Regulatory Architecture of the RCA Gene Cluster Captures an Intragenic TAD Boundary, CTCF-Mediated Chromatin Looping and a Long-Range Intergenic Enhancer

Jessica Cheng[1], Joshua S. Clayton[2,3], Rafael D. Acemel[4], Ye Zheng[5,6], Rhonda L. Taylor[2,3], Sündüz Keleş[6,7], Martin Franke[4], Susan A. Boackle[8,9], John B. Harley[10,11,12], Elizabeth Quail[1,13], José Luis Gómez-Skarmeta[4†] and Daniela Ulgiati[1*]

[1] School of Biomedical Sciences, The University of Western Australia, Crawley, WA, Australia, [2] Harry Perkins Institute of Medical Research, QEII Medical Centre, Nedlands, WA, Australia, [3] Centre for Medical Research, The University of Western Australia, Crawley, WA, Australia, [4] Centro Andaluz de Biología del Desarrollo, Consejo Superior de Investigaciones Científicas/Universidad Pablo de Olavide, Sevilla, Spain, [5] Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, United States, [6] Department of Statistics, University of Wisconsin-Madison, Madison, WI, United States, [7] Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, United States, [8] Department of Medicine, Division of Rheumatology, University of Colorado Anschutz Medical Campus, Aurora, CO, United States, [9] Rocky Mountain Regional Veterans Affairs Medical Center, Aurora, CO, United States, [10] Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States, [11] Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, United States, [12] US Department of Veterans Affairs Medical Centre, US Department of Veterans Affairs, Cincinnati, OH, United States, [13] School of Molecular Sciences, The University of Western Australia, Crawley, WA, Australia

The Regulators of Complement Activation (RCA) gene cluster comprises several tandemly arranged genes with shared functions within the 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 mechanisms 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 topologically associated domains (TADs). Interestingly, an 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 autoimmune disease susceptibility may be influenced by genetic variants, thus highlighting the important contribution of chromatin topology to gene regulation and complex genetic disease.

Keywords: RCA, B cells, enhancers, chromatin looping, TADs

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 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) on chromosome 1q32.2. 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, 5). 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 roles in innate immune responses, members of RCA gene cluster are involved in the processes of tissue injury, inflammation and apoptosis. Accordingly, they have been implicated in a range of inflammatory and autoimmune disorders (9–12).

A role for complement receptors CR2 and CR1 in the autoimmune disease, Systemic Lupus Erythematosus (SLE) is well established. 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 and CR1 regulate B cell responses by modulating B cell activation and antibody production upon binding of complement-tagged antigens (15, 16). Aberrant expression of CR2 on the surface of B cells has been demonstrated both in mouse models of the disease (17, 18) and SLE patients (19, 20), which functionally contributes to B cell autoreactivity and autoimmune disease susceptibility (21–23). The *CR1* gene contains an 18 kb intragenic segmental duplication, known as 'low copy repeat 1' (LCR1). This repeat results in different structural alleles of *CR1* with recognised association to SLE susceptibility, but the functional role of this large genomic duplication is not well understood (24). The *CR2* gene has also been implicated in SLE at the genetic level through linkage analyses (25–27) and association studies (27–29). A SNP within the first intron of *CR2* (rs1876453) was shown to alter the expression of the neighbouring gene (*CR1*) without influencing CR2 expression (29), 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*, suggesting that CTCF may have a role in co-regulating expression of *CR2, CR1* and the RCA gene cluster (29).

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. Additionally, CTCF has been shown to play a critical role in forming chromatin loops and mediating interactions between distal loci (30). Chromatin loops are organised into genomic compartments known as topologically associated domains (TADs) (31). 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 (32). 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 (33, 34). 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 (33, 35).

Enhancers represent an important class of distal-regulatory elements which are largely responsible for governing cell-type specific gene expression patterns. 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 (36). Importantly, the majority of disease-associated SNPs from genome-wide association studies (GWAS) fall within enhancer regions (37). 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) (38, 39). 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 (36). 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 the identification of enhancer targets. However, these data are still limited by resolution and the physical chromatin interactions detected using these methods may not necessarily be functional. As such, experimental validation of enhancers and physically associating enhancer-gene pairs is imperative to determine their influence on gene expression (36, 40). In addition, large repetitive regions in the genome, such as the LCR in *CR1*, cannot be uniquely aligned and readily analysed using next-generation

sequencing technologies. As a result, these regions are under-represented in high-throughput epigenetic and chromatin conformation capture datasets.

The aim of this investigation was to explore the structural organisation of the RCA gene cluster in order to identity transcriptional 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 discretely separates chromatin interactions in the RCA gene cluster into two domains and co-localises to the intragenic segmental duplication in *CR1*. Importantly, we functionally interrogated a putative long-range enhancer and demonstrated that it co-regulates two genes within a TAD in B cells. Collectively, we have revealed how three-dimensional chromatin organisation plays a role in regulating the RCA gene cluster and have uncovered novel regulatory loci which govern the expression of these genes.

MATERIALS 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 (29). 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
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 (41, 42). Sequences of primers used as 4C viewpoints are listed in **Supplementary Table 1**.

Bioinformatic Datasets and Pipelines
Hi-C data for GM12878 from Rao et al. (43) were visualised as contact heatmaps and virtual 4C signal using the 3D Genome Browser (44) and Juicebox (45). CTCF orientation calls from GM12878 were retrieved from Rao et al. (43) and assessed in the *CR1* segmental duplication 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.

Mapping Hi-C Reads with mHi-C
Multi-mapping Hi-C sequencing reads from Rao et al. (43) were evaluated using the mHi-C pipeline (47) at 5 kb resolution (**Supplementary Table 2**). mHi-C was used as described in Zheng et al. (47) with a novel post-mHiC processing strategy. In brief, the genomic distance effect on the contact probabilities is estimated using the univariate spline model based on uniquely mapping reads. Such prior probabilities information is updated iteratively by the local bin-pairs contact counts leveraging both uniquely mapping reads and multi-mapping reads. The posterior probabilities, as the results of mHi-C, quantify the chance for the candidate bin-pair to be the true origin for each multi-mapping read pair. Instead of applying general filtering based on the posterior score by a fixed threshold, the posterior probabilities are interpreted as fractional Hi-C contact counts to incorporate a more significant number of the multi-mapping reads into the analysis. To examine interaction artefacts due to highly repetitive sequences, a stringent multi-mapping allocation strategy was employed which enforced all the multi-mapping reads assigned to corresponding regions have greater or equal to a 0.99 posterior score. Subsequent to rescuing multi-mapping reads by mHi-C, TAD boundaries are detected by the state-of-the-art TAD caller spectralTAD (48) which provides nested TAD at different levels. The TADs shown in **Figure 2** are first-level TAD boundaries called at 25 kb resolution.

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. 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^{TM} 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$_2$O (Life Technologies). qPCR reactions comprised 1X SYBR Green No-Rox (Bioline), 250 nM forward and reverse primers (**Supplementary Table 1**), 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 (49) 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 and 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 Deletion

CRISPR plasmid constructs were modified from pSpCas9(BB)-2A-GFP (PX458), a gift from Feng Zhang (50) (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. gRNAs were designed using CRISPRscan (51) to select highest scoring sequences with minimal off-target effects (**Supplementary Table 3**). 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 *Acc*65I and *Xba*I 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+ expression using fluorescent activated cell sorting (FACS) on a FACSAria II (BD Bioscience). The GFP+ pool (bulk) was expanded for further analysis. To obtain single cell clones, the bulk sample was plated into 96-well plates at approximately 1 cell per well in conditioned media. Single cell colonies were expanded and cells were cryopreserved. Genomic DNA and total RNA extraction was performed 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

DNA was qualitatively screened for the 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 20 μL, 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. Single cell clones that screened positive for genomic deletion were confirmed using Sanger sequencing. The deletion product from the bulk sample was also analysed using Sanger sequencing.

Flow Cytometry

Cells (1 × 10^6 cells) were harvested and washed with cold staining buffer (PBS with 5% FBS (v/v)) at 300 × *g* for 5 min at 4°C. For surface staining, cells were resuspended in 90 μL staining buffer and incubated with 10 μL of anti-human CD21-PE (Cat #555422, BD Bioscience), anti-human CD55-PE (MHCD5504, Thermo Fisher Scientific) or IgG1κ-PE isotype control (Cat #555749, BD Bioscience) for 20 min on ice. After incubation, cells were washed and resuspended in 0.5 mL staining buffer and

processed using a BD Accuri C6 flow cytometer (BD Bioscience). Data was analysed using FlowJo software V10.8.0 (Tree Star). Samples were run alongside unstained controls.

Statistical Analysis

Differences in transcriptional activity, mRNA expression and mean fluorescence intensity 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. (43) 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 (**Figure 1A**), indicating that chromatin interactions in this region may be directionally constrained and organised to more than one TAD. This pattern of interaction was consistent across the 6 other cell lines also examined at 10 kb resolution in Rao et al. (43) Hi-C dataset (K562, HMEC, NHEK, IMR90, KBM7, HUVEC) (**Supplementary Figure 1**). The *CR1* intragenic duplication (*CR1* exon 5 – 20) is included on the human reference genome, but Hi-C interaction data was filtered out at this region in all datasets due to sequence repetitiveness and sequence unmappability (**Figure 1A**; **Supplementary Figure 1**).

