*5336–5350* *Nucleic Acids Research, 2010, Vol. 38, No. 16* *Published online 26 April 2010*
*doi:10.1093/nar/gkq268*

# An extended set of PRDM1/BLIMP1 target genes links binding motif type to dynamic repression

**Gina M. Doody[1], Matthew A. Care[1,2], Nicholas J. Burgoyne[2], James R. Bradford[3], Maria Bota[2], Constanze Bonifer[1], David R. Westhead[2] and Reuben M. Tooze[1,*]**

[1]Section of Experimental Haematology, Leeds Institute of Molecular Medicine, University of Leeds, Leeds LS9 7TF, [2]Bioinformatics Group, Institute of Molecular and Cellular Biology, University of Leeds, Leeds LS2 9JT and [3]Applied Computational Biology and Bioinformatics, Paterson Institute of Cancer Research, University of Manchester, Manchester, UK

Received October 16, 2009; Revised March 11, 2010; Accepted March 30, 2010

## ABSTRACT

**The transcriptional repressor B lymphocyte-induced maturation protein-1 (BLIMP1) regulates gene expression and cell fate. The DNA motif bound by BLIMP1 *in vitro* overlaps with that of interferon regulatory factors (IRFs), which respond to inflammatory/immune signals. At such sites, BLIMP1 and IRFs can antagonistically regulate promoter activity. *In vitro* motif selection predicts that only a subset of BLIMP1 or IRF sites is subject to antagonistic regulation, but the extent to which antagonism occurs is unknown, since an unbiased assessment of BLIMP1 occupancy *in vivo* is lacking. To address this, we identified an extended set of promoters occupied by BLIMP1. Motif discovery and enrichment analysis demonstrate that multiple motif variants are required to capture BLIMP1 binding specificity. These are differentially associated with CpG content, leading to the observation that BLIMP1 DNA-binding is methylation sensitive. In occupied promoters, only a subset of BLIMP1 motifs overlap with IRF motifs. Conversely, a distinct subset of IRF motifs is not enriched amongst occupied promoters. Genes linked to occupied promoters containing overlapping BLIMP1/IRF motifs (e.g. *AIM2*, *SP110*, *BTN3A3*) are shown to constitute a dynamic target set which is preferentially activated by BLIMP1 knock-down. These data confirm and extend the competitive model of BLIMP1 and IRF interaction.**

## INTRODUCTION

The transcriptional repressor PR domain containing 1 (PRDM1), better known as B-lymphocyte-induced maturation protein-1 (BLIMP1), is essential for normal development and immunity (1). In the immune system, the central role of BLIMP1 during terminal B-cell differentiation is indicated by the failure of plasma cell differentiation in murine *Blimp1* knockout B-cells and the repression of key gene expression programmes following ectopic BLIMP1 expression in human cell lines (2–4). In the absence of *Blimp1* in T-cells, mice develop an inflammatory bowel disease, which reflects a role for BLIMP1 in the control of effector T-cell populations (5,6). This has been underscored by recent data establishing the role for BLIMP1 in the control of specific CD4 and CD8 T-cell populations (7–11).

BLIMP1 represses transcription by binding to specific target sites in DNA and recruiting co-regulatory molecules (12–16). BLIMP1 contains five carboxy-terminal C2H2 Zn-finger domains, the first four of these are highly evolutionarily conserved, and of these, the first two are necessary to confer DNA binding specificity (12,17). To date, only a limited number of regulatory elements have been characterized as direct targets of BLIMP1 and the majority of these have been identified in promoters (18). During plasma cell differentiation, BLIMP1 acts via promoter elements to repress core transcription factors such as *PAX5*, *MYC* and *CIITA* (19–21). However, BLIMP1 can also act via distant regulatory elements, and several important regulatory associations of this type have been recently identified in T cells (22,23).

The function of BLIMP1, interferons (IFNs) and the interferon regulatory factor (IRF) family of transcription factors are closely interlinked, as originally indicated by the isolation of BLIMP1 as a post-induction repressor of the *IFN-β* promoter (24). The optimal DNA sequences bound by human and murine BLIMP1 *in vitro* are essentially identical, and similar to that bound by the IRF family, in particular IRF1 and IRF2 (25–27). The nature of the consensus sequence recognized by BLIMP1 *in vitro* predicts the existence of at least two distinct sets of

*To whom correspondence should be addressed. Tel: +44 113 3438415; Fax: 44 113 3438502; Email: r.tooze@leeds.ac.uk
The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

BLIMP1 target sites. These either show minimal overlap with IRF binding sites (e.g. *PAX5*) or overlap with extended IRF elements (IRF-E; e.g. *PSMB8*, *PSMB10*). This distinction is functionally significant since at promoters containing BLIMP1 sites which overlap with IRF-Es, binding of BLIMP1 and IRFs is mutually exclusive and promoter activity is regulated by the antagonistic action of these transcription factors (28,29). Conversely, not all IRF sites are predicted to be recognized by BLIMP1 (24,26). Two point mutations are sufficient to eliminate BLIMP1 binding to IRF-Es, while preserving IRF binding and IFN responsive regulation (28,29). IRFs play critical roles in mediating responses to cytokines and innate immune stimuli, and may themselves act as positive and negative regulators of promoter activity (30). The similarity in motifs selected *in vitro* suggests that the overlap in sequence recognition by IRFs and BLIMP1 imposes an additional level of control in cells in which these factors are co-expressed. However, in the absence of an unbiased analysis of *in vivo* occupancy, the generality of these predictions is uncertain.

To address this question, we have defined an extended set of promoters directly bound by BLIMP1 in human cells. We use this target set to establish the optimal DNA sequence motifs bound by BLIMP1 *in vivo*, capturing subtle complexity of binding specificity, and a link between motif type and promoter CpG content. We demonstrate that only a subset of occupied promoters is associated with overlapping BLIMP1 and IRF motifs, and conversely that a subset of IRF motifs is not associated with BLIMP1 occupancy. In a model system in which BLIMP1 is co-expressed with IRFs we find that genes linked to overlapping motifs, despite having amongst the highest predicted BLIMP1 binding affinities, are the most labile target genes upon BLIMP1 knock-down. This unbiased experimental approach thus confirms two key predictions of the antagonistic regulation model, providing further evidence for the functional linkage between BLIMP1 and IRFs.

## METHODS

### Cell lines, antibodies and siRNA

U266 and Daudi cells were maintained in RPMI with 10% heat-inactivated FCS. Antibodies were: rabbit antisera to BLIMP1 (28), rabbit polyclonal antibodies to IRF-1 and IRF-2 (Santa Cruz) non-immune rabbit IgG (Upstate Biotechnology) and mouse monoclonal antibody to β-ACTIN (clone AC15, Sigma).

