Accepted Manuscript

Title: Analysis of tandem E-box motifs within human *Complement receptor 2 (CR2/CD21)* promoter reveals cell specific roles for RP58, E2A, USF and localized chromatin accessibility

Author: Mark N. Cruickshank James Dods Rhonda L. Taylor
Mahdad Karimi Emily J. Fenwick Elizabeth A. Quail
Alexander J. Rea V. Michael Holers Lawrence J. Abraham
Daniela Ulgiati

PII: S1357-2725(15)00085-0
DOI: <http://dx.doi.org/doi:10.1016/j.biocel.2015.03.016>
Reference: BC 4588

To appear in: *The International Journal of Biochemistry & Cell Biology*

Received date: 28-1-2015
Revised date: 3-3-2015
Accepted date: 18-3-2015

Please cite this article as: Cruickshank, M. N., Dods, J., Taylor, R. L., Karimi, M., Fenwick, E. J., Quail, E. A., Rea, A. J., Holers, V. M., Abraham, L. J., and Ulgiati, D., Analysis of tandem E-box motifs within human *Complement receptor 2 (CR2/CD21)* promoter reveals cell specific roles for RP58, E2A, USF and localized chromatin accessibility, *International Journal of Biochemistry and Cell Biology* (2015), <http://dx.doi.org/10.1016/j.biocel.2015.03.016>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Analysis of tandem E-box motifs within human *Complement receptor 2 (CR2/CD21)* promoter reveals cell specific roles for RP58, E2A, USF and localized chromatin accessibility

Mark N. Cruickshank¹, James Dods², Rhonda L. Taylor^{2,3}, Mahdad Karimi², Emily J. Fenwick², Elizabeth A. Quail², Alexander J. Rea³, V. Michael Holers⁴, Lawrence J. Abraham³ and Daniela Ulgiati³

*Running Title: *Regulation of CR2/CD21 by E-box binding factors.*

¹Telethon Kids Institute, 100 Roberts Road, Subiaco, Western Australia, ²Biochemistry and Molecular Biology, School of Chemistry and Biochemistry, The University of Western Australia, Crawley, Western Australia; ³School of Pathology and Laboratory Medicine, Centre for Genetic Origins of Health and Disease, The University of Western Australia; ⁴Department of Medicine, University of Colorado, School of Medicine, Aurora, CO.

To whom correspondence should be addressed: Daniela Ulgiati, School of Pathology and Laboratory Medicine, University of Western Australia, M Block QEII, Crawley, W.A. 6009 Australia. Fax: +61 8 9346 2891; Phone: +61 8 9346 1076; E-mail: daniela.ulgiati@uwa.edu.au

Abbreviations: 2DE, Two Dimensional Electrophoresis; bHLH, basic-helix-loop-helix; CAGE, Capped analysis of gene-expression; CHART-PCR, Chromatin accessibility measured by real-time PCR; ChIP, Chromatin immunoprecipitation; EMSA, electrophoretic mobility shift assay; IEF, isoelectric focusing; MALDI-TOF MS, Matrix-assisted laser desorption/ionization time of flight mass spectrometry; NE, nuclear extract; TF, transcription factor.

27

28 **Abstract**

29 Complement receptor 2 (CR2/CD21) plays an important role in the generation of normal B
30 cell immune responses. As transcription appears to be the prime mechanism *via* which surface
31 CR2/CD21 expression is controlled, understanding transcriptional regulation of this gene will
32 have broader implications to B cell biology. Here we report opposing, cell-context specific
33 control of *CR2/CD21* promoter activity by tandem E-box elements, spaced 22bp apart and
34 within 70bp of the transcription initiation site. We have identified E2A and USF transcription
35 factors as binding to the distal and proximal E-box sites respectively in CR2-positive B-cells,
36 at a site that is hypersensitive to restriction enzyme digestion compared to non-expressing
37 K562 cells. However, additional unidentified proteins have also been found to bind these
38 functionally important elements. By utilizing a proteomics approach we have identified a
39 repressor protein, RP58, binding the distal E-box motif. Co-transfection experiments using
40 RP58 overexpression constructs demonstrated a specific 10-fold repression of *CR2/CD21*
41 transcriptional activity mediated through the distal E-box repressor element. Taken together,
42 our results indicate that repression of the *CR2/CD21* promoter can occur through one of the E
43 box motifs *via* recruitment of RP58 and other factors to bring about a silenced chromatin
44 context within CR2/CD21 non-expressing cells.

45

46 **Keywords:** Human, B cells, Gene regulation, Transcription Factors.

47

48

48

49 **1.1 Introduction**

50 The human cell surface receptor Complement Receptor 2 (CR2/CD21) shows a restricted
 51 tissue distribution and is expressed on sub-populations of B cells (Alvarez et al.), follicular
 52 dendritic cells (Reynes et al., 1985), T cells (Levy et al., 1992), epithelial cells (Levine and
 53 Davidson, 2005), basophils (Bacon et al., 1993), mast cells (Gommerman et al., 2000) and
 54 keratinocytes (Hunyadi et al., 1991). During B-lymphopoiesis, CR2/CD21 is first observed
 55 on newly formed immature B cells in the bone marrow (Takahashi et al., 1997, Tedder et al.,
 56 1984) and is up-regulated following negative selection and migration to the periphery as
 57 transitional intermediates (Melamed et al., 1998, Melamed and Nemazee, 1997, Hartley et al.,
 58 1993). CR2/CD21 is undetectable during early stages of development and upon terminal
 59 differentiation into antibody secreting plasma cells.

60

61 Since the developmental timing and level of *CR2/CD21* transcription is crucial for generating
 62 a normal B cell response, we have conducted an analysis of factors regulating the human
 63 *CR2/CD21* proximal promoter. We have previously characterized the basal requirements for
 64 transcription of the *CR2/CD21* gene and identified tandem E-box elements that are crucial for
 65 cell- and B cell stage-specific expression (Ulgiati et al., 2002, Ulgiati and Holers, 2001). E-
 66 boxes often serve as binding sites for E-proteins, including E2A (E12 and E47 isoforms
 67 encoded by the *TCF3* gene), E2-2 (E2-2A and E2-2B isoforms) and HEB proteins.

68

69 E-proteins comprise the class I basic helix-loop-helix (bHLH) transcription factors (TFs)
 70 which are widely expressed in different cell-types (Massari and Murre, 2000) and are known
 71 to be crucial for regulating various aspects of B cell development (Kwon et al., 2008). E-
 72 proteins interact with E-boxes (with a consensus core CANNTG) as homodimers or
 73 heterodimers with other bHLH family proteins and can activate transcription by recruitment
 74 of coactivators p300 and CBP (Guo et al., 2009). While E-proteins are promiscuously
 75 expressed (though not ubiquitous), a wide variety of genes controlled by E-box motifs are

expressed in a highly restricted developmental pattern. Tissue specific interactions with distinct classes of HLH-containing proteins plays a key role in regulating the target preference of E proteins (Murre, 2005); however, the mechanisms preventing wide-spread spurious binding of homodimers are incompletely understood.

Here we investigate tandem E box motifs, spaced 22 bp apart and within 70 bp of the transcription initiation site, revealing a critical role in controlling *CR2/CD21* transcription. These two motifs show opposing function and appear only to be active depending on the cellular context. Furthermore, we have identified E2A transcription factors as binding to one of the E box motifs using EMSA and ChIP analyses; however, as yet unidentified proteins are also binding these functionally important elements. Utilizing a proteomics approach to identify unknown components interacting with E box site 2 motifs, we successfully isolated and identified RP58 (encoded by *RP58/ZNF238/ZBTB18*); and reconstituted its binding activity *in vitro*. Further, *in vitro* transfection studies confirmed repression specifically through the E box site 2 repressor element. Conversely, CR2-expression in B cells was associated with E2A and USF binding via E box sites 1 and 2 respectively and localized chromatin hypersensitivity.

2. Materials and Methods

2.1 Cell culture

Cell cultures were maintained at 37°C with 5% CO₂ in RPMI-1640 supplemented with 10% FCS, 50 U/ml penicillin, 50 µg/ml streptomycin and 4 mM L-Glutamine. Cell lines were obtained from the ATCC (Manassas, VA, USA) including Raji (CCL-86) and K562 (CCL-243).

2.2 B cell isolation from peripheral blood

Peripheral blood was obtained from healthy donors and the PBMC fraction was isolated by Ficoll gradient centrifugation using standard protocols. B cells were purified by positive selection using MACS CD19 MicroBeads (Miltenyi Biotec, Australia, Cat#130-050-301) according to the manufacturer's instructions.

2.3 Construction of *CR2/CD21* mutant reporter constructs

A *NheI/XhoI* fragment of the *CR2/CD21* promoter containing the -315/+75 of the *CR2/CD21* promoter was cloned into the luciferase reporter pGL3-basic vector (Clontech Laboratories, Palo Alto, CA, USA) as described previously (Ulgiati et al., 2002). Site-directed mutagenesis was performed using the QuikChange mutagenesis kit (Stratagene, La Jolla, CA, USA). Correct orientation and sequence of constructs was verified by restriction enzyme digestion and nucleotide sequence analysis.