To further clarify the TAD organisation of the RCA gene cluster, we performed 4C-seq in an analogous B lymphoblastoid cell line (B-0028). It is well established that in the B cell lineage, only membrane-bound RCA members (*CD55, CR2, CR1, CD46*) are expressed, not soluble protein members (*C4BPB* and *C4BPA*) (11). We confirmed this expression pattern using qPCR in the B-0028 cell line (**Supplementary Figure 2**). As CTCF plays an important role in chromatin looping, we selected 4C viewpoints (VP) from CTCF sites utilising B cell ChIP-seq data. (GM12878) Specifically viewpoints were selected which co-localised to regions of highly frequent interactions observed in the Hi-C data (**Figure 1B**). These viewpoints included; the intergenic region between *CD55* and *CR2* (VP1), intron 1 of *CR2* (VP2), which is the CTCF site influenced by SLE-associated SNP rs1876453 (29), 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 the upstream region of this gene cluster (**Figure 1B**). We confirmed enrichment of CTCF at these VPs in the B-0028 cell line using ChIP-qPCR (**Supplementary Figure 3**).

4C maps from VP1, VP2 and VP3 yielded consistent 4C signal peaks at upstream CTCF sites near RCA member *C4BPB* and within non-RCA member *YOD1* (**Figure 1C**, asterisks), corresponding to Hi-C data (**Figure 1A**). These CTCF viewpoints also consistently interacted with the CTCF site within intron 6 of RCA gene *CD55* (**Figure 1C**, asterisks). 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 a 60 kb region 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), whereby CTCF interactions in the RCA gene cluster were also organised to two discrete regions (Supplementary **Figure 4**), potentially representing two TADs with an inter-TAD boundary located within the *CR1* gene itself. However, like Hi-C, reads mapping to the *CR1* segmental duplication were filtered out during 4C-seq data processing.

To refine the location of the TAD boundaries in the RCA gene cluster, we used a customised mHi-C pipeline which probabilistically assigns multi-mapping reads in Hi-C experiments to their most likely genomic position (47). Indeed, mHi-C successfully recovered Hi-C interactions across the *CR1* segmental duplication in the GM12878 cell line at 5 kb resolution (**Figure 2A**). These interactions were visualised as virtual 4C signal from RCA gene 5' upstream promoter regions using Juicebox (45) (**Figure 2B**). In line with our previous observations, interactions from the promoters of upstream RCA genes *C4BPB, C4BPA, CD55, CR2* and *CR1* were localised to a distinct domain region compared to downstream RCA genes *CR1L* and *CD46* (**Figure 2B**). After recovering multi-mapping reads, we used spectralTAD (48) to systematically call TADs. Using this TAD caller, TAD boundaries were assigned at the intergenic region upstream of *YOD1*, intron 11 of *CR1* and the intergenic region downstream of *CD46*, placing RCA genes *C4BPB, C4BPA, CD55, CR2 and CR1* in TAD 1, and *CR1L* and *CD46* in TAD 2 (**Figure 2B**).

To corroborate these findings, we examined CTCF enrichment and motif binding orientation at the TAD boundaries in the RCA gene cluster. Each TAD was flanked by CTCF enrichment in the GM12878 B cell line and convergent CTCF motifs, characteristic of TAD boundaries. Importantly, the inter-TAD boundary was directly located to the *CR1* segmental duplication. Like other highly repetitive, unmappable genomic regions, CTCF enrichment at this region is underrepresented in the high-throughput datasets such as the ENCODE portal following the typical ChIP-seq processing pipeline. Therefore, we examined raw CTCF ChIP-seq signal in the GM12878 cell lines and observed enrichment of CTCF at this repeat element underlined by reverse orientation CTCF motifs at each repeat segment which has not been previously defined (**Figure 2B**). Enrichment of CTCF at the *CR1* repeat segments (*CR1* intron 7, 15 and 23) was observed across all cell lines in this ENCODE CTCF signal data set (**Supplementary Figure 5**). Taken together, our data shows that the RCA gene cluster is divided into two
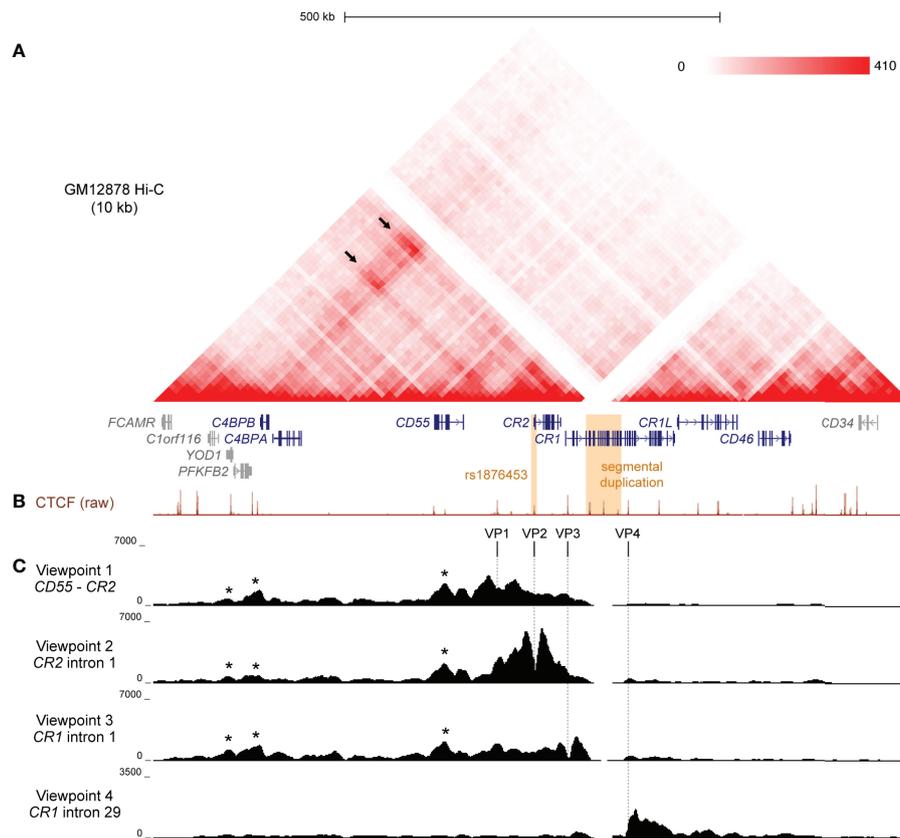
FIGURE 1 | Chromatin conformation of the RCA gene cluster in B cells. **(A)** Hi-C heatmap matrix (10 kb resolution) for the GM12878 B cell line from Rao et al. (41) for the 1 Mb region across the RCA genes (dark blue) on hg19 (chr1:207,120,000-208,130,000). Relative interaction frequencies between two 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. SLE-associated variants are indicated in orange. **(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, VP1), intron 1 of *CR2* (viewpoint 2, VP2), the intron 1 of *CR1* (viewpoint 3, VP3) and intron 29 of CR1 (viewpoint 4, VP4). 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.
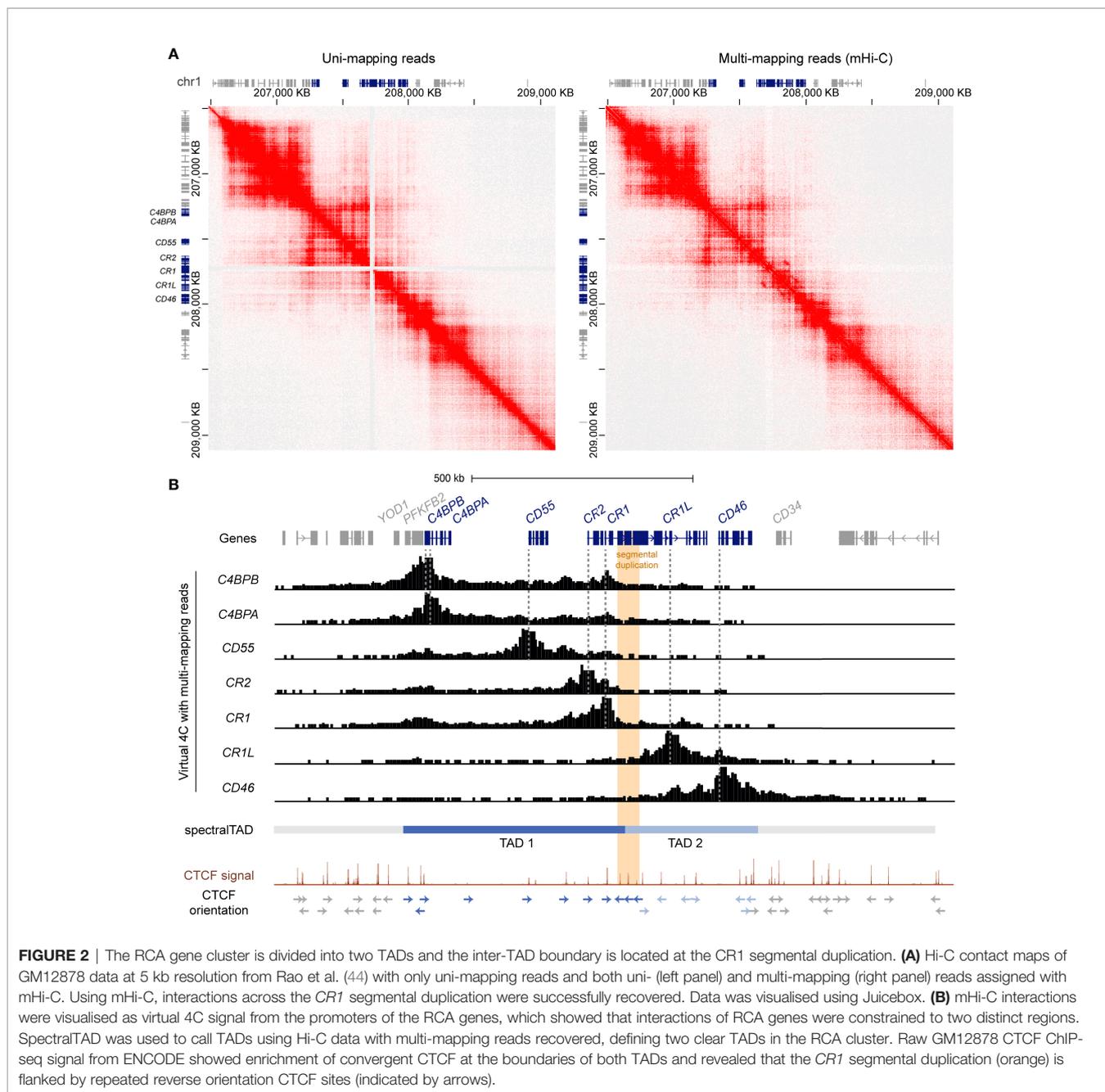
TADs and reveals that a TAD boundary is located within the *CR1* gene at the intragenic segmental duplication.