For siRNA-knockdown, U266 cells were electroporated with 400 pmol of siRNA (200 pmol each BLIMP1 oligo or 400 pmol Control oligo) using a BioRad Gene Pulser at 250 V and 960 μF. Twenty-four hours post-transfection, cells were processed for protein and mRNA or chromatin evaluation. The oligos used were:

BLIMP1 siRNA #1 sense sequence 5′-GGAAAGGACC UCUACCGUU-3′

BLIMP1 siRNA #2 sense sequence 5′-GAUCUGACCC GAAUCAAUG-3′

Control siRNA sense sequence 5′-CUACCUCUAGAAC GGACGU-3′

### ChIP amplification, target validation and mRNA quantification

Chromatin immunoprecipitation and T7 amplification were performed as described except RNA was purified with TRIzol (Invitrogen) (28,31).

Quantitative real-time PCR was performed using SYBR Green MasterMix on an ABI Prism 7500 system (Applied Biosystems). The precipitated material was quantified against a standard curve of input DNA and normalized relative to control immunoprecipitates. RNA was purified with TRIzol (Invitrogen) and subject to DNase I treatment (DNAFree, Ambion). cDNA was generated using oligo-dT primer and Superscript II (Invitrogen). Relative mRNA expression was normalized to *GAPDH*.

### Arrays and hybridization

NimbleGen RefSeq hg18 promoter tiling arrays (design ID 4226, design name 2006-07-18_HG18_RefSeq_promoter) were used. Labelling, hybridization and data acquisition was performed by NimbleGen according to standard procedures with the modification that amplified material from control and specific ChIPs were directly co-hybridized. Three independent control and specific samples from the U266 cell line were evaluated. A dye-swap replicate was included for one pair of samples. For the control Daudi cell line, two independent control and specific samples were hybridized.

### Peak detection and filtering

Peak detection was performed with the MA2C algorithm, using a sliding window of 500 bp and a *P*-value cut-off of 0.001 (32). Data from the negative control Daudi cell line provided experimental assessment of false positive detection rates. For further details, see Supplementary Methods.

### *De novo* motif analysis and matrix scoring

*De novo* motif detection was performed using Weeder version 1.3 and Oligo-analysis (33,34). Sequence logos were drawn using enoLOGOS (35). The STAMP web-based tool for comparison of position weight matrices (PWM) was used with default parameters (36).

Matches to PWMs were scored using RSAT matrix-scan program (37). Distributions of all matches at a range of predicted *P*-value thresholds were evaluated in the MA2C defined peak regions and in $10^6$ randomly selected sets of promoter regions of matched number, length and position. Enrichment and statistical significance was determined by comparison of observed frequency in peak regions and expected frequency derived from $10^6$ randomizations. For further details, see Supplementary Methods.

### Promoter CpG content and conservation

Please see Supplementary Methods section for details of analyses of promoter CpG content, and evolutionary conservation.

### Expression analysis

Gene expression was analyzed using Nimblegen arrays (HG18_60mer_expr, hg18, NCBI Build 36); four BLIMP1 knockdown and four control siRNA samples were analyzed twice, giving a total of 16 data sets. Labelling, hybridization and data acquisition was performed by NimbleGen according to standard procedures. Following analysis of raw data and comparison of correlation, three BLIMP1 knockdown and three control arrays were removed from analysis due to abnormal signal distributions. Remaining arrays were normalized with robust multichip average (RMA) (38). A linear model was fitted using the R Limma package and significance of differential expression was gauged using Limma empirical Bayes statistics module with probabilities adjusted for multiple testing (Benjamini & Hochberg) (39). The enrichment of groups of BLIMP1 target genes defined by motif type was assessed using a hypergeometric test of all genes showing increases or decreases in expression above/below a range of fold-change cut-offs following BLIMP1 knockdown.

### Luciferase assays

BLIMP1, IRF1 and control pIRES2-EGFP (Clontech) modified by Myc-tag were as previously described (28,40). pGL4-based luciferase reporter constructs for promoters of *AIM2* (hg17_chr1-:155859556-155860496), *BTN3A3* (hg17_chr6+:26,547,952-26,548,891) and *SP110* (hg17_chr2-:230909991–230910891) were purchased from SwitchGear Genomics. For mutated promoters, equivalent insert sequences were synthesized (MrGene) with overlapping BLIMP1/IRF sites identified in Supplementary Table S1 mutated by the insertion of XhoI restriction sites: *AIM2* (TACTTTCGCTT to TACTCGAGAAT and CACTTTCTGTTTC to CACCTCGAGAATC), *BTN3A3* (CTTTCACTTTTC to CTCTCGAGAATC), *SP110* (ACTTTCACTTTTC to ACCTCGAGAATTC).

For luciferase assays, three replicate transfections were performed for each condition. The indicated quantities of expression vectors were co-transfected with reporter luciferase vector and pRL-CMV control. For each condition, the total amount of vector transfected was normalized by inclusion of control pIRES2-EGFP plasmid. Experiments were done using the Promega luciferase assay system and analyzed on a Berthold Lumat LB Luminometer. Each experiment was performed in duplicate with similar results. The data are from a representative experiment displayed as fold increase in relative light units relative to unstimulated cells co-transfected with the empty vector.

### Electrophoretic mobility shift assay

Nuclear extracts were prepared as previously described from COS cells transfected with expression vectors for BLIMP1 or IRF1 (28,40). For electrophoretic mobility shift assay (EMSA), the double-stranded probes used contained the following sequences, only one strand is shown:

*AIM2* (5′-ATGAGACCCAAGCGAAAGTATAGACATTTTA-3′)
*BTN3A3* (5′-TGAAATGAAAAGTGAAAGTACATTGAATTGT-3′)
*SP110* (5′-AGAAAAGAAAAGTGAAAGTTTTCGGAGTTTG-3′)
*IVNS1ABP* (5′-CGCACAGGGCGGCGAAAGTGAGGGTGCGGGG-3′)
*RTN4IP1* (5′-AAGTGGAGGGGGTGAAAGTGAGGGAGGAGAA-3′)
*TLN1* (5′-CCTCCAACGCAGCGAAAGTGACGCACCCCAC-3′)

For methylated oligonucleotides, the following sequences were used with symmetric methylation of both strands, only the upper strand is shown:

*AIM2* F (5′-ATGAGACCCAAGC(Me)GAAAGTATAGACATTTTA-3′)
*TLN1* F (5′-CCTCCAAC(Me)GCAGC(Me)GAAAGTGACGCACCCCAC-3′)

DNA probes [$^{32}$P]-labelled with T4 polynucleotide kinase were incubated with nuclear extract in the presence of poly [dI:dC] (Amersham) for 30 min at room temperature. Supershift was performed by the addition of antisera to the extract prior to mixing with radioactive probe and competition assays included the addition of unlabelled probe to the reaction mixture.