2.4 Transfection and quantitation of promoter activity

Transfections were performed by harvesting cells (Raji or K562) grown to log phase using the Superfect reagent (QIAGEN, Valencia, CA, USA) according to the manufacturer's specifications with plasmid DNA prepared using Endofree Maxiprep-500 columns (QIAGEN). Briefly, 4 µg of plasmid DNA and 120 ng of pRL-TK (*Renilla*) control vector were complexed in combination with Superfect reagent for 10 min at room temperature. The transfection complexes were added to suspensions of cells in 24-well plates to a final density of 5×10^5 cells/ml. Cell lysates from the transfected cells were prepared and assayed for both *Firefly* and *Renilla* luciferase using the Dual-Luciferase Reporter Assay system according to the manufacturer's instructions (Promega, Madison, WI, USA). All transfection data shown are the mean of 3–5 independent transfections performed in triplicate. Promoter activity is expressed as relative *Firefly* luciferase activity normalized against *Renilla* luciferase activity.

129 2.5 *Nuclear Extract (NE) Preparation and Sephacryl S-300 Gel Filtration*

130 5 to 10 L of K562 culture were harvested at a density of 8×10^5 cells/ml in 250 ml centrifuge
 131 bottles by centrifugation at 3500 rpm for 10 minutes at room temperature with a Beckman J2-
 132 MC Centrifuge (JA-14 rotor; Beckman Instruments). NE were prepared by scaling a
 133 previously described method (Li et al., 1991), were snap frozen in liquid nitrogen and stored
 134 at -80°C or proceeded directly to ammonium sulphate precipitation. Crude nuclear proteins
 135 were first purified by cutting in 0.33 g of ammonium sulphate per ml of NE as previously
 136 described (Marshak, 1996). NE was further purified using HiPrep[®] Sephacryl S-300 High
 137 Resolution columns (Amersham Biosciences, USA) in accordance with the manufacturer's
 138 instructions. The fractions were tested for protein content (Bio-Rad, USA) and binding
 139 activity, using EMSA as described (Woo et al., 2002), snap frozen in liquid nitrogen and
 140 stored at -80°C .

141

142 2.6 *Electrophoretic Mobility Shift Assay (EMSA)*

143 Oligonucleotides representing the sense and anti-sense strands of the promoter region of
 144 interest were synthesised with a non-native 5' G-overhang added to allow ^{32}P end-labeling.
 145 Purified double-stranded oligonucleotides were ^{32}P labeled by filling the 5' G-overhang with
 146 ^{32}P -dCTP (Amersham Pharmacia Biotech, Little Chalfont, England) using the Klenow
 147 enzyme (Promega Corp., Madison, WI, USA). 1-5 μg of crude NEs or 0.1-1 μg of Sephacryl
 148 S-300 Gel Filtrated fractions were pre-incubated for 10 minutes on ice with 1 μg (0.5 μg
 149 where indicated) of poly dI-dC (poly [dI-dC].poly [dI-dC]; Sigma-Aldrich) in a final reaction
 150 volume of 20 μl which also consists of 4% Ficoll, 20 mM HEPES (pH 7.9), 1 mM EDTA, 1
 151 mM DTT and 50 mM KCl. Following the pre-incubation, 200 fmol of ^{32}P labeled probe
 152 oligonucleotides were added and the binding reaction was incubated for an additional 30
 153 minutes before loading onto a pre-ran (3 hours) 6% non-denaturing polyacrylamide gel
 154 containing 0.25 x Tris-Taurine/EDTA. Gels were electrophoresed at 150 V for 3 hours,

vacuum dried and exposed to a phosphoimaging plate (Fuji Photo Film Co.), and processed by the Fuji BAS 1000 Phosphoimaging System (Fuji Photo Film Co.).

157

2.7 *Molecular Weight (MW) determination by Coupled SDS-PAGE and EMSA*

50 µg of Sephacryl S-300 Gel Filtrated fractions or 250 µg of crude NEs were first concentrated using a Micro-con 10 micro-concentrator (Amicon inc., USA) in accordance with the manufacturer's recommendations. The concentrated sample was denatured and resolved with an 8% SDS polyacrylamide slab gel as previously described (Laemmli, 1970). The lane containing resolved NE was uniformly excised (5 mm interval). The nuclear protein was eluted and renatured by modifying a previously described method (Ossipow et al., 1993). Each gel slice was diced and crushed into 1.5 volumes of renaturation buffer (3% Triton X-100, 20 mM Hepes, 100 mM NaCl, 5 mg/ml BSA, 3 mM ZnCl₂, 3 mM MgCl₂, 2 mM DTT, 0.1 mM PMSF, 0.1 mM Benzamidine-HCl) and incubated overnight at 4° C. The remaining gel piece containing MW markers that was resolved concurrently with the nuclear protein was stained and dried. A standard curve, taking in account gel shrinking and swelling which occurs during staining, was constructed in order to determine the MW interval of each gel piece. The nuclear proteins eluted from each gel slices were recovered and assayed for DNA binding activity in EMSA.

173

2.8 *pI determination by coupled isoelectric focusing (IEF) and EMSA*

Sephacryl S-300 fractionated NE (200 µg) was concentrated by using Microspin™ G-25 columns (Amersham Biosciences, USA) and resuspended in rehydration solution (8 M urea, 4% CHAPS, 0.5% IPG buffer, 2 mM tributyl phosphine) followed by IEF analysis as described previously (Woo et al., 2002). Three successive concentration and reconstitution cycles in rehydration solution was followed for buffer exchange and removal of salts for IEF. The samples (250 µl) were then loaded onto IPG Drystrips (13 cm, pH 3-10L or 4-7L; Amersham Biosciences, USA) and loaded onto an IPGphor (Amersham Biosciences, USA)

182 before commencing the following protocol: 12 h rehydration, 100 V for 100 Vh, 250 V for
 183 250 Vh, 500 V for 1000 Vh, 1000 V for 2000 Vh, 8000 V to 60 000 Vh. Following IEF,
 184 strips were rocked in urea free equilibration solution (50 mM tris, 30% glycerol (v/v), 2%
 185 SDS (w/v), 2 mM tributyl phosphine) for 30 minutes. IPG strips were dissected uniformly and
 186 diced into 2 volumes of renaturation buffer. The nuclear proteins eluted from each gel slices
 187 were recovered and assayed for DNA binding activity in EMSA. The pI of each gel piece
 188 was determined from the supplied standard curve (Amersham Bioscience).

190 2.9 Two Dimensional Electrophoresis (2DE)

191 Following IEF, strips were rocked in equilibration solution (6 M urea, 50 mM tris, 30%
 192 glycerol (v/v), 2% SDS (w/v), 2 mM tributyl phosphine) for 30 min. Strips were rinsed
 193 briefly in SDS running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS [w/v]) prior to being
 194 sealed at the top of a 10% SDS polyacrylamide gel with 0.5% agarose in SDS running buffer.
 195 Electrophoresis was performed at 20 mA until the dye front reached the anodic end of gels.
 196 Gels were subsequently stained using the Silver Staining Kit, Protein (Amersham
 197 Biosciences, Sweden) as per manufacturer's instructions. Alternatively, colloidal coomassie
 198 was used if mass spectrometry (MS) was being performed.

200 2.10 MALDI-TOF MS and Data Analysis

201 The excised protein spots of interest were analyzed by Australian Proteome Analysis Facility
 202 (APAF) for MALDI-TOF MS. Samples were subjected to a 16 h tryptic digest at 37° C.
 203 Peptides were extracted from the gel using a 50% (v/v) acetonitrile, 1% (v/v) trifluoroacetic
 204 acid solution. A 1 µL aliquot was spotted onto a sample plate with 1 µL of matrix (α -cyano-
 205 4-hydroxycinnamic acid, 8 mg/mL in 40% acetonitrile (v/v), 1% trifluoroacetic acid) and
 206 MALDI-TOF analysis was performed on a Micromass Tofspec Time of Flight mass
 207 spectrometer. The peptide masses obtained from MALDI-TOF spectra were analyzed using

208 Swiss-Prot databases utilizing the MS-FIT database tool located at the Protein Prospector
209 website (Clauser et al., 1999).