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 (30). As CTCF-mediated chromatin looping was identified in the RCA gene cluster in B cells (**Figure 1**), we speculated that enhancer elements were present in this region that may regulate these genes in this cell lineage. To identify putative enhancers in the RCA, we leveraged a well-known enhancer database, GeneHancer. This integrates enhancer datasets from multiple consortium-based projects and other functional datasets to generate enhancer predictions and identify their potential gene-targets (52). Confidence scores for each enhancer prediction

(GeneHancer score) and enhancer-gene prediction (gene-association score) were computationally assigned, based on the level of evidence retrieved. A strength of this database is that predicted enhancers can be classified as "double elite" if both their GeneHancer and gene-association scores were derived from more than one source of data, thus representing a prediction which is more likely to be functional (52). Numerous predicted enhancers on GeneHancer were identified across TAD 1 and TAD 2, but only a subset of these were classified as "double elite" (**Figures 3A, B**; **Supplementary Table 4**).

Enhancers are important in cell-type specific regulation of gene expression and act by looping to their target gene promoters (53). To identify active enhancers that were most likely functional in B cells, we examined epigenetic marks characteristic of enhancers, such as H3K27ac and DNase I hypersensitivity (DHS) within candidate regions. We identified four candidate B cell enhancers (BENs) in TAD 1 that showed strong consistent H3K27ac enrichment and DHS in both B cell

FIGURE 2 | The RCA gene cluster is divided into two TADs and the inter-TAD boundary is located at the CR1 segmental duplication. **(A)** Hi-C contact maps of GM12878 data at 5 kb resolution from Rao et al. (44) with only uni-mapping reads and both uni- (left panel) and multi-mapping (right panel) reads assigned with mHi-C. Using mHi-C, interactions across the *CR1* segmental duplication were successfully recovered. Data was visualised using Juicebox. **(B)** mHi-C interactions were visualised as virtual 4C signal from the promoters of the RCA genes, which showed that interactions of RCA genes were constrained to two distinct regions. SpectralTAD was used to call TADs using Hi-C data with multi-mapping reads recovered, defining two clear TADs in the RCA cluster. Raw GM12878 CTCF ChIP-seq signal from ENCODE showed enrichment of convergent CTCF at the boundaries of both TADs and revealed that the *CR1* segmental duplication (orange) is flanked by repeated reverse orientation CTCF sites (indicated by arrows).

lines and primary B cells (**Figure 3C**). 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 contained binding sites for numerous transcription factors (based on ENCODE ChIP-seq data) including those important in B cell development, such as early B cell factor 1 (EBF1) (54) and PAX5 (55), and general regulatory factors (eg. EP300, CTCF and RNA polymerase II) (**Table 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, 52, 56). 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 3D**; **Supplementary Table 5**). However, only interactions between BEN-1 and *CD55*, BEN-2 and *CR2*, and BEN-3 and *CR2*, represented
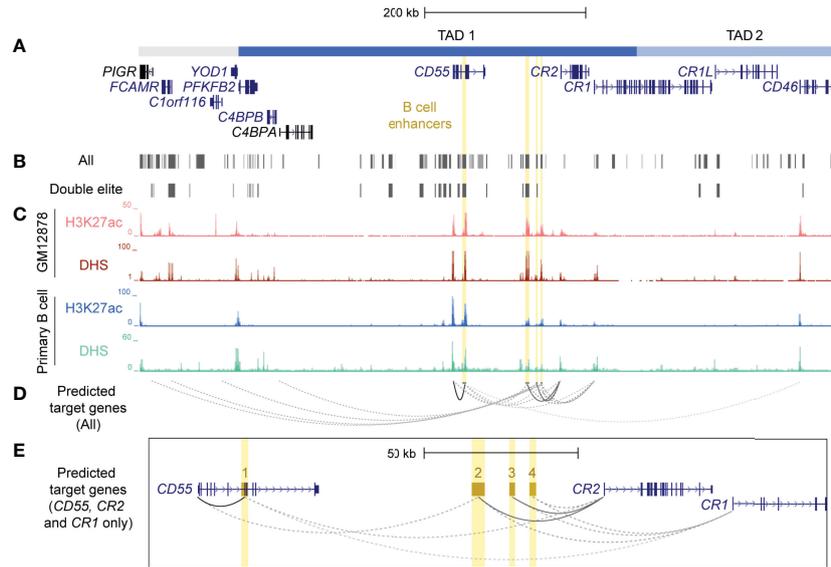
FIGURE 3 | Identification and prioritisation of candidate B cell enhancers in TAD 1. **(A)** The structural organisation of the RCA genes (dark blue) and upstream genes (grey) *PIGR, FCAMR, C1orf116* and *YOD1* on hg19 (chr1:207,104,491-207,978,031). **(B)** Putative enhancers were identified by GeneHancer 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 human 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). **(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. Scores determined by GeneHancer are listed in **Supplementary Table 5**. **(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 on 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.

high-confidence ("elite") associations, being identified by more than one contrasting method (**Figure 3E**; **Supplementary Table 5**). Of these, only BEN-1 was predicted to regulate a gene (*CD46*) located downstream of the intragenic TAD boundary in *CR1* (**Figure 3D**). However, this predicted

interaction had the lowest score among all gene-enhancer predictions for these BENs (**Figure 3D**; **Supplementary Table 5**). Although these gene-enhancer interactions were based on bioinformatic predictions, this highlighted the potential for the RCA genes to be co-regulated in B cells.

TABLE 1 | Candidate B cell enhancers (BENs) on GeneHancer were identified from multiple enhancer databases, contained numerous transcription factor binding sites (TFBS) and were predicted to regulate several 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.

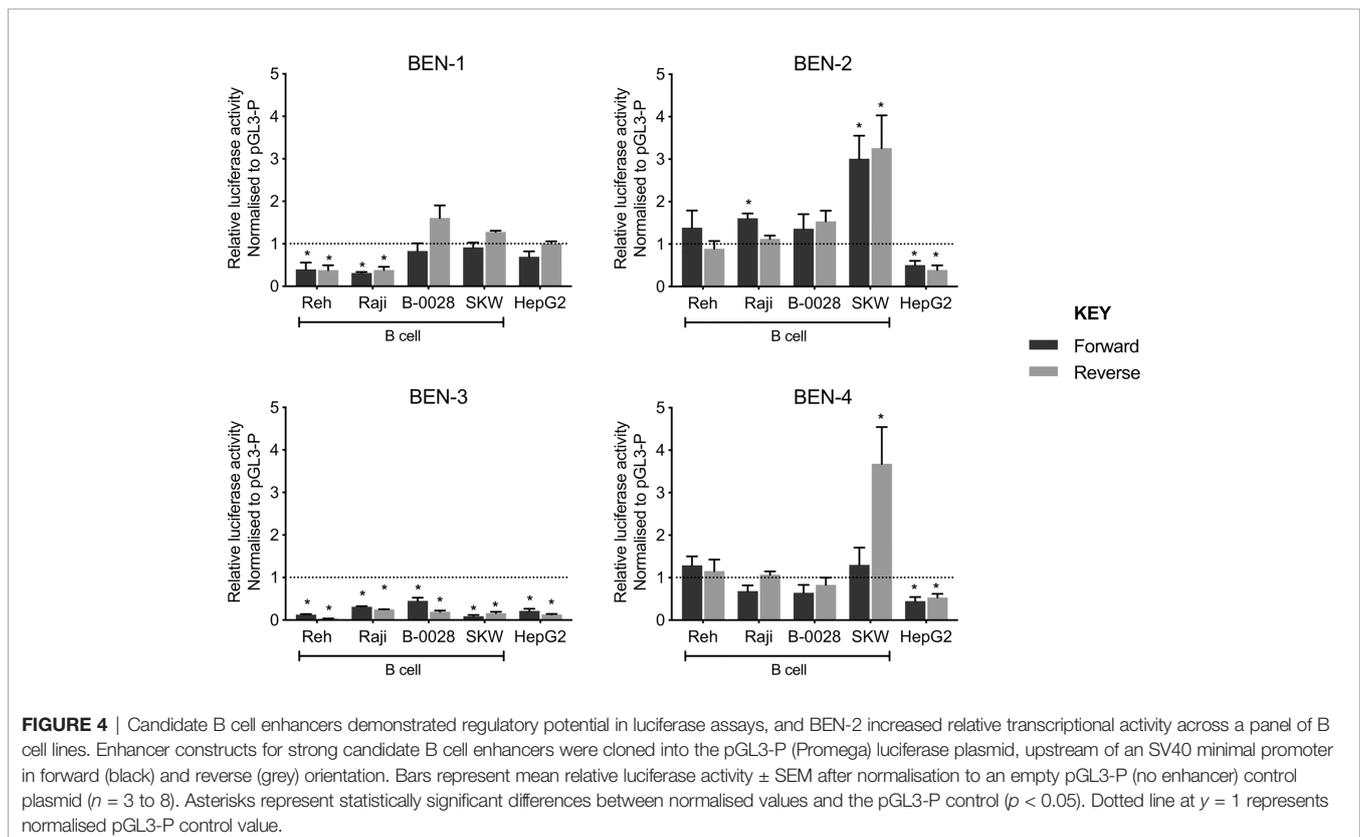## Candidate B Cell Enhancers in the RCA Gene Cluster Were Functional *in Vitro*