## RESULTS

### Defining BLIMP1 occupied promoters

To establish an extended unbiased set of BLIMP1 targets, we employed NimbleGen RefSeq promoter tiling arrays, to examine ChIP samples generated from myeloma cell line U266. The arrays include predicted promoters for 19 178 human RefSeq genes from on average 2.2 kb upstream to 0.5 kb downstream of the transcriptional start site (TSS), at density of 1 tile/100 bp. ChIP samples were subject to T7 linear amplification and equivalent enrichment of known targets was demonstrated in amplified material prior to further analysis (Supplementary Figure S1) (31). Representative primary data for promoters of *STAT6*, *GLI3* and *SP110* are shown (Figure 1A–C).

To test the overall false positive detection rate for analysis, we performed BLIMP1 ChIP-chip from a non-expressing cell line, Daudi. Peak detection was performed using model-based analysis of two-colour arrays (MA2C) and at a $P = 0.001$ cut-off only a single peak across all tiled promoter regions was identified in duplicate samples from the Daudi cell line (32). At this cut-off, 286 promoter regions were identified as BLIMP1 occupied across the four U266 data sets (Supplementary Table S1), and did not include the single Daudi peak. This is similar to the number of high-confidence targets identified for the transcription factor XBP-1 in plasma cells (41). Relaxing

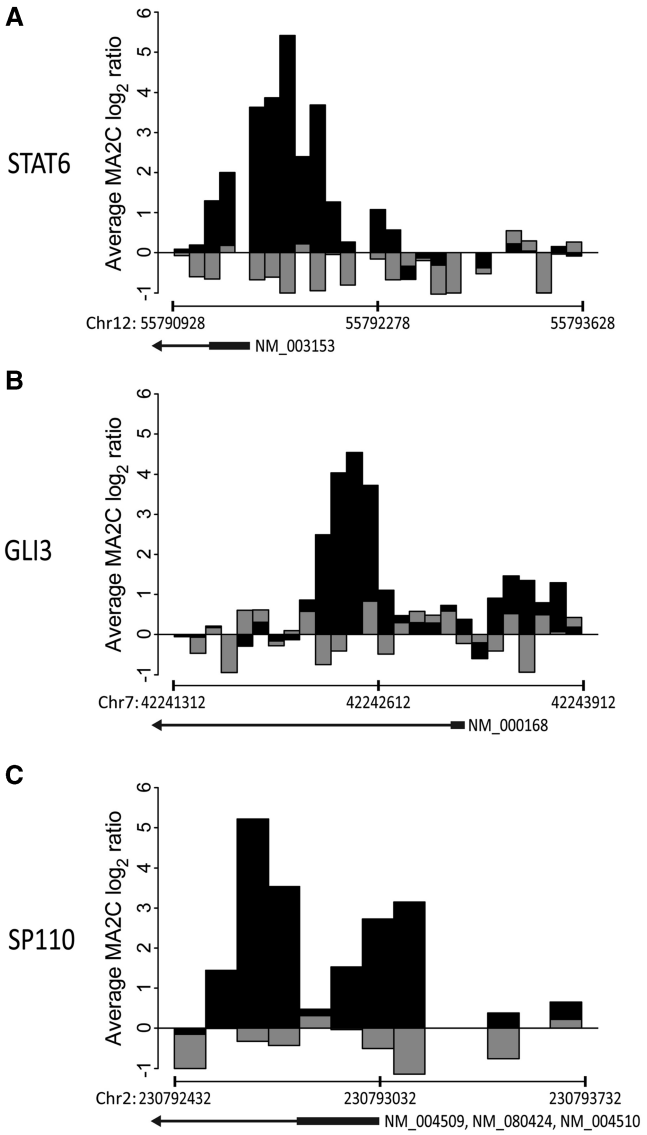


**Figure 1.** ChIP-chip identifies BLIMP1 promoter occupancy. Representative primary ChIP-chip data from genomic regions including the promoters for *STAT6* (**A**), *GLI3* (**B**) and *SP110* (**C**). Each panel shows the average of U266 (black bars) and Daudi (grey bars) data sets. The *y*-axis shows the log2 ratio of signal derived from hybridization of specific BLIMP1 against control rabbit Ig ChIP samples. The *x*-axis shows the genomic coordinates of micro-array tiles. The position of the TSS for each gene is displayed.

MA2C analysis stringency to identify additional occupied promoters, resulted in a disproportionate increase in false positive peaks in the Daudi data set (data not shown). A low false positive detection rate was prioritized, and only the results of stringent settings were used for further study.

Four genes (*ELMOD3*, *RGS3*, *CCL4L1* and *HIST2H3C*) were associated with two distinct regions of promoter occupancy. Forty peak regions were in bi-directional promoters, including the two peaks associated with *ELMOD3*. Thus, a total of 322 unique genes were linked with the 286 occupied promoter regions (Supplementary Table S1). Overall, the BLIMP1 peaks demonstrated a strong clustering with 60% of peaks present within 0.5 kb of the predicted TSS (data not shown), consistent with preferential BLIMP1 binding in the vicinity of predicted basal promoter elements.

Amongst the 286 high-confidence peaks identified in the U266 data were the known BLIMP1 target sites of *CIITA-pIII, MYC, PSMB8*, *PSMB10* and *TAPBP*. To validate new targets, we chose a set of 27 peak regions across a range of MA2C scores reflecting the robustness of ChIP-chip signal (MA2C peak score for all targets: range 3.8 to 0.76 mean peak score 1.22; MA2C peak score for tested targets: 3.8 to 0.77, mean peak score 1.65) to test for BLIMP1 occupancy by conventional qPCR analysis. The enrichment observed for the known target *PSMB8* ($3.2 \pm 0.6$) was set as a threshold for a confirmed interaction, equivalent to the weakest qPCR enrichment observed at previously defined targets (29). Amongst the test set, 25/27 (92.6%) sequences were enriched greater than or equal to this 3-fold threshold (Figure 2A). To further assess the validity of the ChIP-chip data, a predicted peak region was analysed in detail by qPCR. *STAT6* was chosen, since this gene has previously been reported as repressed on BLIMP1 over-expression (3). Comparison of results between ChIP-chip and qPCR analyses of this promoter demonstrated a high degree of similarity in profile. Moreover, the pattern of binding was also confirmed in an alternate myeloma cell line, H929 (Figure 2B).

### *De novo* motif detection confirms BLIMP1 binding specificity

Two *de novo* motif discovery programs, Weeder and Oligo-analysis, were employed to determine whether a common BLIMP1 binding motif could be identified amongst our peak regions, without prior knowledge of motif composition. These programs rely on different algorithms and have performed well in head-to-head comparison with alternate programs (33,34,42,43). As a control, five random sets of promoter sequences, matched to the BLIMP1 peak regions for length and distribution relative to TSS, were analysed. Amongst the 286 peak regions, the Weeder program identified 14 distinct motifs of between 6 and 12 bp in length (Figure 3A and Supplementary Table S2). By visual inspection, 13/14 of these matrices appeared related, representing variations on a common theme. This relationship was confirmed by analysis with the STAMP web-tool, which compares similarities between position weight matrices (PWMs) and PWM databases (Supplementary Figure S2A) (36). The results obtained with the Weeder program were further confirmed by independent analysis of the peak regions with the Oligo-analysis program, which identified an extended match to the BLIMP1 consensus 'AAGTGAAAGTGAA', as the most significantly over-represented motif across the peak regions (data not shown) (34).