210

211 2.11 Chromatin immunoprecipitation (ChIP)

212 ChIPs were performed based on the Upstate Biotechnology (Lake Placid, NY, USA) protocol
213 with modifications. Briefly, ChIP assays used 10^7 cells per immunoprecipitation, cross-
214 linking with 1% formaldehyde in PBS at room temperature for 15 min and then adding
215 glycine to 0.125 M for 5 min. After washing with ice-cold PBS twice, nuclei were isolated by
216 incubating cells for 15 min in 10 ml NP-40 lysis buffer (10 mM Tris pH 8.0, 10mM NaCl,
217 3mM $MgCl_2$, 0.5% NP-40, 0.15 mM spermine and 0.5 mM spermidine) and dounce
218 homogenizing 10 times on ice. Nuclei were collected, lysed in SDS lysis buffer (5×10^7 cells
219 per ml; 1% SDS, 10 mM EDTA, 50 mM Tris, pH 8.1) and sonicated 5 times on ice for 30 s
220 with output set at 3.5. Soluble chromatin was isolated by centrifugation, diluted as described
221 and incubated overnight with 20 μ g antibodies; α -E12 (Santa Cruz, Cat#sc-762X), α -E47
222 (Santa Cruz, Cat#sc-763X), α -USF1 (Santa Cruz, Cat#sc-8983X), α -USF2 (Santa Cruz,
223 Cat#sc-861X), α -CD19 (IgG control; BD Biosciences PharMingen, Cat#555410) or no
224 antibody. An additional untreated aliquot of chromatin was used to monitor DNA shearing by
225 agarose gel electrophoresis and to generate standard curves for ChIP quantitation. Immune
226 complexes were collected with 60 μ l Salmon Sperm DNA/Protein A/G-agarose beads
227 (Upstate Biotechnology) for 1 h at 4°C and washed according to the manufacturer's
228 specifications. Bound complexes were extracted from antibody twice with 250 μ l elution
229 buffer (1% SDS, 0.1 M $NaHCO_3$) and cross-links were reversed by adding 25 μ l 4 M NaCl
230 and heating to 65°C for 6 h. DNA was extracted by adding 20 μ l 1M Tris-HCl (pH 6.4), 20 μ l
231 0.5 M EDTA (pH 8.1) and 2 μ l 20 mg/ml Proteinase K and incubating for 10 h at 42°C,
232 before using the QIAamp DNA blood mini-kit (QIAGEN) to elute DNA in 100 μ l $0.5 \times$ TE.
233 Q-PCR was performed using 4 μ l template, with the CFX96 Real-Time PCR Detection
234 System (Bio-Rad Laboratories) and analyzed using the CFX Manager Software v1.0 (Bio-Rad

Laboratories). Primer pairs CR2pA, GATA-1-HS3.5 and DOLKp (Table I) were used to detect locus-specific enrichment with an annealing temperature of 55°C.

237

2.12 Chromatin accessibility measured by real-time PCR (CHART-PCR)

Chromatin accessibility assays were performed as previously described (Rao et al., 2001) using the restriction enzyme *DdeI* or *BstNI* with minor modifications. Approximately 1.25×10^6 nuclei in 100 μ l NEBuffer2 (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, 1 mM Dithiothreitol, pH 7.9) $\pm$ 75U *DdeI* (New England Biolabs, Ipswich, MA) were incubated at 37°C for 45 min. For *BstNI* accessibility, nuclei were resuspended in 100 μ l NEBuffer1 (10mM Bis-Tris-Propane-HCl, 10mM, MgCl₂, 1mM DTT) with BSA (0.1 mg/ml) and incubated at 50 °C for 45 min. Reactions were terminated with 100 μ l 40 mM EDTA adding placed on -20°C cooling block. DNA was isolated using the QIAamp DNA blood mini-kit (QIAGEN), eluted in 150 μ l 0.1 $\times$ TE, supplemented with 0.008% BSA (w/v). 5 μ l of DNA recovered from “Cut” (+*DdeI*) and “Uncut” samples were used in Q-PCR assays to measure the relative abundance of amplicons within the *CR2/CD21* promoter containing *DdeI* sites (CTNAG sites located at 1143/-1139, -284/-280 and -55/-51 using primers shown in Table I: CR2pF, CR2pD and CR2pB) or a *BstNI* site (CCWGG site located at -77/-73 assayed using CR2pB). Chromatin accessibility ratios were calculated using the Rotor-Gene™ v6.1, Comparative Quantitation method to determine the ratio of “Uncut”/“Cut” for 3 independent pairs of samples.

255

2.12 Western blot

Cells were lysed in 1x RIPA lysis buffer (sc-24948) according to the manufacturer’s instructions (Santa Cruz Biotechnology, Santa Cruz, CA, USA). Protein (20 ug) was mixed 1:1 with 2x SDS sample buffer, boiled for three minutes and separated on 10% NuPAGE Bis-Tris gels (Invitrogen) under reducing conditions then electroblotted onto nitrocellulose membranes. Membranes were blocked with 5% (wt/vol) skim milk powder (SMP) in TBST

(10 mM Tris, pH 7.4, 150 mM NaCl, 0.05% Tween 20) and probed with primary antibody diluted 1:200 (RP58; sc-248479 and E2A; sc-416) or 1:50,000 (β actin; A1978) in TBST containing 5% SMP. After washing five times with TBST, membranes were incubated with horseradish peroxidase–conjugated donkey anti-goat IgG (RP58; sc2033) or goat anti-mouse IgG (E2A; A0168) diluted 1:10,000, or goat anti-mouse IgG (β actin; A0168) diluted 1:5,000 in TBST containing 5% SMP. The membranes were washed five times with TBST and developed using Pierce ECL Plus Western Blotting Substrate (Thermo Scientific, IL, USA) and exposure to Hyperfilm (Amersham Biosciences).

2.14 Statistical analysis

All values described represent the mean $\pm$ S.D. and *P*-values were calculated by two-tailed unpaired t-tests unless stated otherwise. Statistics and graphs were generated using GraphPad Prism v4 (GraphPad, San Diego CA, USA).

3. Results

3.1 Deletion and mutation analysis of tandem E-box sites

We have previously demonstrated that E-box motif 2 is crucial for silencing the *CR2/CD21* promoter in non-expressing cells (Ulgiati and Holers, 2001). On the other hand, E-box motif 1 binds the transcriptional activator USF1 *in vitro* and is required for basal transcription in CR2-expressing cells (Ulgiati et al., 2002). Since the activity of the tandem E-boxes show opposing functions, we wished to examine potential interactions between E-boxes using both CR2-expressing and –non-expressing cell-types. To examine the effect of tandem E-boxes, mutant -315/+75 constructs were generated which contained a deletion in either of the E-boxes or introduced mutations into either of the E-box motifs (Figure 1D). In agreement with our previous studies, deletion or mutation of E-box site 1 resulted in a dramatic reduction in transcriptional activity in Raji cells (Figure 1E; deletion: 4.1-fold decrease *P* < 0.0001 and

289 mutation: 1.9-fold decrease $P < 0.0001$). The reduction in transcriptional activity of these
 290 constructs was significant in K562 cells, however muted likely reflecting the low basal
 291 activity in the non-expressing cell-type (Figure 1F; deletion: 2.2-fold decrease $P < 0.0001$ and
 292 mutation: 1.3-fold decrease $P = 0.0162$). In contrast, mutation or deletion of E-box site 2
 293 resulted in a dramatic increase in transcriptional activity in K562 cells (6-7 fold increase; $P <$
 294 0.0001), confirming our previous finding that this site likely binds repressor proteins. The E-
 295 box 2 mutant and deletion constructs also showed elevated transcriptional activity in Raji
 296 cells, however, the effect was at a lower magnitude (1.5-2 fold; $P < 0.0001$). These results
 297 highlight that the tandem E-box sites within the *CR2/CD21* promoter play opposing functions
 298 and that the effect of these sites is dependent on the cellular context (Figure 1). We next
 299 generated two mutant -315/+75 constructs that contained a deletion of one E-box and
 300 introduced an *Mlu* I site into the other E-box (Figure 1A). Transfection of these constructs in
 301 Raji (CR2-positive) cells revealed that intact tandem E-boxes were necessary for basal
 302 transcriptional activity in Raji cells since both deletion ($P < 0.0001$) and mutant ($P < 0.0001$)
 303 constructs showed approximately 50% reduction in promoter activity (Figure 1B). Similarly,
 304 repression of these constructs in K562 (CR2-negative) also required intact E-box motifs since
 305 both mutant constructs showed similarly (2.5-3 fold; both $P < 0.0001$) elevated transcriptional
 306 activity in K562 cells (Figure 1C). Thus, opposing activities of tandem E-box elements is
 307 specific to cell-type and additive within the same cellular context.

308

309 3.2 Linker-scanning mutagenesis reveals critical nucleotides for E-box site 1 310 activator function

311 Since the E-box motifs show opposing functions and these are cell-context dependent, we
 312 wished to decipher which specific nucleotides were important for mediating these effects. We
 313 have previously determined E-box site 2 is critically dependent on the region -67 to -59
 314 (GCAGGTGTG) to function as a repressor in CR2-non-expressing cells by linker-scanning
 315 mutagenesis. A similar approach was used to determine the specific nucleotides mediating the

316 activator function of E-box site 1. A series of mutant constructs were generated by
 317 introducing variant nucleotides at positions -52/-50, -47/-45, -42/-40 and -37/-35 (Figure 2A).
 318 All mutant constructs together with the wild-type -315/+75 were transfected into Raji and
 319 K562 cells (Figure 2B). Again, we noted that each of these constructs showed very low level
 320 transcriptional activity in K562 cells with significantly lower levels in all mutants. The most
 321 dramatic reduction in transcriptional activity (50-70% reduced activity) occurred in Raji cells
 322 using either the -47/-45 (2.4 fold reduction; $P = 0.0229$) or -42/-40 (3.5 fold reduction; $P =$
 323 0.0058) mutant constructs. We conclude, therefore, that nucleotides -47 to -40 of the
 324 *CR2/CD21* promoter (AGCACGTG) are crucial for the activator function of E-box site 1
 325 (Figure 2).

