTGCCAAGTACAC
BEN-3 forward FWD	GCGACGCGTTCCACAGAGCCAACAGCATTC
BEN-3 forward REV	GGAAGATCTAGCCAGAAGCACAGCTGTATG
BEN-3 reverse FWD	GGAAGATCTTCCACAGAGCCAACAGCATTC
BEN-3 reverse REV	GCGACGCGTAGCCAGAAGCACAGCTGTATG
BEN-4 forward FWD	CGGGGTACCTGCAGATGGAGGTTCCTAGAG
BEN-4 forward REV	GCGACGCGTTGTGCTGTTCATAAGCCATCC
BEN-4 reverse FWD	GCGACGCGTTGCAGATGGAGGTTCCTAGAG
BEN-4 reverse REV	CGGGGTACCTGTGCTGTTCATAAGCCATCC
ACTB mRNA FWD	ACCTTCTACAATGAGCTGCG
ACTB mRNA REV	CCTGGATAGCAACGTACATGG
CD55 mRNA FWD	TGCAACCATCTCCTTCTCATG
CD55 mRNA REV	GGTGCTGGACAATAAATTTCTCTG
CR2 mRNA FWD	TGCCTGTAAAACCAACTTCTC
CR2 mRNA REV	AGCAAGTAACCAGATTCACAG
CD46 mRNA FWD	TCAGTAGCAATTTGGAGCGG
CD46 mRNA REV	AGGTGCAGGATCACAACTATAAG
BEN-2 ChART FWD	AAAGTCCCATGCAACACTGG
BEN-2 ChART REV	AGCAAGGTTCAGAGATGTGC
4C VP1 CTCF ChIP FWD	AGGCCATTGTCACACTGAAAC
4C VP1 CTCF ChIP REV	GTGGTGACCCTGATGATGTG
4C VP2 CTCF ChIP FWD	TAGCTTTGAGGGACCACTGC
4C VP2 CTCF ChIP REV	AATTCTGGAGGTCCCAGCTC
4C VP3 CTCF ChIP FWD	TTCATCCACAACAGCAGAGC
4C VP3 CTCF ChIP REV	TGCCTGGTAAAGCTTAATTCG
4C VP4 CTCF ChIP FWD	ACCACTGAGCTGGGAAGATG
4C VP4 CTCF ChIP REV	TTTTGGTCAGCAGGATTGTG
BEN-2 H3K27ac ChIP FWD	GCAGCACATAAGGGTTCCAG
BEN-2 H3K27ac ChIP REV	GCAGGGCAGAAGAAGGAATG
CRISPR sub-cloning FWD	TGCTCTAGAGCTGGCCTTTTGCTCACATG
CRISPR sub-cloning REV	TTGGGTACCGCCATTTGTCTGCAGAATTG
CRISPR deletion FWD	TGCAGCTGGAAGCCATTATAC
CRISPR deletion REV	ATAGGGTCTCATTCTGTTGCC
CRISPR non-deletion FWD	AGGTTCAGGGAGTCTGAGGTG
CRISPR non-deletion REV	TTCTGCCCTTGTCCACCAAAC

Supplementary Table 2: CRISPR gRNAs sequences (g1-4) for enhancer deletion.

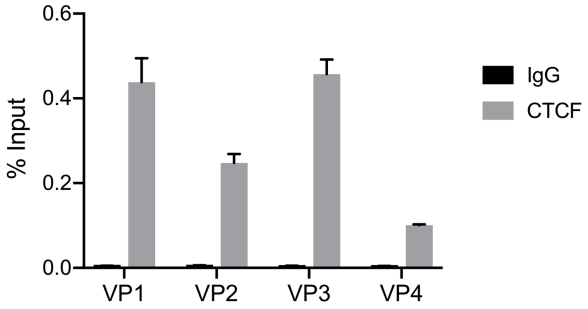
Guide	Sequence (5' to 3')	Genomic co-ordinates (hg38)
g1	GGACATGGACATAGACACTC	chr1:207411153-207411175
g2	AGCTCTCAGACAGTGGGTTA	chr1:207415757-207415779
g3	GACTAATAGAAGGGGGAAGG	chr1:207411175-207411197
g4	CGCTAGTCTGCCGAAGCTCC	chr1:207415664-207415686



Supplementary Figure 1: Hi-C and TAD predictions across several cell-types corresponded to 4C-seq maps and indicated that the RCA gene cluster is divided into two TADs.