Analysis against TRANSFAC v11.3 demonstrated that the most significant single match amongst the 13 related matrices was to the TRANSFAC BLIMP1 motif (M001066) (Supplementary Figure S2B). However, even greater significance was observed when aligning against the high-affinity *in vitro* weight matrix for BLIMP1

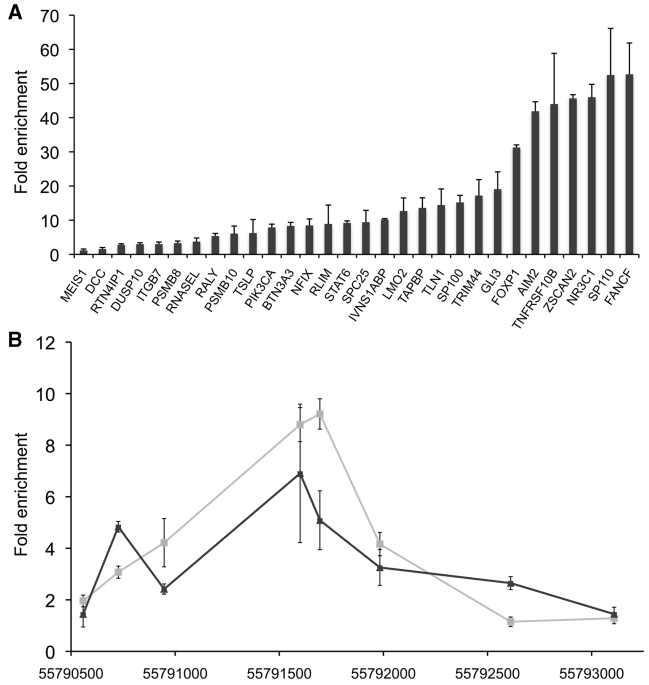


**Figure 2.** Validation of BLIMP1 occupied promoters. (**A**) Real-time PCR quantification of BLIMP1 promoter occupancy in triplicate ChIP samples from U266 cells. Data are displayed as mean fold enrichment ± SD relative to control rabbit IgG. Twenty-seven new peak regions as well as three known targets (*PSMB8*, *PSMB10* and *TAPBP*) are shown. The known target *PSMB8*, was used as threshold (3.2 ± 0.6) to confirm interaction. (**B**) Quantification of BLIMP1 binding across the identified peak region spanning the *STAT6* promoter by real-time PCR. The *x*-axis shows co-ordinates on human chromosome 12 (hg18). Grey lines represent the mean fold enrichment ± SD from three ChIP samples relative to control IgG from U266 myeloma cells. Black lines show the equivalent data from an additional myeloma cell line H929 myeloma. The background enrichment at unoccupied promoter regions is below 2-fold.

defined by Kuo and Calame, which was not contained in the version of TRANSFAC tested. Nine of 13 matrices aligned more significantly with this matrix than with any matrix in the TRANSFAC database. Notably, other than the alignment with BLIMP1 matrices, the most significant matches were observed to PWMs of the IRF family. Amongst the five random promoter sets no matches to BLIMP1 or IRF matrices of similar significance were observed (data not shown).

## Motifs are highly enriched across a wide range of thresholds

In order to distinguish between the individual Weeder weight matrices (WWMs), we first asked how significant the motif enrichment was in occupied promoter regions for each WWM. To answer this question, we used a boot-strapping approach to compare the observed occurrence of motif matches in our peak regions to the expected occurrence of motif matches across all RefSeq promoters. Using the RSAT matrix-scan program the RefSeq promoters were scanned with each WWM, and the high affinity *in vitro* weight matrix for BLIMP1 (26,44). For each matrix, the expected distribution of motif matches was then determined by boot-strapping with random sets of promoter regions, matching the set of BLIMP1 occupied regions for number, length and position relative to TSS. This was performed across a full range of quality thresholds with $10^6$ random sets, generating accurate distributions of motif matches. Comparison with the observed distribution in the BLIMP1 bound peak regions established the degree of enrichment and the statistical significance for each WWM (Supplementary Figure S3A and data not shown). Indeed, the family of related WWMs was enriched in the bound promoter regions with very high significance (Table 1). This provides independent proof that the WWMs represent a high-quality matrix set with a very significant association between the presence of these motifs and BLIMP1 binding. However, the results also demonstrate that relative enrichment does not provide a clear single choice of optimal matrix.

## Combinations of motifs capture distinct aspects of binding specificity

Given the similarity and overall enrichment of the set of WWMs, we next asked whether individual matrices were redundant. We noted that the WWMs could be divided into subclasses (Figure 3A). Ten of 13 related motifs contained a common GAAAGT core, which is an extension of the GAAAG core selected by BLIMP1 *in vitro* (26). With increasing motif length, this core was flanked by 5′ or 3′ extensions. Amongst longer motifs two distinct sub-divisions were evident. First, matrices were characterized either by a strict 5′ GTG element, or a more degenerate 5′ GNG element. Secondly, WWMs of either type were asymmetric with either a dominant 5′ or 3′ extension. Notably, the two most enriched matrices identified in the boot-strapping analysis derived from either extreme of 5′ and 3′ motif extension (Figure 3A and Table 1).

To confirm the validity of this *in silico* prediction, we examined the ability of BLIMP1 to bind to oligonucleotides representing variants of the 5′ and 3′ motif extension types. The corresponding motifs for *AIM2, SP110* and *BTN3A3* (5′ motif extension type) as well as *TLN1*, *RTN4IP1* and *IVNS1ABP* (3′ extension type) were tested in EMSA with BLIMP1 containing nuclear extracts. In each case, a single complex with BLIMP1 was detected whose specificity was demonstrated by antibody supershift and competition with unlabelled probe (Figure 3B). These data confirm that motifs of both the 5′ and 3′ extension type are directly bound by BLIMP1.