326

327 3.3 *E-box spacing does not effect cell-specific promoter activity*

328 We have noted that the tandem E-boxes within the *CR2/CD21* promoter are separated by 22
 329 bp which corresponds to two turns of the DNA double helix. To determine the effect of the
 330 spacing of the E-boxes, we generated reporter constructs containing either a 6 bp insertion or
 331 a 6 bp deletion at the mid-point between the two E-boxes (Figure 3A, B). We reasoned that if
 332 the E-boxes required specific phasing for cooperative binding of protein complexes, altering
 333 this by adding or removing 6 bp, would interrupt their respective functions since 6 bp
 334 corresponds to a half turn of the DNA helix. We note that these constructs do not show cell-
 335 specific activity as described for deletion/mutation reporters. In contrast, we found that the
 336 deletion construct, generating a spacing of 1.5 turns between E-boxes, resulted in a modest
 337 decrease in transcriptional activity in both Raji and K562 cells (Figure 3C; 1.4 fold reduction;
 338 $P = 0.0067$). In contrast, the construct containing an insertion resulting in a spacing of 2.5
 339 turns between E-box motifs, showed no effect when compared to wild-type constructs in
 340 either cell line ($P = 0.5752$). Our data, therefore, does not indicate that precise spacing and/or
 341 phasing between the two E-boxes contributes to cell-specificity (Figure 3). However, these
 342 results suggest that additive effects of E-box sites mediating promoter activation are restricted

when a small deletion juxtaposes these motifs closely together, while repressor functions are unaffected.

3.4 Identification of interaction between RP58 and E-box site 2

We have previously examined the repressor E-box site 2 using EMSA and have shown that a DNA oligonucleotide probe spanning -73 to -52 (Table 1; E-box site 2) of the *CR2/CD21* promoter specifically binds to proteins prepared from K562 NEs (Ulgiati and Holers, 2001). We observed specific binding of a high molecular weight complex in addition to three specific lower order bands (A-C) representing distinct protein complexes. Our supershift and cross-competition experiments demonstrated that E2A proteins were present in complex A (Ulgiati and Holers, 2001), while additional unidentified proteins were important in mediating the repressor effect *via* complex C. We therefore wished to investigate the identity of proteins which form this lower order complex.

To this end, we fractionated K562 NEs using Sephacryl S-300 HR columns. EMSA was performed using the oligonucleotide probe spanning -73 to -52 of the *CR2/CD21* promoter (CR2/CD21 E-box site 2; Table I), incubated either with crude NE or protein fractions of decreasing sizes. As expected three major complexes were detected using crude NE (Figure 4A, NE; A-C), while the fractionated extracts showed varying intensities of these bands in concordance with the sizes of proteins separated by chromatography (Figure 4A, lanes 1-12). Therefore, following fractionation of K562 NEs, the separated nuclear proteins re-capitulated the banding patterns observed in crude extract. Importantly, low molecular weight fractions were recovered which formed complex C but not other higher weight complexes (Figure 4A lanes 9-12).

As a step to identify protein(s) that form complex C, fractionated extracts showing this activity were separated by SDS polyacrylamide gel electrophoresis (PAGE) and bands corresponding to molecular weights 43-61 kDa were excised, renatured and used in binding reactions for EMSA analysis. We determined that the activity of proteins forming complex C

was within the size range 53-57 kDa by PAGE (Figure S1A). Next, we separated fractionated extracts by 2-dimensional PAGE to further purify protein(s) that retain the activity to form complex C. By excising gel slices at the appropriate MW but with differing iso-electric points, we determined that the activity of protein(s) generating complex C could be recovered within the pI range 4.9-5.3 (Figure S1B). We next subjected the protein fractions that form complex C only, to 2-dimensional PAGE to separate acidic proteins (Figure 4B). We selected a rectangular region of the gel based on the predicted size (53-57 kDa) and pI (4.9-5.2) of protein(s) that form complex C and sliced this into 20 segments (Figure 4B; expanded panel). Protein was then recovered from each of the 20 segments and subsequently used for EMSA using the E-box site 2 probe. As expected the crude NE formed complexes A-C, while the S300 fraction formed complex C alone. Of note, segments of the gel labeled 5 and 6 (Figure 4C), formed complex C alone indicating that these portions contain proteins responsible for this activity. Protein spots 5 and 6 were analyzed by MALDI-TOF MS analysis. Monoisotopic peaks were searched against the SwissProt and NCBI databases for human proteins 1-100kDa in size with a maximum of one missed cleavage, unmodified cysteines and with a mass tolerance of 11ppm. The database was searched using the MS-FIT database tool located at the ProteinProspector website. The top match for protein spot 5 resulted in the identification of RP58 (MOWSE score of 3012; %Peptide matched of 24%). Therefore, these results suggest that RP58 is a sequence-specific repressor protein capable of binding to E-box site 2 within the *CR2/CD21* proximal promoter.

3.5 *RP58 represses CR2/CD21 promoter activity in vitro via E-box site 2*

Since RP58 was identified as a repressor protein capable of binding E-box site 2, we wished to investigate its function with respect to the *CR2/CD21* proximal promoter. We generated luciferase reporter constructs containing the -315 to +75 *CR2/CD21* promoter with mutations at either E-box site or both sites (double mutant) and co-transfected these with an RP58 over-expression construct. As a control, we co-transfected the RP58 over-expression construct with

an SV40-luciferase reporter (Figure 5). We found similar levels of repression (~12-fold decrease) using the wild-type 315-*CR2/CD21*-luciferase reporter and the corresponding E-box site 1 mutant ($P = 0.647$). In contrast, RP58-mediated repression of the 315-*CR2/CD21*-luciferase reporter containing the E box site 2 mutant was abolished (~5-fold decrease; $P = 0.001$), demonstrated by similar repression compared to the control SV40-luciferase reporter (~5-fold decrease; $P = 0.0009$). We observed similarly muted repression using the 315-*CR2/CD21*-luciferase reporter containing mutations in both E-box sites (~5-fold decrease; $P = 0.0101$). These results clearly define E-box site 2 within the *CR2/CD21* promoter as mediating repression by RP58.

3.6 *E12 and E47 activate the CR2/CD21 proximal promoter via tandem E-boxes*

We have previously found that *in vitro*, E2A proteins (E12 and E47) interact with E-box site 2 but not E-box site 1 (Ulgiati and Holers, 2001, Ulgiati et al., 2002). To investigate the effect of E2A expression on *CR2/CD21* promoter activation we performed co-transfection experiments using E2A (E12/E47) and E47 over-expression vectors together with luciferase reporter constructs containing wild-type 315luc or constructs containing mutations in either E-box. Transfections were first performed in CR2-non-expressing K562 and CR2-expressing Raji cells to compare the effect of E2A and E47 with that of RP58 overexpression. In both cell-types RP58 mediated repression (K562: 2.3-fold repression, $P < 0.0001$; Raji: 8.0-fold repression, $P < 0.0001$) while E47 or E2A transactivated the wild-type -315luc constructs (K562: 7.5-fold and 5.4-fold activation, $P < 0.0001$; Raji: 2.4-fold and 2.3-fold activation, $P < 0.0001$). These results suggest that E2A proteins function as activators of the *CR2/CD21* promoter.

We next tested the activating effect of E2A and E47 over-expression using E-box mutant constructs in K562 cells and found that activation was completely muted in constructs harboring mutations in E-box site 2 (Figure 7B, C). This data is in agreement with our previous experiments that demonstrated that E2A proteins bind E-box 2 *in vitro*. On the other

hand, reporter constructs harboring mutations in E-box site 1 were still activated, however, at approximately half the level of wild-type constructs by both E2A and E47 overexpression (Figure 7B, C; E47: $P = 0.1371$; E2A: $P = 0.1267$). Taken together, these results suggest that E2A factors compete for binding at E-box site 2 with repressor proteins (including RP58) while on the other hand, E-box 1 binds activators which function cooperatively with E2A.

3.7 Binding of E12, E47, USF1 and USF2 at the CR2/CD21 promoter correlates with chromatin accessibility in Raji cells

Since over-expression of E2A activates *CR2/CD21* proximal promoter activity *in vitro* we next wanted to determine the physiological relevance of this interaction. We performed ChIP assays to examine *in vivo* occupancy of the *CR2/CD21* promoter by E12 and E47 in CR2-expressing Raji cells and CR2-non-expressing K562 cells. Both E12 and E47 are expressed at comparable levels in these cell lines (data not shown). Our ChIP analysis confirmed that both E12 and E47 bind the *CR2/CD21* promoter *in vivo* in Raji cells but not K562 cells (Figure 7A, B). As a control, we confirmed that both E12 and E47 bind an enhancer at the *GATA-1* locus in K562 cells (primer pair GATA-1 HS3.5; previously reported by Valverde-Garduno *et al.* (Valverde-Garduno *et al.*, 2004)) but they do not bind here in Raji cells (Figure 7C, D).