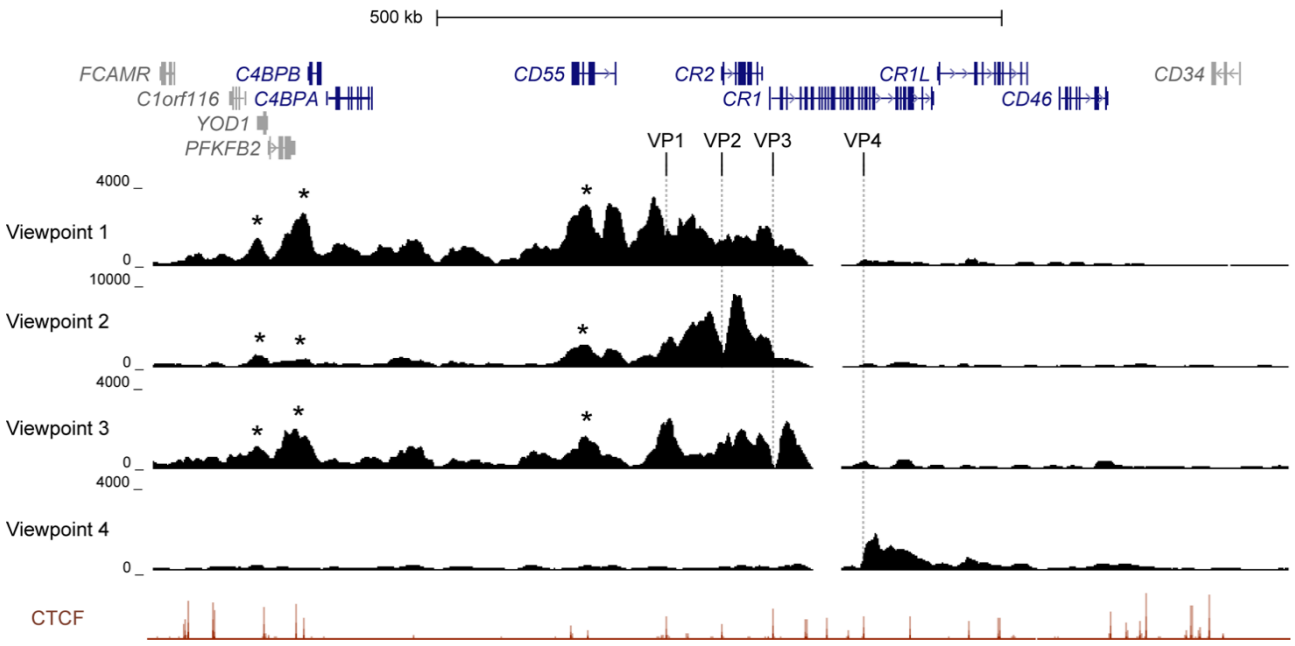
A. Raw Hi-C data at 10 kb resolution from Rao *et al.* for the 1 Mb region across the RCA genes (dark blue) on hg19 (chr1:207,120,000-208,130,000), as shown in Figure 1. Highly frequent chromatin interactions between *C4BPB, C4BPA, CD55, CR2* and *CR1* in GM12878 were observed in all cell lines assessed in this dataset (arrows).

B. Predicted TAD co-ordinates from Dixon *et al.* varied between different cell lines, but RCA genes *C4BPB, C4BPA, CD55, CR2* and the *CR1* promoter were consistently situated in the same TAD (dotted lines). A boundary at exon 5 of *CR1* (hg19 chr1:207,700,000) was identified in NHEK, IMR90, KBM7 and HUVEC cell lines.