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 4**). Interestingly, transcriptional 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 4**). BEN-4 displayed some enhancer activity in 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 4**). 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 4**). Notably, transcriptional activity of BEN-2 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 4**). Together, these data indicated that BEN-2 was the most likely candidate BEN to be active in the B cell lineage.
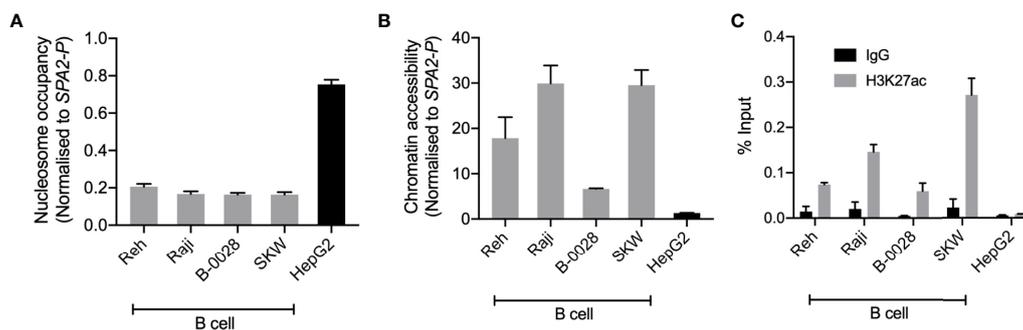
To support the functional role of BEN-2 in this cell type, we quantified epigenetic marks characteristic of enhancer regions, such as chromatin accessibility and H3K27ac enrichment, in the panel of cell lines used in the luciferase assays. Nucleosome occupancy was measured using MNase, a micrococcal nuclease which cannot bind and digest nucleosome-bound DNA. At BEN-2, nucleosome occupancy was consistently lower in the B cell lines than in the non-B cell control, HepG2 (**Figure 5A**). Conversely, chromatin accessibility was consistently high across all B cell lines, but inaccessible in HepG2 (**Figures 5A, B**), 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 5C**). Together, 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 Decreased *CR2* and *CD55* Expression at the Transcript and Protein Levels

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 3**). CRISPR deletion machinery was delivered using a plasmid-based method into the Raji mature B cell line. This cell line expresses *CD55*, *CR2* and *CD46*, although *CR1* is not expressed at levels detectable by qPCR (**Figure 6A**). This pattern of gene expression is in

FIGURE 4 | 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 (black) and reverse (grey) 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 5 | 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 and non-B control (HepG2, liver). **(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.

accordance with other B cell lines, such as B-0028 (**Supplementary Figure 2**). To efficiently delete the BEN-2 region, we modified the PX458 CRISPR plasmid to express two gRNA sequences that cut either side of BEN-2 (**Figure 6B**). The CRISPR plasmids, containing a GFP marker, were delivered into Raji cells and successfully transfected GFP-positive cells were enriched by fluorescence activated cell sorting (FACS). The

resultant polyclonal GFP+ bulk population (bulk) was expanded and used for single cell cloning by limiting dilution. Successful enhancer deletion was qualitatively assessed using PCR (**Figures 6B, C**). Indeed, PCR indicated that BEN-2 was deleted in a proportion of cells in the bulk population (**Figure 6C**). After screening expanded single cell clones, we successfully isolated a population containing a homozygous
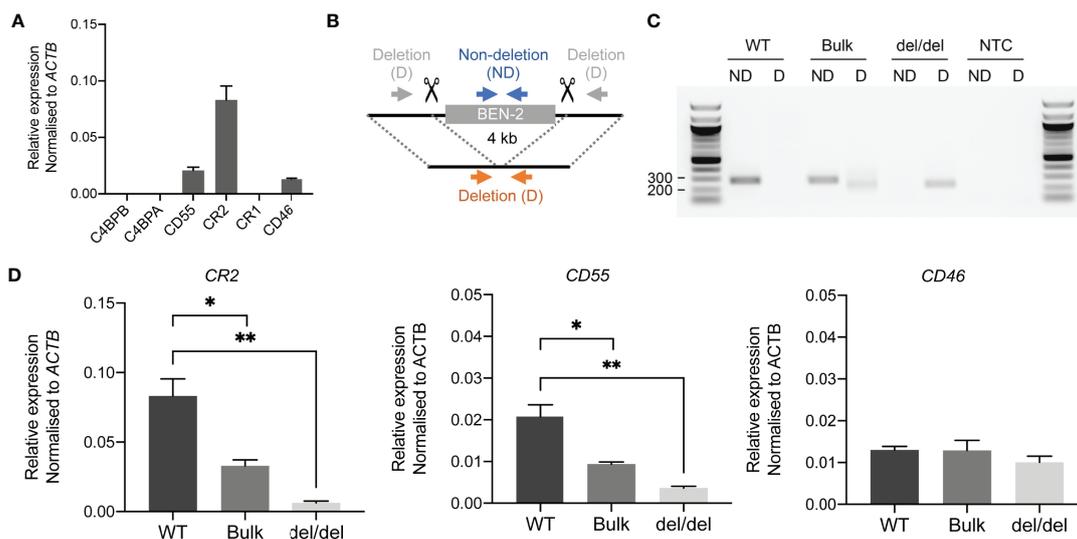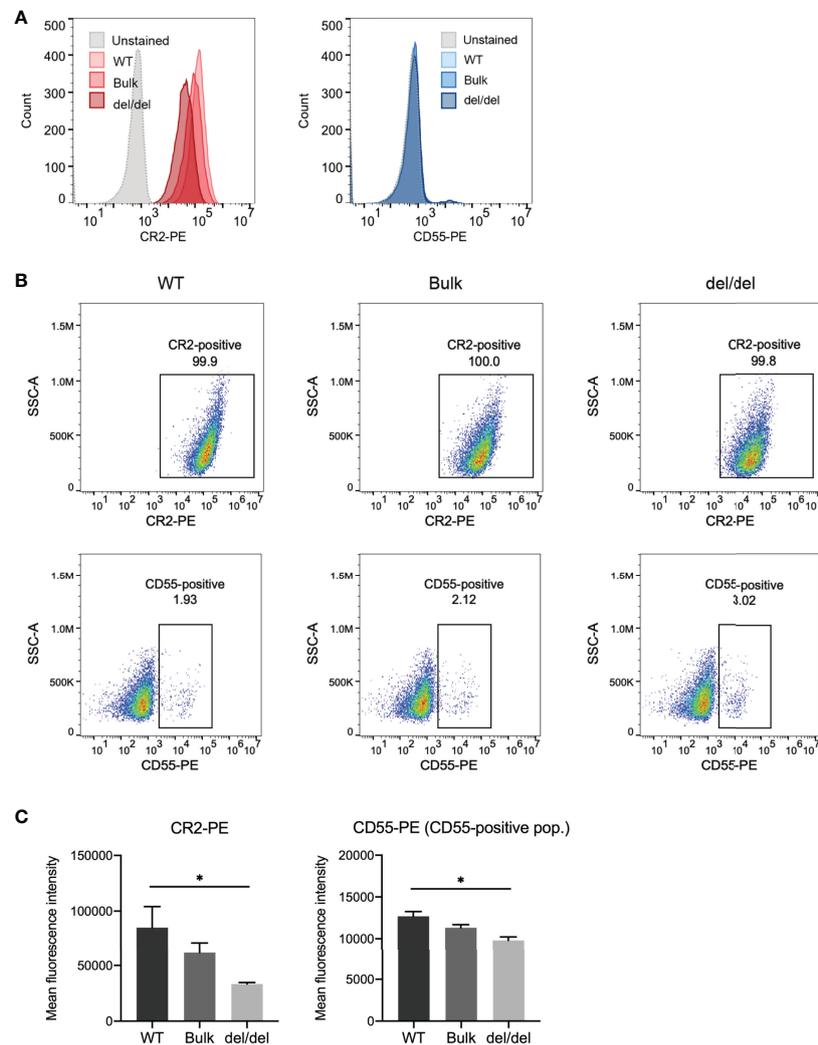