We have previously shown by *in vitro* EMSA experiments that the *CR2/CD21* site 1 E-box motif binds USF-family TFs (Ulgiati *et al.*, 2002), while the site 2 E-box binds E2A proteins (Ulgiati and Holers, 2001). We performed ChIP assays targeting USF1 and USF2 in both CR2-expressing Raji cells and CR2-non-expressing K562 cells. Our results show specific enrichment of *CR2/CD21* promoter sequences from immunoprecipitated chromatin using either α -USF1 or α -USF2 in Raji cells but not K562 cells (Figure 7A, B). This interaction was cell-specific because both α -USF1/2 antibodies enriched promoter fragments (primer pair DOLKp) defined as constitutive USF-binding sites at the *DOLK* promoter (Gallagher *et al.*, 2009, Rada-Iglesias *et al.*, 2008) in both K562 and Raji cells (Figure 7E, F).

451

452 We next performed chromatin accessibility assays using restriction enzymes that target
 453 regions in the vicinity of the tandem E-box elements to enable fine-mapping of
 454 hypersensitivity. We observed significantly enhanced chromatin accessibility (8-fold
 455 increase; $P = 0.0297$) in CR2-expressing Raji cells at a *Dde* I sites spanning -55 to -51 (primer
 456 pair CR2pB) when compared to K562 cells (Figure 7G). Enhanced chromatin accessibility of
 457 these sites in Raji cells was sequence-specific, since low levels of chromatin accessibility
 458 were observed across the -284 to -280 and -1143 to -1139 *Dde*I sites in both Raji and K562
 459 cell types (Figure 7G). We also identified enhanced accessibility at a *Bst*NI-site located at -
 460 77/-73 of the *CR2/CD21* promoter in Raji cells (3.5-fold increase; $P = 0.0115$) compared to
 461 an intronic site at +634/+638 (Figure 7G).

462 We conclude that E2A and USF proteins bind the *CR2/CD21* proximal promoter in a cell-
 463 specific manner correlating with enhanced surrounding tandem E-box sites in CR2/CD21-
 464 expressing Raji cells but not in non-expressing K562 cells.

465

466 3.8 *RP58 and E2A are expressed in human B cells*

467 The pivotal role of E2A in lymphoid development is well established while the role of RP58
 468 in B cell development has not been defined. As a first step to determine a potential role for
 469 RP58 in lymphoid cells, we assessed protein expression in cell lines and peripheral human B
 470 cells. Protein from Raji and K562 cell lines as well as primary human B cells was probed for
 471 E2A and RP58. Immunodetection of E2A revealed a band of 70 kDa, abundant in the cell
 472 lines Raji and K562, and primary human B cells (Figure 8A, upper panel). Immunodetection
 473 of RP58 generated a band of approximately 60 kDa in both Raji and K562 cells and was
 474 detectable at low levels in primary B cells (Figure 8A, center panel). Finally, we confirm
 475 expression of the loading control β actin, generating the expected 42 kDa band in all samples
 476 (Figure 8A, lower panel).

477 We next searched the FANTOM5 database (Forrest et al., 2014, Andersson et al., 2014) to
 478 confirm *TCF3*, *CR2/CD21* and *RP58* promoter activity in CD19 B cells. We also wished to
 479 explore their activity in context with a panel of genes activated during early B-lymphoid
 480 development (Kohn et al., 2012). We downloaded tag counts for promoters of interest using
 481 FANTOM5 Table Extraction Tool (<http://fantom.gsc.riken.jp/5/tet>), for 65 human blood
 482 lineage samples (excluding treated and expanded specimens), 15 muscle samples and 11 brain
 483 samples. Promoter activity was defined at the gene-level by summing normalized tag counts
 484 arising from alternative transcriptional start sites (Table S1). Using these normalized count
 485 data, we found a 2.7-fold increase in *TCF3* promoter activity ($P = 5.63 \times 10^{-11}$), a 44-fold
 486 increase in *CR2/CD21* promoter activity ($P = 3.67 \times 10^{-24}$) and a 4.4-fold increase in *RP58*
 487 promoter activity ($P = 2.83 \times 10^{-10}$) in CD19 B cells ($n=11$) compared to non B blood lineage
 488 cells ($n = 54$; including stem cells ($n = 9$), monocytes ($n = 12$), CD4 T-cells ($n = 22$) and CD8
 489 T-cells ($n = 11$). We also confirm B lineage specific activity of the *CD19* promoter and the
 490 *PAX5* promoter (Figure 8B). We next compared *RP58* promoter activity in blood lineage cells
 491 with tissues where *RP58* has a known role, brain and muscle. We found highest expression of
 492 CAGE-tags in brain tissue ($n=3$ neuron; $n=4$ cerebellum; $n=4$ hippocampus) with lower levels
 493 in muscle tissues ($n=9$ skeletal muscle; $n=3$ skeletal muscle satellite cells; and $n=3$ myoblast
 494 cells). *RP58* activity was 38-fold lower in CD19 B cells compared to brain samples ($P <$
 495 0.0001), and 2.6-fold greater than muscle samples ($P = 0.0157$). Next, we interrogated these
 496 samples using a gene-set defined by Kohn and colleagues (2012) as up-regulated in B-
 497 lymphoid progenitors ($n = 26$ genes). We found B cell specific promoter activity for 10 of
 498 these genes (*CD79A*, *CD22*, *CD19*, *TCF3*, *VPREB3*, *CD24*, *EBF1*, *CD79B*, *PAX5* and
 499 *MEF2A*), together with *RP58* and *CR2/CD21* at a false discovery rate of 1% (Holm-Šidák
 500 multiple correction; Figure 8D). These results reveal similar patterns of *RP58*, *TCF3* and
 501 *CR2/CD21* promoter activity among human blood lineage cells and confirm specific up-
 502 regulation of *RP58* in CD19 B cells. As a corollary to these observations, we also note co-
 503 expression of mouse orthologs of *RP58* (*Zfp238*) and *CR2/CD21* (*Cr2*) within a set of genes

defined as “the major early B module, relevant to development” by the Immunological Genome Project (Figure S2; <http://www.immgen.org/>) (Painter et al., 2011).

4. Discussion

Since *CR2/CD21* expression is tightly controlled during B cell ontogeny, detected only on immature B cells that escape negative selection and mature B cells, factors regulating *CR2/CD21* expression are likely to play wider roles in B cell development. Our laboratory has previously discovered several critical elements involved in cell type-specific repression (Makar et al., 2001, Ulgiati and Holers, 2001) and basal activation of human *CR2/CD21* (Ulgiati et al., 2002) as well as mapping cell-type specific chromatin structural variation across the *CR2/CD21* proximal promoter (Cruickshank et al., 2008, Cruickshank et al., 2009). We have now used a proteomic approach, which identified RP58 as a novel factor that binds the repressor E-box site 2. Our data also shows that E2A proteins interact at this site and can activate *CR2/CD21* promoter activity. Furthermore, we identify a role for E-box site 1 in mediating USF-activation of *CR2/CD21* promoter activity, however we find that this activity was independent of E-box site 2. Both E2A and USF binding at the *CR2/CD21* promoter was cell-specific and correlated with enhanced chromatin accessibility within the vicinity of the tandem E boxes in CR2-expressing cells compared to non-expressing cells. Finally, analysis of 65 blood lineage samples from the FANTOM5 database revealed enhanced activity of the *RP58* promoter in human CD19 B cells.

RP58 is a sequence-specific zinc-finger transcriptional repressor that associates with heterochromatin (Aoki et al., 1998). RP58 preferentially binds E-box motifs and also interacts with DNMT2a to repress transcription (Meng et al., 2000, Fuks et al., 2001). The functions of RP58 have been predominantly examined in the nervous system (Kang et al., 2001, Ohtaka-Maruyama et al., 2007, Okado et al., 2009) and more recently during myogenesis (Di Padova

et al., 2007, Yokoyama et al., 2009), however, there is limited information regarding its specific role in other tissues. In both humans and mice, RP58 is highly expressed in the brain with distinct regional expression levels. On the other hand lower level expression of RP58 has been detected in most tissues, including skeletal muscle in which it plays a profound developmental role orchestrating late events during the myogenesis. Furthermore, RP58 has been defined within a signature gene-expression profile distinctive of t(1;19)(q23;p13.3) translocated pediatric B cell acute lymphoblastic leukemia (Ross et al., 2003). This translocation results in expression of a chimeric protein, E2A-PBX1, comprising the DNA-binding domain of the PBX1 protein fused to the E2A transactivation domain (Aspland et al., 2001). These observations raise the possibility that deregulated RP58 expression may perturb normal B cell development. Our proteomic approach to identify factors binding the E-box near the *CR2/CD21* transcription start site, provides the first evidence that RP58 may play a role as a repressive TF in lymphoid cells. In addition, enhanced *RP58* promoter activity in human CD19 B cells further supports an unexpected role for this TF in B lymphoid cells. Finally, we noted co-expression of *CR2/CD21* and *RP58* mouse orthologs (*Cr2* and *Zfp238*) within a module of genes up-regulated in early B cells (Jojic et al., 2013), suggesting that specific activation of these genes during B ontogeny is conserved. Taken together, these data implicate RP58 as a regulator of B lymphoid cells in humans and mice.