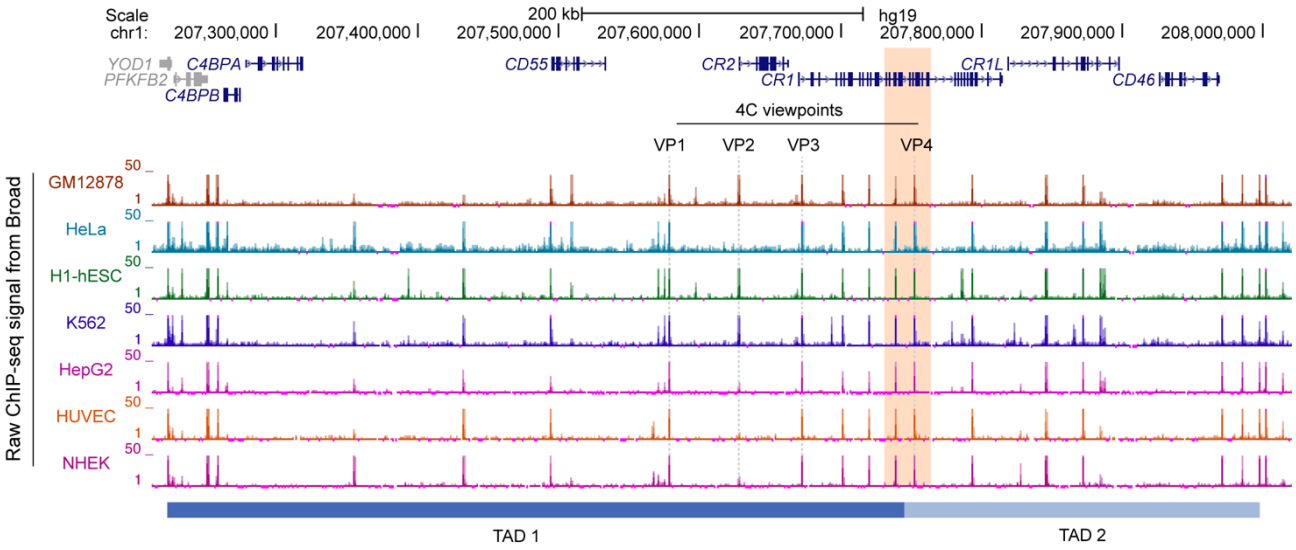


Supplementary Figure 2: CTCF enrichment at 4C viewpoints in the B-0028 B lymphoblastoid cell line was confirmed by ChIP-qPCR using the percent input method.

Grey bars indicate H3K27ac enrichment at the target locus, and black bars show enrichment using a non-specific IgG control antibody. All data are presented as mean ± SEM from at least 3 biological replicates.



Supplementary Figure 3: 4C-seq maps from CTCF viewpoints in the RCA gene cluster were replicated in the B-0056 lymphoblastoid cell line.
Maps were generated from four viewpoints on CTCF sites, as indicated by GM12878 ChIP-seq signal for CTCF from ENCODE, in the intergenic region between *CR2* and *CD55* (viewpoint 1), intron 1 of *CR2* (viewpoint 2), the intron 1 of *CR1* (viewpoint 3) and intron 29 of CR1 (viewpoint 4).



Supplementary Figure 4: Divergent CTCF sites at the inter-TAD boundary (orange) were constitutive across multiple cell types.
Raw ChIP-seq signal for CTCF from ENCODE/Broad was visualised using the UCSC Genome Browser on hg19. CTCF sites used as 4C viewpoints in TAD 1 (VP1, VP2, VP3) and TAD 2 (VP4) of the RCA gene cluster are indicated.

Supplementary Table 3: Double elite putative enhancers in TAD 1 and TAD 2 of the RCA gene cluster, and their identified gene targets from GeneHancer.