FIGURE 6 | CRISPR deletion of BEN-2 decreased *CR2* and *CD55* mRNA expression in the Raji mature B cell line. **(A)** Transcript abundance of RCA genes in Raji mature B cell line as measured by qPCR. Values were normalised to the β-actin gene (*ACTB*) using the ΔCt method. Bars represent mean relative expression ± SEM from at least 3 biological replicates. **(B)** Schematic of enhancer deletion and screening strategy using CRISPR-Cas9. Deletion (4 kb) of BEN-2 was mediated by two gRNAs that cut either side of the enhancer region. Plasmids were modified from PX458 to express the two guides, 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 occurred (orange arrows). PCR primers that amplify within the enhancer region (non-deletion; ND) were used as a control (blue arrows). **(C)** PCR deletion screen of wild-type Raji DNA (WT), polyclonal GFP+ bulk population (bulk) and monoclonal single-cell clone containing a homozygous deletion of BEN-2 (del/del). PCRs were run alongside a no-template control (NTC). **(D)** Transcript abundance of RCA genes in TAD 1 (*CR2* and *CD55*) and TAD 2 (*CD46*) were measured by qPCR. Values were normalised to the β-actin gene (*ACTB*) using the ΔCt method. Bars represent mean relative expression ± SEM from 3 biological replicates. Asterisks represent statistically significant differences between WT, bulk and del/del samples (*$p < 0.05$, **$p \leq 0.01$).

deletion of the BEN-2 region (**Figure 6C***)*. We confirmed the genotype of the del/del population using Sanger sequencing (**Supplementary Figure 6**).

Remarkably, the homozygous deletion of BEN-2 in Raji cells significantly decreased *CR2* transcript abundance by approximately 90% relative to WT levels ($p = 0.0034$, $n = 3$) and *CD55* transcript abundance by approximately 80% of WT levels ($p = 0.0039$, $n = 3$). We also measured transcript abundance of *CR2* and *CD55* in the bulk population where a proportion of cells contained the enhancer deletion. Accordingly, both transcripts were significantly decreased compared to WT but to a lesser extent than del/del population; *CR2* transcript

abundance decreased by approximately 70% of WT levels ($p = 0.0184$, $n = 3$) and *CD55* by approximately 50% of WT levels ($p = 0.0167$, $n = 3$) (**Figure 6D**). We also measured transcript abundance of *CD46*, which was not predicted to be targeted by BEN-2 on GeneHancer and localised to the neighbouring TAD (**Figure 3**). Enhancer deletion did not alter *CD46* transcript abundance (**Figure 6C**). These data confirmed that BEN-2 is a functional enhancer in B cells and demonstrated that BEN-2 regulates *CD55* and *CR2* within this cellular context.

As *CR2* and *CD55* transcript levels were significantly decreased with BEN-2 deletion in the Raji cell line, we next determined if surface protein expression of these receptors was

FIGURE 7 | CRISPR deletion of BEN-2 decreased surface expression of CR2 and CD55 in the Raji B cell line. **(A)** Cell surface expression of CR2 and CD55 protein was determined using flow cytometry. Cells were labelled with PE-conjugated CR2 antibody, PE-conjugated CD55 antibody or PE-conjugated IgG (isotype control) to confirm positive expression. Samples were run alongside unstained controls (not shown). For each sample, 10000 events were collected. Representative histograms from the WT, bulk and del/del samples stained with CR2-PE or CD55-PE, as well as unstained WT control. **(B)** Representative dot plot from the WT, bulk and del/del samples stained with CR2-PE or CD55-PE. The CR2- and CD55-positive gates were plotted using the unstained control. **(C)** Bars represent mean fluorescence intensity ± SEM from 3 biological replicates. For CD55, the geometric mean of the CD55-positive population was assessed. Asterisks represent statistically significant differences between WT and del/del samples (*$p < 0.05$).

concomitantly affected. We used flow cytometry to assess CR2 surface expression in the WT, bulk and del/del populations. Indeed, the cells in all populations were CR2-positive, but the surface expression of CR2 in the del/del population was observably decreased relative to the WT and bulk populations (**Figures 7A, B**). The geometric mean of the del/del population was significantly decreased by approximately 3-fold in the del/del population relative to the WT control (p = 0.0257, n = 3) (**Figure 7C**). Surface staining using a CD55 antibody showed approximately 2-3% of the population of Raji cells were CD55-positive (**Figures 7A, B**). Thus, we examined the geometric mean of the CD55-positive population and found that the overall CD55 surface expression was also significantly decreased to approximately 75% of WT levels (p = 0.074, n = 3) (**Figure 7C**). In line with transcript abundance data (**Figure 6D**), CR2 and CD55 surface expression in the bulk population was decreased to levels between WT and del/del; 72% and 90%, respectively, of WT levels. This confirmed that the reduction in CR2 and CD55 transcript expression with CRISPR deletion of BEN-2 reduced surface protein levels. We thus further validated that BEN-2 regulates these genes both at the level of mRNA and subsequent protein expression in a B cell context.

DISCUSSION

In this manuscript, we have explored the chromatin architecture of the RCA gene cluster, an important immunomodulatory region. We provide multiple insights into genes and variants within this cluster from a biological, evolutionary and disease perspective. We show for the first time the co-localisation of the RCA genes is in part due to the requirement of shared long-range regulatory elements. Using high-resolution 4C-seq maps in B cell models, we showed that several distal CTCF sites in the RCA gene cluster engage in chromatin looping, including the CTCF site modulated by the SLE-associated SNP rs1876453 (29). We further shine a light onto the enigma of the unmappable segmental duplication within *CR1* (24). Overall, we reveal extensive and complex mechanisms by which the RCA gene cluster may influence gene expression and thereby autoimmunity.

While Hi-C interaction data are typically sufficient to identify TADs across the genome, the utility of these data to examine the structural arrangement of the RCA gene cluster was hindered by the sequence repetitiveness of the intragenic segmental duplication in *CR1*. Our strategy combined high-resolution 4C-seq and a unique Hi-C pipeline (mHi-C) in conjunction with publicly available Hi-C datasets to uncover a highly intriguing TAD arrangement within the RCA gene cluster in the B cell lineage. We successfully utilised 4C-seq in two representative B cell lines from multiple viewpoints. Further, we were able to show consistent maps of CTCF interactions in the RCA gene cluster and chromatin looping which was constrained to one of two distinct regions across these loci. Importantly, when we examined two CTCF viewpoints within intron 1 and intron 23 of the *CR1* gene, we found that

interactions from these viewpoints were directionally constrained in opposite directions (upstream and downstream of *CR1*, respectively). These data indicated that the inter-TAD boundary was located within the *CR1* gene. We corroborated this finding using mHi-C, a novel Hi-C processing pipeline which assigns multi-mapping reads to the most likely genomic location. We thus resolved interactions at this previously unmappable element and refined the location of the inter-TAD boundary to intron 11 of *CR1*, falling within the segmental duplication. Previous studies have shown that TAD boundaries are enriched within housekeeping genes (31, 57) and may also be located near gene promoters (58). To our knowledge, a TAD boundary located well within the body of an expressed protein-coding gene has not been delineated to this resolution prior to our study.

The tandem segmental duplication in *CR1*, also known as 'low copy repeat 1' (LCR1), results in the duplication of 8 exons and introns in *CR1* and has been shown to alter the number of functional domains in the protein (59, 60). There are multiple co-dominant *CR1* alleles in the population defined by copies of LCR1. CR1-A/F (one copy of LCR1) and CR1-B/S (two copies of LCR1) alleles are most common. Alleles that contain zero copies and three copies of LCR1 have also been documented (61). Importantly, this repeat element is known to be associated with SLE and, more recently, late-onset Alzheimer's disease (24, 62). Despite the large size and nature of this repeat element, its biological implication is undefined. Our findings strongly indicate that the LCR1 repeat element in *CR1* co-localises with a TAD boundary and thus has a role in regulating gene expression in the RCA cluster. This is consistent with the presiding hypothesis that complex diseases develop as a result of dysregulated gene expression rather than functional protein changes. Based on the loop extrusion model of TAD formation, it is likely that the increasing copy number of LCR1 results in increased numbers of CTCF sites (reverse orientation) in TAD 1, thereby increasing insulation of chromatin interactions within this TAD (32). Experiments are currently on-going in our laboratory to establish whether the LCR1 copy number influences chromatin interactions and gene expression in the RCA cluster.

The RCA gene cluster is an exemplar gene cluster as its members are co-localised in the human genome and share protein structure and function. In addition, the genes, gene orientation and gene order of the RCA cluster are well conserved in many species across evolutionary time, including *Xenopus tropicalis* and chicken (63, 64). The RCA gene cluster in mice is also conserved but its members are separated across two chromosomal positions located more than 6 Mb apart (65). This matches closely to the TAD organisation of the gene cluster we describe here. It has been observed that breaks in synteny between species commonly occur at TAD boundaries (66), thus the TAD boundary we identified in *CR1* may represent the breakpoint region for the genomic rearrangement of the RCA gene cluster in humans and mice.