Previous reports indicate that RP58 competes with E proteins for binding at E-box motifs (Aoki et al., 1998). Our data support this hypothesis since RP58 and E2A proteins bind the same E-box element within the *CR2/CD21* proximal promoter, however, function in repression and activation respectively. Since RP58 functions as a repressor, our chromatin accessibility data is also consistent with a model where RP58 binds to E box motifs to bring about a closed chromatin environment. We speculate that, RP58 prevents E2A binding at lymphoid-restricted E-box motifs by regulating chromatin conformation. Further *in vivo* genomic approaches are necessary to determine the scope of this reciprocal relationship. Additionally, mouse models with lineage- and stage-specific knock-out of RP58 in hematopoietic cells are required to determine the precise role of this TF in B cell

559 development. Since E2A, USF and RP58 are widely expressed but also play specific roles in
560 cellular differentiation (Yokoyama et al., 2009, Rada-Iglesias et al., 2008, Massari and Murre,
561 2000, Kee, 2009, Henthorn et al., 1990, Bhattacharya and Baker, 2011, Anantharaman et al.,
562 2011), genomic strategies may also determine how these factors control E-box activity in the
563 context of cell-specific chromatin configurations and also DNA sequence organization.

564

565 In summary, using a novel proteomic approach, we identified RP58 as a direct repressor of
566 the *CR2/CD21* promoter and have functionally characterized binding of bHLH factors to
567 tandem E-box sites that also interact with RP58. This is the first evidence that RP58, a known
568 regulator of neurogenesis and myogenesis, may also play a functional role in the
569 hematopoietic system. Our results suggest that E2A, USF and RP58 proteins are co-regulators
570 of E-box motifs within the human *CR2/CD21* promoter that may function to program
571 localized chromatin structure and gene activation of this key marker of B cell development.

572

572

573 Acknowledgements

574 This work was supported by the National Health and Medical Research Council of Australia
575 (to D.U. and L.J.A.), the Children's Leukaemia and Cancer Research Foundation, Western
576 Australia (to M.N.C) and the Cancer Council Western Australia (to M.N.C). We thank
577 Professor Jim Hagman and Professor Tony Kouzarides for providing E2A and RP58
578 expression constructs respectively.

579

580 References

581

- 582 Alvarez, J. D., Yasui, D. H., Niida, H., Joh, T., Loh, D. Y. & Kohwi-Shigematsu, T.
583 (2000). The MAR-binding protein SATB1 orchestrates temporal and spatial
584 expression of multiple genes during T-cell development. *Genes Dev*, *14*, 521-
585 535.
- 586 Anantharaman, A., Lin, I. J., Barrow, J., Liang, S. Y., Masannat, J., Strouboulis, J.,
587 Huang, S. & Bungert, J. (2011). Role of helix-loop-helix proteins during
588 differentiation of erythroid cells. *Mol Cell Biol*, *31*, 1332-1343.
- 589 Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M.,
590 Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li,
591 K., Schwarzfischer, L., Glatz, D., Raithel, J., Lilje, B., Rapin, N., Bagger, F.
592 O., Jorgensen, M., Andersen, P. R., Bertin, N., Rackham, O., Burroughs, A.
593 M., Baillie, J. K., Ishizu, Y., Shimizu, Y., Furuhashi, E., Maeda, S., Negishi,
594 Y., Mungall, C. J., Meehan, T. F., Lassmann, T., Itoh, M., Kawaji, H., Kondo,
595 N., Kawai, J., Lennartsson, A., Daub, C. O., Heutink, P., Hume, D. A., Jensen,
596 T. H., Suzuki, H., Hayashizaki, Y., Muller, F., Forrest, A. R., Carninci, P.,
597 Rehli, M. & Sandelin, A. (2014). An atlas of active enhancers across human
598 cell types and tissues. *Nature*, *507*, 455-461.
- 599 Aoki, K., Meng, G., Suzuki, K., Takashi, T., Kameoka, Y., Nakahara, K., Ishida, R. &
600 Kasai, M. (1998). RP58 associates with condensed chromatin and mediates a
601 sequence-specific transcriptional repression. *J Biol Chem*, *273*, 26698-26704.
- 602 Aspland, S. E., Bendall, H. H. & Murre, C. (2001). The role of E2A-PBX1 in
603 leukemogenesis. *Oncogene*, *20*, 5708-5717.
- 604 Bacon, K., Gauchat, J. F., Aubry, J. P., Pochon, S., Graber, P., Henchoz, S. &
605 Bonnefoy, J. Y. (1993). CD21 expressed on basophilic cells is involved in
606 histamine release triggered by CD23 and anti-CD21 antibodies. *Eur J*
607 *Immunol*, *23*, 2721-2724.
- 608 Bhattacharya, A. & Baker, N. E. (2011). A network of broadly expressed HLH genes
609 regulates tissue-specific cell fates. *Cell*, *147*, 881-892.
- 610 Clauser, K. R., Baker, P. & Burlingame, A. L. (1999). Role of accurate mass
611 measurement (+/- 10 ppm) in protein identification strategies employing MS
612 or MS/MS and database searching. *Analytical chemistry*, *71*, 2871-2882.

- 613 Cruickshank, M., Fenwick, E., Abraham, L. J. & Ulgiati, D. (2008). Quantitative
614 differences in chromatin accessibility across regulatory regions can be directly
615 compared in distinct cell-types. *Biochem Biophys Res Commun*, 367, 349-355.
- 616 Cruickshank, M. N., Fenwick, E., Karimi, M., Abraham, L. J. & Ulgiati, D. (2009).
617 Cell- and stage-specific chromatin structure across the Complement receptor 2
618 (CR2/CD21) promoter coincide with CBF1 and C/EBP-beta binding in B
619 cells. *Mol Immunol*, 46, 2613-2622.
- 620 Di Padova, M., Caretti, G., Zhao, P., Hoffman, E. P. & Sartorelli, V. (2007). MyoD
621 acetylation influences temporal patterns of skeletal muscle gene expression. *J*
622 *Biol Chem*, 282, 37650-37659.
- 623 Forrest, A. R., Kawaji, H., Rehli, M., Baillie, J. K., de Hoon, M. J., Haberle, V.,
624 Lassman, T., Kulakovskiy, I. V., Lizio, M., Itoh, M., Andersson, R., Mungall,
625 C. J., Meehan, T. F., Schmeier, S., Bertin, N., Jorgensen, M., Dimont, E.,
626 Arner, E., Schmidl, C., Schaefer, U., Medvedeva, Y. A., Plessy, C., Vitezic,
627 M., Severin, J., Semple, C., Ishizu, Y., Young, R. S., Francescato, M., Alam,
628 I., Albanese, D., Altschuler, G. M., Arakawa, T., Archer, J. A., Arner, P.,
629 Babina, M., Rennie, S., Balwierz, P. J., Beckhouse, A. G., Pradhan-Bhatt, S.,
630 Blake, J. A., Blumenthal, A., Bodega, B., Bonetti, A., Briggs, J., Brombacher,
631 F., Burroughs, A. M., Califano, A., Cannistraci, C. V., Carbajo, D., Chen, Y.,
632 Chierici, M., Ciani, Y., Clevers, H. C., Dalla, E., Davis, C. A., Detmar, M.,
633 Diehl, A. D., Dohi, T., Drablos, F., Edge, A. S., Edinger, M., Ekwall, K.,
634 Endoh, M., Enomoto, H., Fagiolini, M., Fairbairn, L., Fang, H., Farach-
635 Carson, M. C., Faulkner, G. J., Favorov, A. V., Fisher, M. E., Frith, M. C.,
636 Fujita, R., Fukuda, S., Furlanello, C., Furino, M., Furusawa, J., Geijtenbeek, T.
637 B., Gibson, A. P., Gingeras, T., Goldowitz, D., Gough, J., Guhl, S., Guler, R.,
638 Gustincich, S., Ha, T. J., Hamaguchi, M., Hara, M., Harbers, M., Harshbarger,
639 J., Hasegawa, A., Hasegawa, Y., Hashimoto, T., Herlyn, M., Hitchens, K. J.,
640 Ho Sui, S. J., Hofmann, O. M., Hoof, I., Hori, F., Huminiecki, L., et al. (2014).
641 A promoter-level mammalian expression atlas. *Nature*, 507, 462-470.
- 642 Fuks, F., Burgers, W. A., Godin, N., Kasai, M. & Kouzarides, T. (2001). Dnmt3a
643 binds deacetylases and is recruited by a sequence-specific repressor to silence
644 transcription. *EMBO J*, 20, 2536-2544.
- 645 Gallagher, P. G., Nilson, D. G., Steiner, L. A., Maksimova, Y. D., Lin, J. Y. &
646 Bodine, D. M. (2009). An insulator with barrier-element activity promotes
647 alpha-spectrin gene expression in erythroid cells. *Blood*, 113, 1547-1554.
- 648 Gommerman, J. L., Oh, D. Y., Zhou, X., Tedder, T. F., Maurer, M., Galli, S. J. &
649 Carroll, M. C. (2000). A role for CD21/CD35 and CD19 in responses to acute
650 septic peritonitis: a potential mechanism for mast cell activation. *J Immunol*,
651 165, 6915-6921.
- 652 Guo, C., Hu, Q., Yan, C. & Zhang, J. (2009). Multivalent binding of the ETO
653 corepressor to E proteins facilitates dual repression controls targeting
654 chromatin and the basal transcription machinery. *Mol Cell Biol*, 29, 2644-
655 2657.
- 656 Hartley, S. B., Cooke, M. P., Fulcher, D. A., Harris, A. W., Cory, S., Basten, A. &
657 Goodnow, C. C. (1993). Elimination of self-reactive B lymphocytes proceeds
658 in two stages: arrested development and cell death. *Cell*, 72, 325-335.
- 659 Henthorn, P., Kiledjian, M. & Kadesch, T. (1990). Two distinct transcription factors
660 that bind the immunoglobulin enhancer microE5/kappa 2 motif. *Science*, 247,
661 467-470.