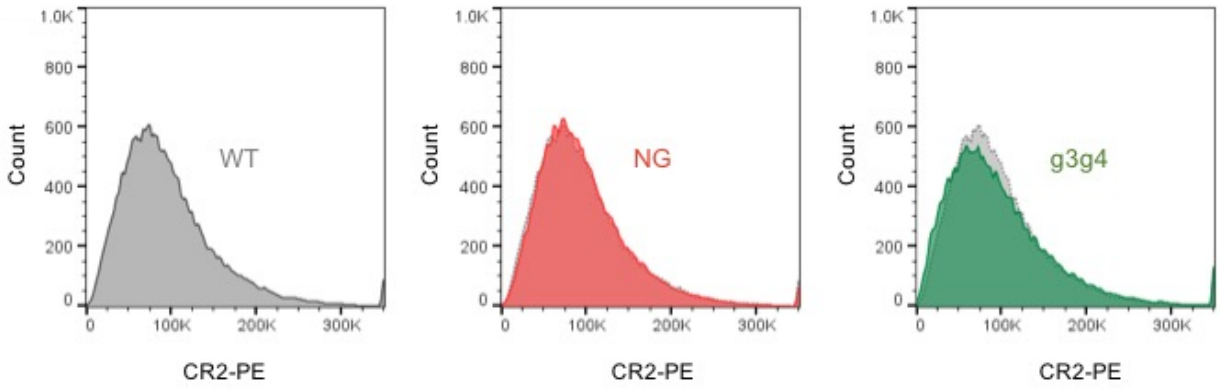
	GeneHancer ID	Relative location	Predicted gene targets
TAD 1	GH01J207064	*PFKFB2*	*C1orf116 FCAMR PIGR IL20 PFKFB2 YOD1 C4BPB*
	GH01J207066	*PFKFB2*	*C1orf116 FCAMR PIGR IL20 PFKFB2 YOD1 C4BPB*
	GH01J207070	*PFKFB2*	*PFKFB2 C1orf116 FCAMR PIGR C4BPA IL20 YOD1 C4BPB*
	GH01J207077	*PFKFB2 – C4BPB*	*PFKFB2 CR1 C4BPA FCAMR PIGR C4BPB YOD1*
	GH01J207204	*C4BPA – CD55*	*CD55 C4BPAP3 C4BPAP2*
	GH01J207239	*C4BPA – CD55*	*CD55 CR2 IL19 C4BPA ENSG00000237074*
	GH01J207280	*C4BPA – CD55*	*CD55 CR2 C4BPA ENSG00000237074*
	GH01J207298	*C4BPA – CD55*	*CD55 ENSG00000237074*
	GH01J207304	*C4BPA – CD55*	*CD55 CR2 ENSG00000237074*
	GH01J207306	*C4BPA – CD55*	*CR2 CD55 ENSG00000237074*
	GH01J207309	*C4BPA – CD55*	*CD55 EIF2D CR2 ENSG00000237074*
	GH01J207316	*C4BPA – CD55*	*CDCA4P4 CD55 CD46 EIF2D ENSG00000234981 CR2 CR1 ENSG00000237074*
	GH01J207327	*CD55*	*CD55 CR2 LOC105372881*
	*GH01J207333**	*CD55*	*CD55 EIF2D CR2 CR1 CD46 LOC105372881*
	GH01J207363	*CD55 – CR2*	*EIF2D CD55 CR2 LOC105372881*
	GH01J207407	*CD55 – CR2*	*CR1 CD55 CR2 C4BPB C4BPA CD46 YOD1 ENSG00000283044 LOC105372881*
	*GH01J207411**	*CD55 – CR2*	*CR2 CR1 CD55 ENSG00000283044 LOC105372881*
	*GH01J207424**	*CD55 – CR2*	*CR2 CR1 C1orf116 C4BPA LOC105372880*
TAD 2	GH01J207625	*CR1*	*CR1 C4BPA CD46P1 ENSG00000236911 GC01M207468*
	GH01J207648	*CR1L*	*CR1L C4BPA MIR29B2CHG CD46P1 CDCA4P3 GC01M207468*
	GH01J207649	*CR1L*	*CR1L C4BPA MIR29B2CHG CR2 CD46P1 CDCA4P3 GC01M207468*
	GH01J207755	*CD46*	*CD46 MIR29B2CHG CDCA4P4*

* These predicted enhancers were functionally investigated in this paper.

Supplementary Table 4: Strong candidate B cell enhancers (BENs) were predicted to regulate genes on GeneHnacer using eQLT analyses, promoter interactions from capture Hi-C (CHi-C) and/or enhancer RNA and mRNA co-expression (eRNA).

			BEN-1	BEN-2	BEN-3	BEN-4
Non-RCA	*PIGR*	Method				CHi-C
		Score				9.7
	FCAMR	Method				CHi-C
		Score				9.7
	C1orf116	Method			CHi-C	
		Score			10.2	
	PFKFB2	Method				
		Score				
RCA	*C4BPA*	Method			CHi-C	
		Score			9.4	
	CD55	Method	eQTL eRNA	CHi-C		CHi-C
		Score	*27.2**	10.3		10.4
	CR2	Method	eQTL	CHi-C eRNA	CHi-C eRNA	CHi-C
		Score	9.6	*18.7**	*18.1**	12.3
	CR1	Method	eRNA	CHi-C	CHi-C	CHi-C
		Score	7.2	11.8	11.1	11.1
	CR1L	Method				
		Score				
	CD46	Method	eRNA			
		Score	3.6			

* These predicted gene-enhancer interactions were identified using more than one method on GeneHancer.



Supplementary Figure 5: CR2 surface protein expression corresponded to CR2 transcript abundance with the no guide (NG) control and BEN-2 deletion with g3g4.
Cell surface protein expression of CR2 protein was determined using flow cytometry. Cells were labelled with PE-conjugated CR2 antibody or PE-conjugated IgG (isotype control) to confirm CR2-positive expression. Samples were also run alongside unstained and single-colour compensation controls (not shown). For each sample, 50,000 events were collected. Results were consistent across three biological replicates.