Our study revealed that the RCA gene cluster is divided into two TADs in the B cell lineage; TAD 1 consists of *C4BPB,*

C4BPA, CD55, CR2 and *CR1*, and TAD 2 consists of *CR1L* and *CD46*. This grouping does not reflect the organisation of active RCA genes in B cells (*C4BPB* and *C4BPA* are not expressed in B cells) or other recognised cell types. Nonetheless, our findings in the RCA gene cluster correspond to the *Six* (67), *HoxA* (68, 69) and *HoxD* (70, 71) homeobox gene clusters which have also been shown to be separated into two distinct TADs and regulatory regions. Unlike the aforementioned gene clusters, which have highly-restricted and distinct expression patterns (72), the RCA gene cluster is unique in that its members are expressed in numerous cell types and across various stages of cell development. Each member of the RCA has a distinct expression pattern; *CD55* and *CD46* are expressed across nearly all cell types, including non-immune cells (11), whereas *CR2* and *CR1* are predominantly expressed on B cells (73) and erythrocytes (74), respectively. Thus, it will be interesting to examine the RCA gene cluster in other cellular contexts to examine the TAD structure and enhancer-gene landscape and determine if these differ from the organisation we define in this study. Like GM12878, the TAD arrangement of the RCA gene cluster was unable to be explicitly resolved based on high-resolution Hi-C data in non-B cell lines (**Supplementary Figure 1**). Multiple strategies of experimentation, such as mHi-C and 4C-seq as we have utilised here, will be required to overcome this caveat.

Gene-gene interactions are a long-recognised and important contributor to complex disease susceptibility, often overlooked due to the difficulty in identifying such interactions (75). In this study, we mapped such interactions between genes in the RCA gene cluster at the molecular level using high-resolution chromatin interaction maps and have successfully identified direct chromatin interactions between CTCF sites in *CD55, CR2* and *CR1*. Binding of CTCF at intron 1 of *CR2* was modulated by an SLE-associated SNP (rs1876453) and shown to influence the expression of its neighbouring gene, *CR1*, in B cells (29). Our data now shows that *CR2* and *CR1* form part of the same CTCF-mediated chromatin network in the RCA gene cluster and support the hypothesis that these genes are co-regulated. It is possible that expression of *CR1* is also co-regulated by BEN-2 in TAD 1 as has been predicted by CHi-C in the GM12878 B cell line (56).

We also uncovered a direct relationship between *CR2* and *CD55*, showing that these genes are co-regulated by an intergenic enhancer, BEN-2. This marks the very first long-range regulatory element identified in this TAD and gene cluster. We characterised this enhancer through multiple lines of evidence *in vitro* and *in vivo* in several B cell models; by characterising transcriptional activity and chromatin marks. Importantly, BEN-2 upregulated transcriptional activity across a panel of B cell lines and reduced it in the non-B HepG2 cell line in our reporter gene assays. This indicates that B cell-specific factors are largely involved in BEN-2 transcription functioning. Further, BEN-2 influenced the transcriptional activity of the luciferase reporter to differing extents across the B cell lines tested. This may be attributed to by the different representative stages of each cell line; namely pre-B (Reh), mature B (Raji and B-0028) and

terminally differentiated plasma cells (SKW) (76–78). Alternatively, the Epstein-Barr virus (EBV) infection status of the B cell lines may account for these differences. EBV infection is one of the most common methods for producing immortalised lymphoblastoid cell lines, with its viral factors predominantly binding to distal enhancer elements (79). In fact, CR2 is an EBV receptor on B cells (80). BEN-2 had the highest transcriptional effect in the EBV-positive plasma cell line SKW, suggesting that this enhancer may cause the strongest upregulation of *CR2* in the latter stages of development. Additionally, is it possible that EBV may influence the activity of the BEN-2 enhancer; a prospect our lab is currently investigating.

To validate the bioinformatic and *in vitro* data, we also utilised CRISPR genomic deletion and observed significant decreases in the mRNA expression of *CR2* and *CD55*. The reduction in transcript level mediated by this enhancer deletion produced concomitant reduction in surface protein level of both CR2 and CD55, corroborating our qPCR results. In consideration of having one single cell clone containing a complete homozygous deletion of BEN-2 to examine, we also utilised the polyclonal, bulk population generated through FACS, in which a proportion of cells contained the full enhancer deletion, as an additional experimental sample. Importantly, qPCR and flow cytometry of this sample supported our findings, as expression of CR2 and CD55 were consistently at an intermediate level between WT and del/del samples. However, we note some limitations in the cell line utilised in this study. The Raji cell line did not express *CR1* and the surface expression of CD55 was heterogenous within the population as assessed by flow cytometry. Distinct clonal populations of Raji cells with differing CD55 surface protein expression have been previously reported (81). Importantly, a CD55-negative clone of Raji was found to have detectable levels of *CD55* mRNA, suggesting that this clone may have altered translational or post-translational mechanisms causing loss of CD55 surface expression (81). Regardless, in conjunction with the putative bioinformatic data for BEN, our experimental data revealed a role for BEN-2 in the expression of these complement regulators and improve our understanding of the complex transcriptional control of these genes. Based on bioinformatic data, it is possible that BEN-2 may also be active in other immune cell types, including neutrophils, macrophages and dendritic cells, (**Supplementary Figure 7**) and control the expression of the RCA cluster members in other cellular contexts.

The surface expression of CD55 (82, 83) and CR2 (19, 20) on B cells is significantly decreased in SLE patients. Genetic variation, such as SNPs, in the BEN-2 region may influence expression of both *CR2* and *CD55* in tandem, thereby exacerbating the effect of variants in their contributions to autoimmunity. The genes of the RCA cluster may also be co-regulated by the other candidate B cell enhancers we identified here. Strategies to map gene-gene interactions and define relationships between genes such as those utilised in this study may open important avenues to better understand how complex diseases, like SLE and Alzheimer's, are influenced by genetic variation in this region.

We have established for the first time that the RCA gene cluster is transcriptionally co-regulated and comprises of a complex network of enhancer-gene and gene-gene interactions. We have also defined the regulatory architecture of the RCA in the B cell lineage, revealing novel mechanisms by which the RCA gene cluster is controlled and expanding the scope for future investigations in the context of evolution, immunity and complex genetic disease.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**. 4C-seq data were deposited in the Gene Expression Omnibus (GEO) database under accession number GSE140127. The mHi-C pipeline can be accessed at https://github.com/yezhengSTAT/mHiC.

AUTHOR CONTRIBUTIONS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.901747/full#supplementary-material

REFERENCES

1. Holers VM. Complement and Its Receptors: New Insights Into Human Disease. *Annu Rev Immunol* (2014) 32(1):433–59. doi: 10.1146/annurev-immunol-032713-120154
2. Carroll MC, Isenman DE. Regulation of Humoral Immunity by Complement. *Immunity* (2012) 37(2):199–207. doi: 10.1016/j.immuni.2012.08.002
3. West EE, Kolev M, Kemper C. Complement and the Regulation of T Cell Responses. *Annu Rev Immunol* (2018) 36(1):309–38. doi: 10.1146/annurev-immunol-042617-053245
4. Ricklin D, Reis ES, Lambris JD. Complement in Disease: A Defence System Turning Offensive. *Nat Rev Nephrol* (2016) 12(7):383–401. doi: 10.1038/nrneph.2016.70
5. Hourcade D, Holers VM, Atkinson JP. The Regulators of Complement Activation (RCA) Gene Cluster. *Adv Immunol* (1989) 45:381–416. doi: 10.1016/S0065-2776(08)60697-5
6. Hourcade D, Garcia AD, Post TW, Taillon-Miller P, Holers VM, Wagner LM, et al. Analysis of the Human Regulators of Complement Activation (RCA) Gene Cluster With Yeast Artificial Chromosomes (YACs). *Genomics* (1992) 12(2):289–300. doi: 10.1016/0888-7543(92)90376-4
7. Pardo-Manuel de Villena F, Rodríguez de Córdoba S. C4BPAL2: A Second Duplication of the C4BPA Gene in the Human RCA Gene Cluster. *Immunogenetics* (1995) 41(2-3):139–43. doi: 10.1007/BF00182326
8. Logar CM, Chen W, Schmitt H, Yung C, Birmingham DJ. A Human CR1-Like Transcript Containing Sequence for a Binding Protein for Ic4 is Expressed in Hematopoietic and Fetal Lymphoid Tissue. *Mol Immunol* (2004) 40:831–40. doi: 10.1016/j.molimm.2003.09.010
9. Ermert D, Blom AM. C4b-Binding Protein: The Good, the Bad and the Deadly. Novel Functions of an Old Friend. *Immunol Lett* (2016) 169:82–92. doi: 10.1016/j.imlet.2015.11.014
10. Dho SH, Lim JC, Kim LK. Beyond the Role of CD55 as a Complement Component. *Immune Netw* (2018) 18(1):1–13. doi: 10.4110/in.2018.18.e11
11. Zipfel PF, Skerka C. Complement Regulators and Inhibitory Proteins. *Nat Rev Immunol* (2009) 9(10):729–40. doi: 10.1038/nri2620
12. Liszewski MK, Atkinson JP. Complement Regulator CD46: Genetic Variants and Disease Associations. *Hum Genomics* (2015) 9:7. doi: 10.1186/s40246-015-0029-z
13. Erdei A, Isaák A, Török K, Sándor N, Kremlitzka M, Prechl J, et al. Expression and Role of CR1 and CR2 on B and T Lymphocytes Under Physiological and Autoimmune Conditions. *Mol Immunol* (2009) 46(14):2767–73. doi: 10.1016/j.molimm.2009.05.181
14. Deng Y, Tsao BP. Genetic Susceptibility to Systemic Lupus Erythematosus in the Genomic Era. *Nat Rev Rheumatol* (2010) 6(12):683–92. doi: 10.1038/nrrheum.2010.176
15. Dempsey PW, Allison MED, Akkaraju S, Christopher C, Fearon DT, Dempsey PW, et al. C3d of Complement as a Molecular Adjuvant: Bridging Innate and Acquired Immunity. *Science* (1996) 271(5247):348–50. doi: 10.1126/science.271.5247.348
16. Jozsi M, Prechl J, Bajtay Z, Erdei A. Complement Receptor Type 1 (CD35) Mediates Inhibitory Signals in Human B Lymphocytes. *J Immunol* (2002) 168(6):2782–8. doi: 10.4049/jimmunol.168.6.2782
17. Andrews BS, Eisenberg RA, Theofilopoulos AN, Izui S, Wilson CB, McConahey PJ, et al. Spontaneous Murine Lupus-Like Syndromes. Clinical

and Immunopathological Manifestations in Several Strains. *J Exp Med* (1978) 148(5):1198–215. doi: 10.1084/jem.148.5.1198