- 662 Hunyadi, J., Simon, M., Jr., Kenderessy, A. S. & Dobozy, A. (1991). Expression of
663 complement receptor CR2 (CD21) on human subcorneal keratinocytes in
664 normal and diseased skin. *Dermatologica*, 183, 184-186.
- 665 Jojic, V., Shay, T., Sylvia, K., Zuk, O., Sun, X., Kang, J., Regev, A., Koller, D., Best,
666 A. J., Knell, J., Goldrath, A., Joic, V., Cohen, N., Brennan, P., Brenner, M.,
667 Kim, F., Rao, T. N., Wagers, A., Heng, T., Ericson, J., Rothamel, K., Ortiz-
668 Lopez, A., Mathis, D., Benoist, C., Bezman, N. A., Sun, J. C., Min-Oo, G.,
669 Kim, C. C., Lanier, L. L., Miller, J., Brown, B., Merad, M., Gautier, E. L.,
670 Jakubzick, C., Randolph, G. J., Monach, P., Blair, D. A., Dustin, M. L.,
671 Shinton, S. A., Hardy, R. R., Laidlaw, D., Collins, J., Gazit, R., Rossi, D. J.,
672 Malhotra, N., Kreslavsky, T., Fletcher, A., Elpek, K., Bellemarte-Pelletier, A.,
673 Malhotra, D. & Turley, S. (2013). Identification of transcriptional regulators in
674 the mouse immune system. *Nat Immunol*, 14, 633-643.
- 675 Kang, J. S., Yang, Y. C., Liu, H. L., Li, Y. H., Du, Y. C. & Li, R. X. (2001). Cloning
676 and Distribution of rRP58, A Novel Neuronal Gene. *Sheng Wu Hua Xue Yu*
677 *Sheng Wu Li Xue Bao (Shanghai)*, 33, 563-568.
- 678 Kee, B. L. (2009). E and ID proteins branch out. *Nat Rev Immunol*, 9, 175-184.
- 679 Kohn, L. A., Hao, Q. L., Sasidharan, R., Parekh, C., Ge, S., Zhu, Y., Mikkola, H. K.
680 & Crooks, G. M. (2012). Lymphoid priming in human bone marrow begins
681 before expression of CD10 with upregulation of L-selectin. *Nat Immunol*, 13,
682 963-971.
- 683 Kwon, K., Hutter, C., Sun, Q., Bilic, I., Cobaleda, C., Malin, S. & Busslinger, M.
684 (2008). Instructive role of the transcription factor E2A in early B
685 lymphopoiesis and germinal center B cell development. *Immunity*, 28, 751-
686 762.
- 687 Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the
688 head of bacteriophage T4. *Nature*, 227, 680-685.
- 689 Levine, M. & Davidson, E. H. (2005). Gene regulatory networks for development.
690 *Proc Natl Acad Sci U S A*, 102, 4936-4942.
- 691 Levy, E., Ambrus, J., Kahl, L., Molina, H., Tung, K. & Holers, V. M. (1992). T
692 lymphocyte expression of complement receptor 2 (CR2/CD21): a role in
693 adhesive cell-cell interactions and dysregulation in a patient with systemic
694 lupus erythematosus (SLE). *Clin Exp Immunol*, 90, 235-244.
- 695 Li, Y. C., Ross, J., Scheppler, J. A. & Fianza, B. R., Jr. (1991). An in vitro
696 transcription analysis of early responses of the human immunodeficiency virus
697 type 1 long terminal repeat to different transcriptional activators. *Molecular*
698 *and cellular biology*, 11, 1883-1893.
- 699 Makar, K. W., Ulgiati, D., Hagman, J. & Holers, V. M. (2001). A site in the
700 complement receptor 2 (CR2/CD21) silencer is necessary for lineage specific
701 transcriptional regulation. *Int Immunol*, 13, 657-664.
- 702 Marshak, D. R. (1996). *Strategies for protein purification and characterization : a*
703 *laboratory course manual*, (Editor ed.). Plainview, N.Y.: Cold Spring Harbor
704 Laboratory Press.
- 705 Massari, M. E. & Murre, C. (2000). Helix-loop-helix proteins: regulators of
706 transcription in eucaryotic organisms. *Mol Cell Biol*, 20, 429-440.
- 707 Melamed, D., Benschop, R. J., Cambier, J. C. & Nemazee, D. (1998). Developmental
708 regulation of B lymphocyte immune tolerance compartmentalizes clonal
709 selection from receptor selection. *Cell*, 92, 173-182.

- Melamed, D. & Nemazee, D. (1997). Self-antigen does not accelerate immature B cell apoptosis, but stimulates receptor editing as a consequence of developmental arrest. *Proc Natl Acad Sci U S A*, 94, 9267-9272.
- Meng, G., Inazawa, J., Ishida, R., Tokura, K., Nakahara, K., Aoki, K. & Kasai, M. (2000). Structural analysis of the gene encoding RP58, a sequence-specific transrepressor associated with heterochromatin. *Gene*, 242, 59-64.
- Murre, C. (2005). Helix-loop-helix proteins and lymphocyte development. *Nat Immunol*, 6, 1079-1086.
- Ohtaka-Maruyama, C., Miwa, A., Kawano, H., Kasai, M. & Okado, H. (2007). Spatial and temporal expression of RP58, a novel zinc finger transcriptional repressor, in mouse brain. *J Comp Neurol*, 502, 1098-1108.
- Okado, H., Ohtaka-Maruyama, C., Sugitani, Y., Fukuda, Y., Ishida, R., Hirai, S., Miwa, A., Takahashi, A., Aoki, K., Mochida, K., Suzuki, O., Honda, T., Nakajima, K., Ogawa, M., Terashima, T., Matsuda, J., Kawano, H. & Kasai, M. (2009). The transcriptional repressor RP58 is crucial for cell-division patterning and neuronal survival in the developing cortex. *Dev Biol*, 331, 140-151.
- Ossipow, V., Laemmli, U. K. & Schibler, U. (1993). A simple method to renature DNA-binding proteins separated by SDS-polyacrylamide gel electrophoresis. *Nucleic acids research*, 21, 6040-6041.
- Painter, M. W., Davis, S., Hardy, R. R., Mathis, D. & Benoist, C. (2011). Transcriptomes of the B and T lineages compared by multiplatform microarray profiling. *J Immunol*, 186, 3047-3057.
- Rada-Iglesias, A., Ameer, A., Kapranov, P., Enroth, S., Komorowski, J., Gingeras, T. R. & Wadelius, C. (2008). Whole-genome maps of USF1 and USF2 binding and histone H3 acetylation reveal new aspects of promoter structure and candidate genes for common human disorders. *Genome research*, 18, 380-392.
- Rao, S., Procko, E. & Shannon, M. F. (2001). Chromatin remodeling, measured by a novel real-time polymerase chain reaction assay, across the proximal promoter region of the IL-2 gene. *Journal of immunology*, 167, 4494-4503.
- Reynes, M., Aubert, J. P., Cohen, J. H., Audouin, J., Tricottet, V., Diebold, J. & Kazatchkine, M. D. (1985). Human follicular dendritic cells express CR1, CR2, and CR3 complement receptor antigens. *J Immunol*, 135, 2687-2694.
- Ross, M. E., Zhou, X., Song, G., Shurtleff, S. A., Girtman, K., Williams, W. K., Liu, H. C., Mahfouz, R., Raimondi, S. C., Lenny, N., Patel, A. & Downing, J. R. (2003). Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood*, 102, 2951-2959.
- Takahashi, K., Kozono, Y., Waldschmidt, T. J., Berthiaume, D., Quigg, R. J., Baron, A. & Holers, V. M. (1997). Mouse complement receptors type 1 (CR1;CD35) and type 2 (CR2;CD21): expression on normal B cell subpopulations and decreased levels during the development of autoimmunity in MRL/lpr mice. *J Immunol*, 159, 1557-1569.
- Tedder, T. F., Clement, L. T. & Cooper, M. D. (1984). Expression of C3d receptors during human B cell differentiation: immunofluorescence analysis with the HB-5 monoclonal antibody. *J Immunol*, 133, 678-683.
- Ulgianti, D. & Holers, V. M. (2001). CR2/CD21 proximal promoter activity is critically dependent on a cell type-specific repressor. *J Immunol*, 167, 6912-6919.