The most enriched matrix variants thus appear to capture distinct aspects of BLIMP1 binding specificity. To directly test the degree of overlap between the WWMs, we compared the four matrices corresponding to GTG- or GNG-based motifs with longest 5′ or 3′ extensions. Each of these four closely related matrices identified different sets of significant sites amongst occupied promoters across a range of thresholds. Even between highly similar matrix pairs, there were considerable differences in binding site prediction. For example, when considering either the top 5 or 15% of matches each

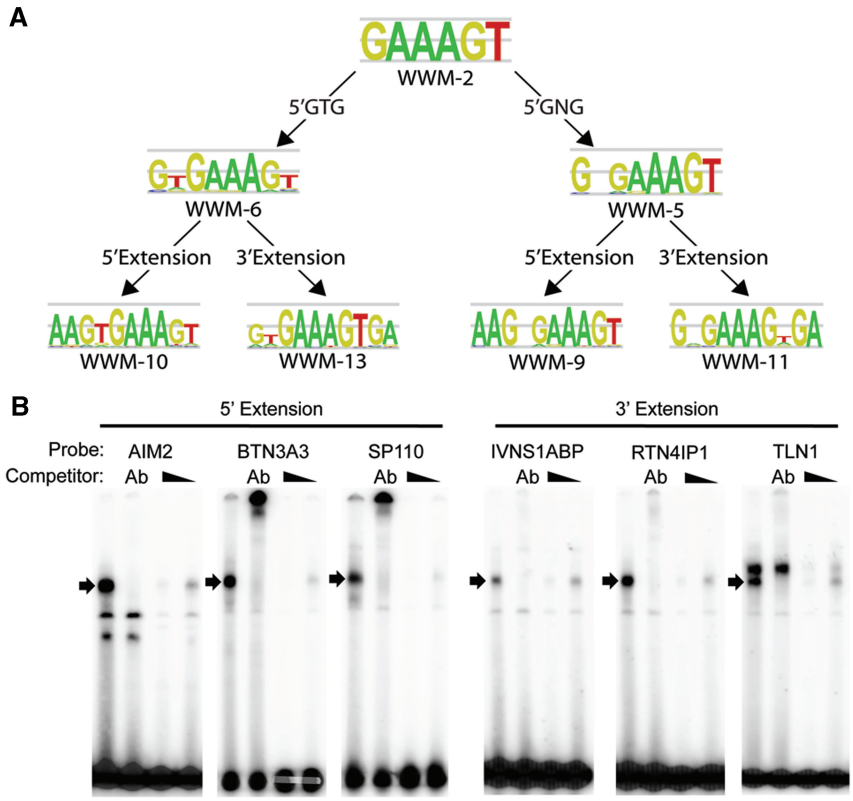


**Figure 3.** *De novo* motif detection of the *in vivo* BLIMP1 binding consensus. (**A**) Sequence logos for 7 out of 14 matrices derived by the Weeder program from occupied promoter regions are shown. These represent two themes around a common core element: a more 'GTG' or less 'GNG' restrictive 5′ triplet, and either additional 5′ or 3′ extensions. (**B**) EMSA was performed with nuclear extracts from *BLIMP1* transfected COS cells and oligonucleotides corresponding to a range of 5′ (*AIM2*, *BTN3A3* and *SP110*) or 3′ extension sites (*IVNS1ABP*, *RTN4IP1*, *TLN1*). BLIMP1 complexes (arrow) were identified by antibody supershift (Ab), and specificity of interaction was confirmed by competition with 100- or 10-fold excess of unlabelled probe (wedge).

**Table 1.** Enrichment of matrices in occupied promoter regions

| Matrix | Median enrichment | Median *P*-value | Median *Z*-score |
|---|---|---|---|
| *In vitro* | 5.63 | <10-6 | 16.96 |
| W10-BO | 4.59 | <10-6 | 15.94 |
| W13-BO | 4.89 | <10-6 | 15.79 |
| W9-BO | 4.15 | <10-6 | 13.57 |
| W5-BO | 3.92 | <10-6 | 13.23 |
| W7-BO | 2.84 | <10-6 | 11.87 |
| W6-BO | 3.09 | <10-6 | 11.57 |
| W4-BO | 3.17 | <10-6 | 11.00 |
| W2-BO | 1.79 | <10-6 | 10.41 |
| W12-BO | 3.34 | <10-6 | 10.33 |
| W11-BO | 2.84 | <10-6 | 8.45 |
| W8-BO | 2.50 | 0.000026 | 7.64 |
| W3-BO | 1.58 | <10-6 | 6.25 |
| W1-BO | 1.29 | 0.001454 | 3.17 |

Enrichment of matches to the family of related WWMs or the *in vitro* matrix for BLIMP1 was determined by comparison to $10^6$ randomly selected matched promoter regions for four motif quality thresholds (top 5, 10, 15 and 20%). Shown are the median enrichment, median *P*-values and median *Z*-scores across all thresholds.

of the WWM quartet (WWM-Q) contributed a distinct as well as overlapping set of binding site predictions (Figure 4A). Thus, none of these matrices could be regarded as redundant.

We next asked whether the *in vitro* derived BLIMP1 matrix contributes substantial information to the predictions of this quartet of matrices. The *in vitro* matrix is robustly enriched in bound promoter regions, to the same degree as the most enriched WWMs (Table 1), and at very stringent motif thresholds added information to binding site predictions of WWM-Q. However, with increasing degeneracy, the contribution of the *in vitro* matrix declined, and the top 15% of motif predictions of the WWM-Q captured all of the motif predictions of the *in vitro* matrix at this threshold (Figure 4B). In contrast, in order to capture the set of targets identified by the top 15% of WWM-Q matches with the *in vitro* matrix alone, 92% of all potential sites across the genome would have to be accepted as matches (data not shown). We conclude that a single matrix is not sufficient to encompass the binding specificity of BLIMP1, that *de novo* motif prediction from *in vivo* binding data captures complexity that is not evident from *in vitro* oligonucleotide selection, and that a quartet of matrices provides a more complete representation of BLIMP1 specificity. We will consider two consequences of the recognition of these motif variants.

## Motif type, CpGs and methylation sensitive DNA binding

The GNG element in WWM-9 and WWM-11 accommodates a group of motifs containing a GCG triplet (Figure 3A and Supplementary Table S1), which may be

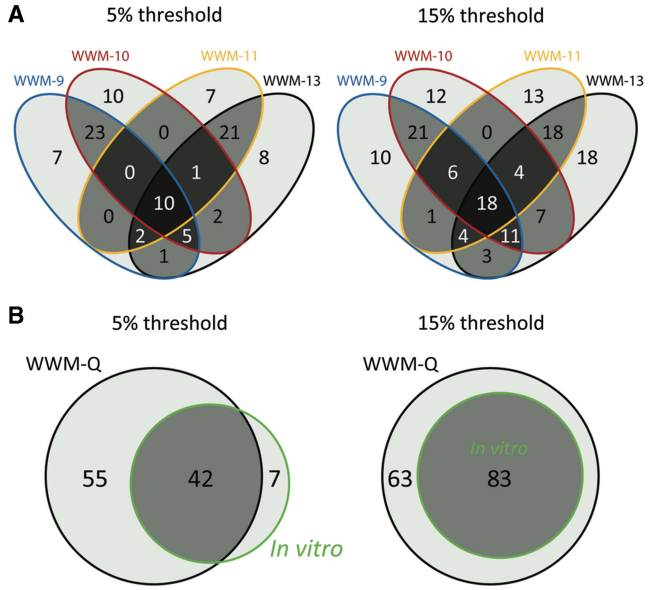


**Figure 4.** Related matrices provide non-redundant information. (**A**) Venn diagrams display the intersection between the sets of genes linked to promoters with significant matches to the quartet of matrices WWM-9, WWM-10, WWM-11 and WWM-13, which represent the extremes of motif variants. Results are shown at two thresholds: top 5% of matrix matches (left) and top 15% of matrix matches (right). (**B**) Displayed are the overlap between the *in vitro* matrix for BLIMP1 and the combined results of WWM-quartet (WWM-Q). Results are displayed for two thresholds: top 5% of matrix matches (left) and top 15% of matrix matches (right).