18. Theofilopoulos AN, Dixon FJ. Murine Models of Systemic Lupus Erythematosus. *Adv Immunol* (1985) 37:269–390. doi: 10.1016/S0065-2776 (08)60342-9

19. Marquart HV, Svendsen A, Rasmussen JM, Nielsen CH, Junker P, Svehag SE, et al. Complement Receptor Expression and Activation of the Complement Cascade on B Lymphocytes From Patients With Systemic Lupus Erythematosus (SLE). *Clin Exp Immunol* (1995) 101(1):60–5. doi: 10.1111/j.1365-2249.1995.tb02277.x

20. Wilson JG, Ratnoff WD, Schur PH, Fearon DT. 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* (1986) 29(6):739–47. doi: 10.1002/art.1780290606

21. Ahearn JM, Fischer MB, Croix D, Goerg S, Ma M, Xia J, 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* (1996) 4(3):251–62. doi: 10.1016/S1074-7613(00)80433-1

22. Marchbank KJ, Kulik L, Gipson MG, Morgan BP, Holers VM. 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* (2002) 169(7):3526–35. doi: 10.4049/jimmunol.169.7.3526

23. Molina H, Holers VM, Li B, Y-f F, Mariathasan S, Goellner J, et al. Markedly Impaired Humoral Immune Response in Mice Deficient in Complement Receptors 1 and 2. *Proc Natl Acad Sci* (1996) 93(8):3357–61. doi: 10.1073/pnas.93.8.3357

24. Nath SK, Harley JB, Lee YH. Polymorphisms of Complement Receptor 1 and Interleukin-10 Genes and Systemic Lupus Erythematosus: A Meta-Analysis. *Hum Genet* (2005) 118(2):225–34. doi: 10.1007/s00439-005-0044-6

25. Boackle SA, Holers VM, Chen X, Szakonyi G, Karp DR, Wakeland EK, et al. Cr2, a Candidate Gene in the Murine Sle1c Lupus Susceptibility Locus, Encodes a Dysfunctional Protein. *Immunity* (2001) 15(5):775–85. doi: 10.1016/S1074-7613(01)00228-X

26. Morel L, Croker BP, Blenman KR, Mohan C, Gilkeson G, Wakeland EK, et al. Genetic Reconstitution of Systemic Lupus Erythematosus Immunopathology With Polygenic Murine Strains. *Proc Natl Acad Sci* (2000) 97(12):6670–5. doi: 10.1073/pnas.97.12.6670

27. Wu H, Boackle SA, Hanvivadhanakul P, Ulgiati D, Grossman M, Lee Y, et al. Association of a Common Complement Receptor 2 Haplotype With Increased Risk of Systemic Lupus Erythematosus. *Proc Natl Acad Sci* (2007) 104 (10):3961–6. doi: 10.1073/pnas.0609101104

28. Douglas KB, Windels DC, Zhao J, Gadeliya AV, Wu H, Kaufman KM, et al. Complement Receptor 2 Polymorphisms Associated With Systemic Lupus Erythematosus Modulate Alternative Splicing. *Genes Immun* (2009) 10:457–69. doi: 10.1038/gene.2009.27

29. Zhao J, Giles BM, Taylor RL, Yette GA, Lough KM, Ng HL, et al. Preferential Association of a Functional Variant in Complement Receptor 2 With Antibodies to Double-Stranded DNA. *Ann Rheumatic Dis* (2016) 75 (1):242–52. doi: 10.1136/annrheumdis-2014-205584

30. Ong C-T, Corces VG. CTCF: An Architectural Protein Bridging Genome Topology and Function. *Nat Rev Genet* (2014) 15(4):234–46. doi: 10.1038/nrg3663

31. Dixon JR, Gorkin DU, Ren B. Chromatin Domains: The Unit of Chromosome Organization. *Mol Cell* (2016) 62(5):668–80. doi: 10.1016/j.molcel.2016.05.018

32. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* (2016) 15 (9):2038–49. doi: 10.1016/j.celrep.2016.04.085

33. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* (2012) 485(7398):376–80. doi: 10.1038/nature11082

34. Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* (2009) 326(5950):289–93. doi: 10.1126/science.1181369

35. Andrey G, Mundlos S. The Three-Dimensional Genome: Regulating Gene Expression During Pluripotency and Development. *Development* (2017) 144 (20):3646–58. doi: 10.1242/dev.148304

36. Shlyueva D, Stampfel G, Stark A. Transcriptional Enhancers: From Properties to Genome-Wide Predictions. *Nat Rev Genet* (2014) 15(4):272–86. doi: 10.1038/nrg3682

37. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* (2012) 337(6099):1190–5. doi: 10.1126/science.1222794

38. Andersson R, Gebhard C, Miguel-escalada I, Hoof I, Bornholdt J, Boyd M, et al. An Atlas of Active Enhancers Across Human Cell Types and Tissues. *Nature* (2014) 507(7493):455–61. doi: 10.1038/nature12787

39. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac Separates Active From Poised Enhancers and Predicts Developmental State. *Proc Natl Acad Sci USA* (2010) 107(50):21931–6. doi: 10.1073/pnas.1016071107

40. Marti-Renom MA, Almouzni G, Bickmore WA, Bystricky K, Cavalli G, Fraser P, et al. Challenges and Guidelines Toward 4D Nucleome Data and Model Standards. *Nat Genet* (2018) 50(10):1352–8. doi: 10.1038/s41588-018-0236-3

41. Fernández-Miñán A, Bessa J, Tena JJ, Gómez-Skarmeta JL. Chapter 21 - Assay for Transposase-Accessible Chromatin and Circularized Chromosome Conformation Capture, Two Methods to Explore the Regulatory Landscapes of Genes in Zebrafish. *Methods Cell Biol* (2016) 135:413–30. doi: 10.1016/bs.mcb.2016.02.008

42. Splinter E, de Wit E, van de Werken HJG, Klous P, de Laat W. Determining Long-Range Chromatin Interactions for Selected Genomic Sites Using 4C-Seq Technology: From Fixation to Computation. *Methods* (2012) 58(3):221–30. doi: 10.1016/j.ymeth.2012.04.009

43. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* (2014) 159(7):1665–80. doi: 10.1016/j.cell.2014.11.021

44. Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, et al. The 3d Genome Browser: A Web-Based Browser for Visualizing 3D Genome Organization and Long-Range Chromatin Interactions. *Genome Biol* (2018) 19(1):151. doi: 10.1186/s13059-018-1519-9

45. Robinson JT, Turner D, Durand NC, Thorvaldsdóttir H, Mesirov JP, Aiden EL. Juicebox. Js Provides a Cloud-Based Visualization System for Hi-C Data. *Cell Syst* (2018) 6(2):256–8.e1. doi: 10.1016/j.cels.2018.01.001

46. Ziebarth JD, Bhattacharya A, Cui Y. CTCFBSDB 2.0: A Database for CTCF-Binding Sites and Genome Organization. *Nucleic Acids Res* (2013) 41:D188–94. doi: 10.1093/nar/gks1165

47. Zheng Y, Ay F, Keles S. Generative Modeling of Multi-Mapping Reads With mHi-C Advances Analysis of Hi-C Studies. *eLife* (2019) 8:e38070. doi: 10.7554/eLife.38070

48. Cresswell KG, Stansfield JC, Dozmorov MG. SpectralTAD: An R Package for Defining a Hierarchy of Topologically Associated Domains Using Spectral Clustering. *BMC Bioinf* (2020) 21(319):1–19. doi: 10.1186/s12859-020-03652-w

49. Cruickshank M, Fenwick E, Abraham LJ, Ulgiati D. Quantitative Differences in Chromatin Accessibility Across Regulatory Regions can be Directly Compared in Distinct Cell-Types. *Biochem Biophys Res Commun* (2008) 367(2):349–55. doi: 10.1016/j.bbrc.2007.12.121

50. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome Engineering Using the CRISPR-Cas9 System. *Nat Protoc* (2013) 8(11):2281–308. doi: 10.1038/nprot.2013.143