- 758 Ulgiati, D., Pham, C. & Holers, V. M. (2002). Functional analysis of the human
759 complement receptor 2 (CR2/CD21) promoter: characterization of basal
760 transcriptional mechanisms. *J Immunol*, *168*, 6279-6285.
- 761 Valverde-Garduno, V., Guyot, B., Anguita, E., Hamlett, I., Porcher, C. & Vyas, P.
762 (2004). Differences in the chromatin structure and cis-element organization of
763 the human and mouse GATA1 loci: implications for cis-element identification.
764 *Blood*, *104*, 3106-3116.
- 765 Woo, A. J., Dods, J. S., Susanto, E., Ulgiati, D. & Abraham, L. J. (2002). A
766 proteomics approach for the identification of DNA binding activities observed
767 in the electrophoretic mobility shift assay. *Molecular & cellular proteomics :
768 MCP*, *1*, 472-478.
- 769 Yokoyama, S., Ito, Y., Ueno-Kudoh, H., Shimizu, H., Uchibe, K., Albini, S.,
770 Mitsuoka, K., Miyaki, S., Kiso, M., Nagai, A., Hikata, T., Osada, T., Fukuda,
771 N., Yamashita, S., Harada, D., Mezzano, V., Kasai, M., Puri, P. L.,
772 Hayashizaki, Y., Okado, H., Hashimoto, M. & Asahara, H. (2009). A systems
773 approach reveals that the myogenesis genome network is regulated by the
774 transcriptional repressor RP58. *Dev Cell*, *17*, 836-848.
- 775
776
777
778

FIGURE LEGENDS

Figure 1: Deletion and mutation analysis of tandem E-box sites reveals cell-type specific effect of tandem E-box motifs. (A) Sequence of E-box double deletion/mutant (D/M) reporter constructs transfected in Raji (B) and K562 (C) cells. (D) Sequence of E-box single deletion or mutant reporter constructs transfected in Raji (E) and K562 (F) cells. Transcriptional activity (y-axis) represents luciferase/renilla luminescence normalized to wild-type (WT) 315-luc constructs $\pm$ SD (n = 5–10).

Figure 2: Linker scanning mutagenesis reveals critical nucleotides for E-box site 1 activator function. (A) Wild-type *CR2/CD21* promoter sequence encompassing E box 2 consensus motif and linker scanning reporter mutants (dashed lines represent sequence identity). (B) Reporter mutants and WT -315luc were transiently transfected into K562 (CR2-negative) and Raji (CR2-positive) cell lines. Results are expressed as the mean fold increase over WT 315 luc $\pm$ SD (n = 5–10).

Figure 3: No detectable cell-specific effects due to spacing of tandem E-box motifs. (A) Schematic depiction of E-box site phasing. (B) Sequence of deletion and insertion constructs that were transiently transfected into (C) Raji (CR2-positive; left panel) and K562 (CR2-negative; right panel) cell lines and normalized against the -315/+75 wild type (WT). Results are expressed as the mean fold increase over 315WT-luc $\pm$ SD (n = 5–10).

Figure 4: Identification of RP58 interacting with E-box site 2 in the *CR2/CD21* promoter. K562 nuclear extracts (NE) were fractionated using Sephacryl S-300 HR columns. (A) EMSA was performed using the oligonucleotide probe spanning -73 to -52 of the *CR2/CD21* promoter, incubated either with crude NE (lane 1) or protein fractions of decreasing sizes (lanes 2-13) identifying three major protein-DNA complexes (A-C).

806 Fractionated K562 NE were separated by 2DE (left panel) and gel slices were excised within
 807 the molecular weight range of 20-60 kDa with differing iso-electric points (right panel).
 808 Protein was recovered from gel slices and subject to EMSA analysis (C; 1-10) together with
 809 crude NE and the S300 protein fraction that retained complex C binding activity. Protein from
 810 gel slices that formed complex C (5 and 6) were subject to mass spectrometry.

811

812 **Figure 5: RP58 represses *CR2/CD21* promoter transcriptional activity *in vitro* via E-box**
 813 **site 2.** Transcriptional activity of either WT 315 luc, E-box mutant series (black bars) or the
 814 SV40 minimal promoter (grey bar/dashed line) co-transfected with RP58-expression
 815 construct. Results are expressed as the mean fold decrease over WT 315 luc $\pm$ SD (n = 5–10).

816

817 **Figure 6: E2A shows opposing activity to RP58 and activates *CR2/CD21* promoter**
 818 **transcriptional activity *in vitro* via E-box site 2.** (A) WT 315 luc was co-transfected with an
 819 empty vector (pcDNA2) or RP58-, E47- or E12/E47-overexpression constructs in K562
 820 (white bars) and Raji (black bars) cell lines. Co-transfection of E47 (B) or E12/E47 (C) with
 821 either WT 315 luc or E-box mutants. Results are expressed as the mean fold decrease over
 822 WT 315 luc $\pm$ SD (n = 5–10).

823

824 **Figure 7: Enhanced chromatin accessibility surrounding tandem E-box elements**
 825 **correlates with cell-specific binding of E12, E47, USF1 and USF2 in B cells.** ChIP assays
 826 using Raji (black bars) and K562 (white bars) cell lines demonstrates specific enrichment of
 827 *CR2/CD21* promoter in Raji cells by antibodies targeting USF1, USF2, E12 and E47
 828 compared to IgG and no antibody controls (A). Results show the mean $\pm$ SEM (n = 2–5)
 829 enrichment as a percentage of total input determined by Q-PCR and (B) gel analysis of PCR
 830 products from a representative experiment (Raji; CR2pA primers). (C) E12 and E47 bind a
 831 *GATA-1* hypersensitive site in K562 cells but not in Raji cells. Results show the mean $\pm$ SEM
 832 (n = 2 [Raji] or n = 3 [K562]) enrichment as a percentage of total input determined by Q-PCR
 833 and (D) gel analysis of PCR products from a representative experiment (K562; GATA1HS3.5

primers). (E) USF1 and USF2 bind the *DOLK* promoter in both K562 and Raji cells. Results show the mean $\pm$ SEM ($n = 2$) enrichment as a percentage of total input determined by Q-PCR and (F) gel analysis of PCR products from a representative experiment (K562 and Raji; *DOLKp* primers). (G) Chromatin accessibility assays were performed using Raji (black bars) and K562 (white bars) cells with the restriction enzyme *Dde* I. Accessibility was determined using Q-PCR with primers flanking *Dde* I sites located at -1143/-1139 (CR2pF), -284/-280 (CR2pD) and -55/-51 (CR2pB), or flanking *Bst*NI sites located at -77/-73 (CR2pB) and +634/+638 (CR2intron) and is plotted as the ratio of uncut:cut (mean $\pm$ SD, $n = 3$ for *Dde* I; $n = 5$ for *Bst* NI).

Figure 8: RP58 and E2A are co-expressed in human B cells. (A) Immunodetection of E2A (upper panel), RP58 (middle panel) or β actin (lower panel) was performed with equal amounts of total protein from K562, Raji or primary human B cells. Bands of the expected sizes, 70 kDa, 58 kDa and 42 kDa were detected with antibodies for E2A, RP58 and β actin respectively. (B) Promoter activity of *RP58*, *TCF3*, *CR2/CD21*, *CD19* and *PAX5* using FANTOM5 normalized robust peak based expression data for CD19 B cells ($n=11$) and blood lineage non B cells ($n = 54$; including stem cells ($n = 9$), monocytes ($n = 12$), CD4 T-cells ($n = 22$) and CD8 T-cells ($n = 11$)). (C) *RP58* promoter activity in stem cells, blood lineage cells (described in B), muscle ($n=9$ skeletal muscle; $n=3$ skeletal muscle satellite cells; and $n=3$ myoblast cells) and brain tissue ($n=3$ neuron; $n=4$ cerebellum; $n=4$ hippocampus).

856 **Table I: Oligonucleotide sequences**

	Forward (5' to 3')	Reverse (5' to 3')	Amplicon
CR2intronic	CTCTATTTTCATATTCATGTTTCAGCA C	CAGTTATGCATATTAACCAAAGT TCAG	237 bp
CR2pF	GATGTGGATTTCGCCTATCCC	CCCTGAAGGTAGTGGTGTAAA GC	164 bp
CR2pD	CAAACTGGGAGATGAATCCAC	CACAAGTTAGTACTCACCT	148 bp
CR2pB	TCCCGCTTGTCCACCCTCA	AGGCGGGCCCTTAAATAGTGT CC	104 bp
CR2pA	GCTCACAGCTGCTTGCTGCT	GCATCGCGGCACAGAACTTT C	224 bp
GATA-1 HS 3.5	CATTGAAGAAGAGTCTCAAACCTT AG	GAGCTGGAAACCCCGAGGCTG TG	241 bp
DOLKp	CCACTTGTATTGGACGTAGAAGG	AGAGGCGGCAGTAGTGTTTG	115 bp
CR2/CD21 E- box site 2	CCGCCTGCAGGTGTGCGCTCAG	CTGAGCGCACACCTGCAGGCG G	N/A (probe)

857

B

Raji

K562

A

WT 315-CR2/CD21luc over-expression assay

B

Figure 7

Figure 8