preferentially associated with CpG rich islands. Examining the overall distribution of CpG sequences in bound promoter regions demonstrated that BLIMP1 occupancy is not generally associated with a bias in CpG-content. Bound promoter regions follow a distribution broadly similar to the biphasic pattern of all RefSeq promoters (Figure 5A and B; 45). However, there was a significant ($P = 0.019$, Wilcoxon signed-rank test) increase in the CpG content of bound promoter regions associated with a high quality motif match of GCG type, relative to those without (Figure 5A and C). Since CpGs tend to be clustered within promoters, this enrichment most likely reflects binding within CpG islands. Indeed, for these promoters, the CpG content of the BLIMP1 bound regions exceeded the CpG content of the corresponding total promoter region (Figure 5D). Thus, while there is no overall skewing in the CpG content of BLIMP1 bound promoter regions, WWM-9 and WWM-11 identify a subset of BLIMP1 motifs preferentially associated with regions of increased CpG content.

Methylation sensitive recognition of CpG containing binding sites is a feature of the zinc-finger transcription factor CTCF, and plays an important role in regulating CTCF occupancy (46). We therefore asked whether BLIMP1 binding to CpG containing motifs shows a similar regulation, using GCG type motifs in the context of either 5′ (*AIM2*) or 3′ (*TLN1*) extensions. Indeed, methylation of a single CpG (*AIM2*) or two CpGs (*TLN1*) efficiently abolished the binding of BLIMP1 (Figure 5E). In addition, even when added at 100-fold excess unlabelled methylated oligonucleotides failed to compete with non-methylated oligonucleotides for BLIMP1 binding. In contrast, methylated oligonucleotides retained the ability to compete efficiently for binding with factors distinct from BLIMP1 (Figure 5E). These experiments confirm the idea that BLIMP1 binding to CpG-containing motifs could potentially be regulated by DNA-methylation *in vivo*.

### Defining the extent of overlap between BLIMP1 and IRF motifs at occupied promoters

BLIMP1 and IRFs share a minimal GAAA core at virtually all sites, but show distinct preferences for surrounding sequence contexts (26,27,47). This potentially allows for the existence of three groups of target sites, one shared set and one set selective for either IRFs or BLIMP1. For IRFs, a strongly preferred context is given by the extended AANNGAAA core sequence. This is supported by *in vivo* and *in vitro* target site selection and crystal structure (26,27,44). However, overlap with this AANNGAAA IRF core is only favoured in BLIMP1 matrices with 5′ extension, since these also have heavy weighting for a 5′ AA doublet. In contrast, the WWMs with 3′ extension (WWM-11 and -13) place no constraint on the equivalent positions and may identify sites unique to BLIMP1. Thus, our *de novo* motif analysis supports the existence of distinct sets of BLIMP1 binding sites, which either do or do not overlap with an extended IRF consensus. Amongst the high-quality (top 15% WWM-Q) BLIMP1 sites identified in occupied promoters 47% contain an AANNGAAA and 53% do not (Supplementary Table S3). In contrast, only 33% of all potential BLIMP1 sites across tiled RefSeq promoters, defined at the same WWM-Q threshold, are of the AANNGAAA type. Overlapping BLIMP1/IRF sites are thus significantly enriched amongst occupied promoter regions (Fischer's exact test $P = 1.59e\text{-}09$). However, these data also confirm that a substantial subset of BLIMP1 occupied promoters is characterized by motifs that do not overlap with the extended IRF consensus.

The intersection between BLIMP1 and IRF DNA binding specificities is likely to be evolutionarily ancient, since both *BLIMP1* and *IRF* orthologues exist in sea urchin, the zinc-finger domains of BLIMP1 are highly conserved, and BLIMP1 proteins encoded by evolutionarily distant orthologues display the same motif restricted ability to inhibit an IFNγ responsive promoter (17,29,48,49). Binding sites selected for regulation by both BLIMP1 and IRFs should therefore also show a higher degree of evolutionary conservation. To address this idea, a set of non-redundant BLIMP1 binding site predictions was generated for the bound promoter regions. Ranked lists were compiled according to conservation scores calculated from the sum of each position of the motif match for either primate or vertebrate conservation. Fischer's exact test was then used to compare the percentage of BLIMP1 sites with or without an IRF overlap in equal fractions from the top (most conserved) or bottom (least conserved) of this list. Across a broad range, overlapping sites showed a significant association