51. Moreno-Mateos MA, Vejnar CE, Beaudoin J-d, Fernandez JP, Mis EK, Khokha MK, et al. CRISPRscan: Designing Highly Efficient sgRNAs for CRISPR-Cas9 Targeting In Vivo. *Nat Methods* (2015) 12(10):982–8. doi: 10.1038/nmeth.3543

52. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Stein TI, et al. GeneHancer: Genome-Wide Integration of Enhancers and Target Genes in GeneCards. *Database* (2017) 2017:bax028. doi: 10.1093/database/bax028

53. Heinz S, Romanoski CE, Benner C, Glass CK. The Selection and Function of Cell Type-Specific Enhancers. *Nat Rev Mol Cell Biol* (2015) 16(3):144–54. doi: 10.1038/nrm3949

54. Busslinger M. Transcriptional Control of Early B Cell Development. *Annu Rev Immunol* (2004) 22(1):55–79. doi: 10.1146/annurev.immunol.22.012703.104807

55. Cobaleda C, Schebesta A, Delogu A, Busslinger M. Pax5: The Guardian of B Cell Identity and Function. *Nat Immunol* (2007) 8(5):463–70. doi: 10.1038/ni1454

56. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, et al. Mapping Long-Range Promoter Contacts in Human Cells With High-Resolution Capture Hi-C. *Nat Genet* (2015) 47(6):598–606. doi: 10.1038/ng.3286

57. Ulianov SV, Khrameeva EE, Gavrilov AA, Flyamer IM, Kos P, Mikhaleva EA, et al. Active Chromatin and Transcription Play a Key Role in Chromosome Partitioning Into Topologically Associating Domains. *Genome Res* (2016) 26:70–84. doi: 10.1101/gr.196006.115

58. Paliou C, Guckelberger P, Schöpflin R, Heinrich V, Esposito A, Chiariello AM, et al. Preformed Chromatin Topology Assists Transcriptional Robustness of Shh During Limb Development. *Proc Natl Acad Sci* (2019) 116:1–10. doi: 10.1073/pnas.1900672116

59. Crehan H, Holton P, Wray S, Pocock J, Guerreiro R, Hardy J. Complement Receptor 1 (CR1) and Alzheimer's Disease. *Immunobiology* (2012) 217(2):244–50. doi: 10.1016/j.imbio.2011.07.017

60. Wong WW. Structural and Functional Correlation of the Human Complement Receptor Type 1. *J Invest Dermatol* (1990) 94(6 Suppl):64S–7S. doi: 10.1111/1523-1747.ep12875150

61. Moulds JM, Reveille JD, Arnett FC. Structural Polymorphisms of Complement Receptor 1 (CR1) in Systemic Lupus Erythematosus (SLE) Patients and Normal Controls of Three Ethnic Groups. *Clin Exp Immunol* (1996) 105(2):302–5. doi: 10.1046/j.1365-2249.1996.d01-748.x

62. Kucukkilic E, Brookes K, Barber I, Guetta-Baranes T, Morgan K, Hollox EJ. Complement Receptor 1 Gene (CR1) Intragenic Duplication and Risk of Alzheimer's Disease. *Hum Genet* (2018) 137(4):305–14. doi: 10.1007/s00439-018-1883-2

63. Oshiumi H, Shida K, Goitsuka R, Kimura Y, Katoh J, Ohba S, et al. Regulator of Complement Activation (RCA) Locus in Chicken: Identification of Chicken RCA Gene Cluster and Functional RCA Proteins. *J Immunol* (2005) 175(3):1724–34. doi: 10.4049/jimmunol.175.3.1724

64. Oshiumi H, Suzuki Y. Regulator of Complement Activation (RCA) Gene Cluster in Xenopus Tropicalis. *Immunogenetics* (2009) 61:371–84. doi: 10.1007/s00251-009-0368-9

65. Kingsmore BYSF, Vikj DP, Kurtz CB, Leroy P, Tack IIBF, Weis JH, et al. Receptor-Related Genes in the Mouse. *J Exp Med* (1989) 169(4):1479–84. doi: 10.1084/jem.169.4.1479

66. Szabo Q, Bantignies F, Cavalli G. Principles of Genome Folding Into Topologically Associating Domains. *Sci Adv* (2019) 5(4):eaaw1668–eaaw. doi: 10.1126/sciadv.aaw1668

67. Gómez-Marín C, Tena JJ, Acemel RD, López-Mayorga M, Naranjo S, de la Calle-Mustienes E, et al. Evolutionary Comparison Reveals That Diverging CTCF Sites are Signatures of Ancestral Topological Associating Domains Borders. *Proc Natl Acad Sci* (2015) 112(24):7542–7. doi: 10.1073/pnas.1505463112

68. Berlivet S, Paquette D, Dumouchel A, Langlais D, Dostie J, Kmita M. Clustering of Tissue-Specific Sub-TADs Accompanies the Regulation of HoxA Genes in Developing Limbs. *PLoS Genet* (2013) 9(12):e1004018–e. doi: 10.1371/journal.pgen.1004018

69. Rousseau M, Crutchley JL, Miura H, Suderman M, Blanchette M, Dostie J. Hox in Motion: Tracking HoxA Cluster Conformation During Differentiation. *Nucleic Acids Res* (2014) 42(3):1524–40. doi: 10.1093/nar/gkt998

70. Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, et al. A Switch Between Topological Domains Underlies HoxD Genes Collinearity in Mouse Limbs. *Science* (2013) 340(6137):123167–. doi: 10.1126/science.1234167

71. Rodríguez-Carballo E, Lopez-Delisle L, Zhan Y, Fabre PJ, Beccari L, El-Idrissi I, et al. The HoxD Cluster Is a Dynamic and Resilient TAD Boundary Controlling the Segregation of Antagonistic Regulatory Landscapes. *Genes Dev* (2017) 31(22):2264–81. doi: 10.1101/gad.307769.117

72. Montavon T, Duboule D. Landscapes and Archipelagos: Spatial Organization of Gene Regulation in Vertebrates. *Trends Cell Biol* (2012) 22(7):347–54. doi: 10.1016/j.tcb.2012.04.003

73. Weis JJ, Tedder TF, Fearon DT. Identification of a 145,000 Mr Membrane Protein as the C3d Receptor (CR2) of Human B Lymphocytes. *Proc Natl Acad Sci USA* (1984) 81(3):881–5. doi: 10.1073/pnas.81.3.881

74. Pascual M, Schifferli JA. The Binding of Immune Complexes by the Erythrocyte Complement Receptor 1 (CR1). *Immunopharmacology* (1992) 24(2):101–6. doi: 10.1016/0162-3109(92)90016-6

75. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the Missing Heritability of Complex Diseases. *Nature* (2009) 461(7265):747–53. doi: 10.1038/nature08494

76. Koziner B, Stavnezer J, Al-Katib A, Gebhard D, Mittelman A, Andreeff M, et al. Surface Immunoglobulin Light Chain Expression in Pre-B Cell Leukemias. *Ann New Y Acad Sci* (1986) 468(1):211–26. doi: 10.1111/j.1749-6632.1986.tb42041.x

77. Benjamin D, Magrath IT, Maguire R, Janus C, Todd HD, Parsons RG. Immunoglobulin Secretion by Cell Lines Derived From African and American Undifferentiated Lymphomas of Burkitt's and non-Burkitt's Type. *J Immunol* (1982) 129(3):1336–42.

78. Ralph P, Saiki O, Welte K, Maurer DH, Welte K. IgM and IgG Secretion in Human B-Cell Lines Regulated by B-Cell-Inducing Factors (BIF) and Phorbol Ester. *Immunol Lett* (1983) 7(1):17–23. doi: 10.1016/0165-2478(83)90049-4

79. Jiang S, Zhou H, Liang J, Gerdt C, Wang C, Ke L, et al. The Epstein-Barr Virus Regulome in Lymphoblastoid Cells. *Cell Host Microbe* (2017) 22(4):561–73.e4. doi: 10.1016/j.chom.2017.09.001

80. Fingeroth JD, Weis JJ, Tedder TF, Strominger JL, Biro PA, Fearon DT. Epstein-Barr Virus Receptor of Human B Lymphocytes is the C3d Receptor CR2. *Proc Natl Acad Sci USA* (1984) 81(14):4510–4. doi: 10.1073/pnas.81.14.4510

81. Harris CL, Morgan BP. Characterization of a Glycosyl-Phosphatidylinositol Anchor-Deficient Subline of Raji Cells. An Analysis of the Functional Importance of Complement Inhibitors on the Raji Cell Line. *Immunology* (1995) 86(2):311–8.

82. Alegretti AP, Mucenic T, Merzoni J, Faulhaber GA, Silla LM, Xavier RM. Expression of CD55 and CD59 on Peripheral Blood Cells From Systemic Lupus Erythematosus (SLE) Patients. *Cell Immunol* (2010) 265(2):127–32. doi: 10.1016/j.cellimm.2010.07.013

83. García-Valladares I, Atisha-Fregoso Y, Richaud-Patin Y, Jakez-Ocampo J, Soto-Vega E, Elías-López D, et al. Diminished Expression of Complement Regulatory Proteins (CD55 and CD59) in Lymphocytes From Systemic Lupus Erythematosus Patients With Lymphopenia. *Lupus* (2006) 15(9):600–5. doi: 10.1177/0961203306071916