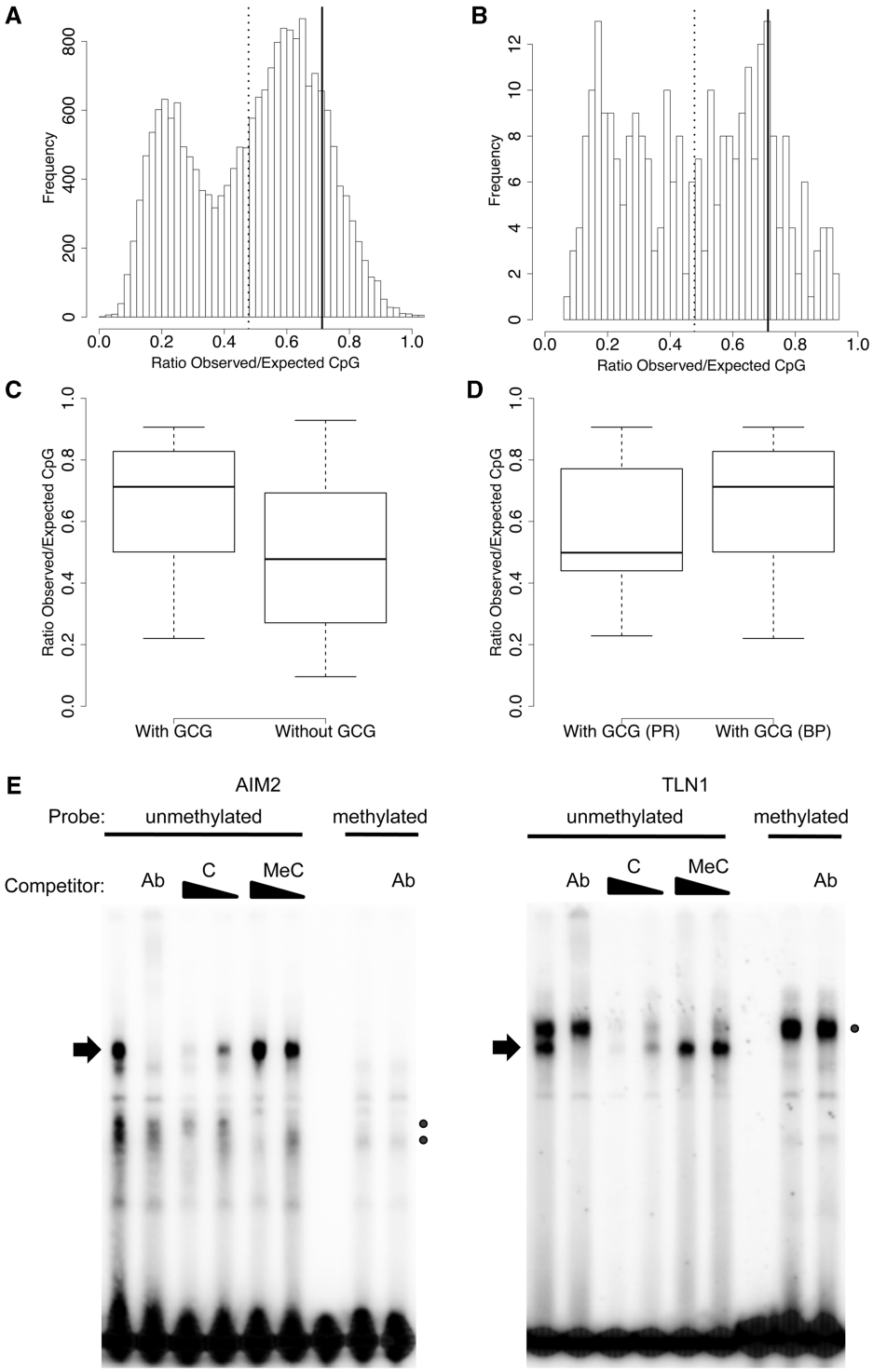


**Figure 5.** Motif type, CpGs and methylation sensitive DNA binding. (**A**) Graph of CpG content of all tiled promoter regions, and (**B**) graph of the distribution of CpG content in all BLIMP1 occupied promoters. Median CpG ratio of occupied promoter regions containing a top 15% WWM-Q match with GCG triplet (solid line) or a top 15% WWM-Q match without GCG triplet (dashed line). *y*-Axes, number of promoter regions; *x*-axes, ratio of observed to expected CpG content [number of CpG/(Number of C×Number of G)×sequence length], bin size = 0.02. (**C**) Box-plot of CpG content in bound peak regions containing a top 15% BLIMP1 motif match with GCG triplet (left) and a top 15% BLIMP1 motif match without GCG triplet (right), and (**D**) box-plot comparing tiled promoter (PR) regions to bound peak (BP) regions for occupied promoters with a top 15% BLIMP1 motif match containing a GCG triplet. *y*-Axes, ratio of observed to expected CpG content described above. (**E**) EMSA was performed with nuclear extracts from *BLIMP1* transfected COS cells and oligonucleotide probes corresponding to *AIM2* (left panel) or *TLN1* (right panel) sites, respectively. Within each panel, the left six lanes show EMSAs with a labelled unmethylated probe, whereas the right two lanes show EMSAs done with probes containing symmetrically methylated CpGs. The expected BLIMP1 complexes (arrows) were observed with the labelled unmethylated probe and confirmed by antibody supershift (competitor: Ab) but did not form with methylated probes. To confirm loss of BLIMP1 binding to methylated probes, cold competition was performed with a 100- or 10-fold excess of unmethylated unlabelled competitor (competitor: C) or methylated unlabelled competitor (competitor: MeC). Circles indicate other protein–DNA complexes distinct form BLIMP1 whose binding was not abolished by DNA methylation.

with better conservation scores (Table 2). To establish how motif quality related to conservation and motif type in occupied promoters, we examined the *P*-value distributions of the two BLIMP1 motif types in the top and bottom fractions of the ranked list. Similar distributions were observed for both primate and vertebrate conservation and demonstrated that relative to other sites, overlapping BLIMP1/IRF sites with higher conservation scores are skewed away from high and toward low *P*-values (Supplementary Figure S4). Thus, conserved BLIMP1/IRF sites have a tendency to be better quality motif matches than non-conserved sites. Such skewing was not evident for sites lacking IRF overlap. These findings support the contention that functional antagonism is subject to evolutionarily conservation.

We have previously shown that mutation of position 3 and 9 of the *in vitro* derived BLIMP1 consensus from G to C in the context of overlapping BLIMP1/IRF-E sites abrogates the ability of BLIMP1 to bind to DNA, but does not alter binding of IRF1/IRF2 or regulation by IFNγ (28,29). Amongst the top 15% of IRF1, IRF2 or ISGF-3 (ISRE) sites in human RefSeq promoters (defined using TRANSFAC M00062, M00063 and M00258), respectively, 21%, 14% and 22% of matches contain a C at both critical positions that prevent BLIMP1-dependent regulation. This predicts that BLIMP1 is excluded from a significant fraction of IRF sites. If this holds true across the genome, sequences related to the consensus AACTGAAACT should not be enriched amongst BLIMP1 occupied promoters. We therefore repeated the boot-strapping analysis of promoter regions using versions of the WWM-Q or the *in vitro* BLIMP1 matrix in which the corresponding position weighting was switched from G to C (Supplementary Table S2). In contrast to the original matrices, these 'mutated' matrices were minimally enriched amongst bound promoter regions, indicating that BLIMP1 distinguishes between these closely related sites (Figure 6 and Supplementary Figure S3A and B). Thus, the results we previously observed with a limited set of promoters can be extrapolated across the genome (28,29). This confirms that a subset of IRF sites remains unoccupied by BLIMP1, and that selectivity is exerted on both sides of the antagonistic BLIMP1/IRF module at the level of *cis*-regulatory element occupancy *in vivo*.

## Overlapping BLIMP1/IRF motifs identify sites of dynamic repression

The antagonistic regulation model predicts that in cells expressing competing BLIMP1 and IRFs, genes whose promoters contain overlapping BLIMP1/IRF motifs should be enriched amongst genes that are increased in expression after BLIMP1 knockdown. U266 cells express IRFs, including IRF1, which we have shown to activate gene expression in direct competition with BLIMP1-dependent repression (28,29). To verify that IRFs can compete with BLIMP1 for occupancy at target promoters in U266 cells, we assessed the binding of IRF1 to a selection of promoters linked to predicted overlapping BLIMP1/IRF sites. As expected IRF1 binding was detected by ChIP across the range of promoters tested (Supplementary Figure S5), demonstrating that these represent potential *in vivo* IRF targets and that this cell line therefore provides a model system in which to test the prediction.

We next examined gene expression changes after BLIMP1 knockdown in U266 cells (Figure 7A). First, 15 targets were tested by qPCR of which 12 showed an average increase of expression of ≥1.5-fold, with six genes showing ≥1.9-fold increases in expression (Figure 7B). We then examined global changes in gene expression by micro-array. BLIMP1 target genes showing the most significant changes included *BCAT1, LMO2, BTN3A3, CIITA, CSF1* and *SP110* (Supplementary Table S4). To assess whether the presence of a BLIMP1/IRF motif was associated with an increase in gene expression, we subdivided BLIMP1 occupied promoters into three groups and asked whether motif type was predictive of a change in gene expression. The groups of genes considered were those linked with: (i) promoters containing a top 15% match to the WWM-Q with IRF overlap ($n = 86$), (ii) promoters containing a top 15% match to the WWM-Q without IRF overlap ($n = 76$), (iii) all other genes linked to occupied promoters not included in groups (i) or (ii) ($n = 160$). Gene expression changes on BLIMP1 knockdown were separated into increased or decreased expression relative to control siRNA-treated samples, and then according to fold-change greater than or equal to cut-offs starting from 1.4-fold. A hypergeometric test was used to assess

**Table 2.** Conservation analysis for BLIMP1 motif matches in occupied promoters

| Percentage of ranked list | Primate PhyloP conservation | | | Vertebrate PhyloP conservation | | |
|---|---|---|---|---|---|---|
| | Top percentage with IRF | Bottom percentage with IRF | FETpVal | Top percentage with IRF | Bottom percentage with IRF | FETpVal |
| 10 | 17.03 | 7.42 | 1.01*E*-04 | 15.11 | 9.07 | 1.66*E*-02 |
| 20 | 15.11 | 8.10 | 3.85*E*-05 | 15.52 | 9.20 | 3.20*E*-04 |
| 30 | 13.92 | 8.61 | 1.08*E*-04 | 14.38 | 8.33 | 1.05*E*-05 |
| 40 | 13.32 | 9.00 | 2.54*E*-04 | 12.98 | 9.00 | 7.13*E*-04 |

The table shows results for PhyloP conservation and is divided by primate or vertebrate score. Rows are divided by incremental conservation fractions, while columns show the percentage of BLIMP1 motif matches with IRF overlap in the corresponding top (most) or bottom (least) conserved fractions. The significance of enrichment was calculated using Fischer's exact test.

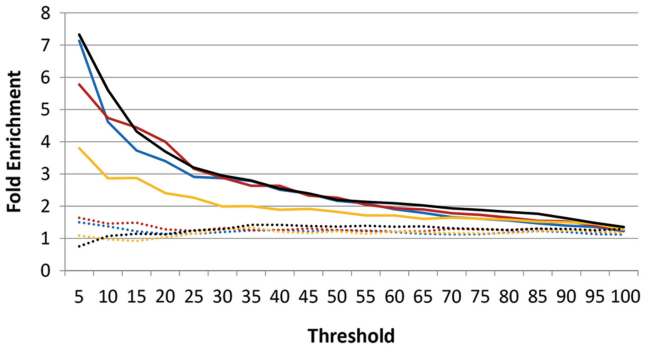


**Figure 6.** *In silico* 'mutagenesis' identifies potential IRF motifs not subject to antagonistic regulation. The graph displays the fold enrichment (*y*-axis) of the WWM-Q (solid lines) and 'mutated' variants (broken lines) across 20 motif thresholds (*x*-axis) compared to $10^6$ matched random promoter regions. Values for original WWM-Qs are as in Table 1: WWM-9 (blue lines), WWM-10 (red lines), WWM-11 (yellow lines) and WWM-13 (black lines).

the probability of finding the observed distribution of target genes associated with each of the BLIMP1 motif types amongst the genes in each expression change category (Figure 7C). Strikingly, target genes linked to group (i) were significantly associated with a positive change in gene expression on BLIMP1 knockdown across all fold-change cut-offs tested. In contrast, genes in groups (ii) or (iii) showed no significant association with a positive change in gene expression. When examining all genes with reduced expression following BLIMP1 knockdown no consistently significant association with any of the groups was seen (data not shown).

To test whether the observed change in gene expression corresponded to the predicted shift in promoter occupancy from BLIMP1 to IRF, ChIP was performed following BLIMP1 knockdown. This demonstrated that loss of BLIMP1 protein expression was consistently linked both to a drop in promoter occupancy by BLIMP1 and to an increase in promoter occupancy by IRFs (Figure 7D). We next confirmed by EMSA that the predicted overlapping binding motifs present in dynamically regulated promoters (*AIM2*, *BTN3A3* and *SP110*) were indeed IRF sites (Figure 8A). Consistent with the motif sequences, IRF1 bound as a single complex to the *AIM2* site (long arrow). In contrast, two complexes of different mobilities were observed for the more extended *BTN3A3* and *SP110* sites (long and short arrows), with the slower mobility complexes consistent with cooperative binding by two IRF1 proteins (27,29). To demonstrate the functional significance of these sites for promoter regulation, luciferase constructs in which the overlapping BLIMP1/IRF sites were retained or mutated were co-transfected with BLIMP1 and/or IRF1 expression vectors (Figure 8B). The wild-type promoters were induced by IRF1 and repressed by BLIMP1. Promoter activation by IRF1 was efficiently repressed by BLIMP1 at low doses (left panels), but with higher amounts of IRF1 repression was partially overcome (middle panels). The presence of identified binding motifs was essential both for BLIMP1-dependent promoter repression, and for IRF-dependent promoter enhancement, since neither factor mediated these effects in promoters in which binding sites were disrupted by mutagenesis (right panels).

Finally, we considered that the presence of an overlapping BLIMP1/IRF motif alone, rather than the presence of such a motif in an occupied promoter could be the important factor in predicting response following BLIMP1 knockdown. We therefore used the same criteria (WWM-Q top15% with IRF overlap) to identify all genes linked to promoters containing BLIMP1/IRF motifs, and repeated the analysis of the global gene-expression data. This group of all potential target genes showed minimal association with a positive change in gene expression (Supplementary Figure S6). Moreover, removing BLIMP1 occupied promoters from this gene set eliminated any residual association indicating that it is unlikely that there is a substantial group of dynamic target genes amongst this gene set that we have not detected. We conclude that it is the combination of BLIMP1 occupancy and motif type rather than motif type per se that predicts a positive change in gene expression on BLIMP1 knockdown in this system. The association between occupancy and promoter motif type with response to BLIMP1 knockdown, the observed shift in transcription factor occupancy after knockdown and the requirement for individual binding sites in reporter assays, therefore confirm the central prediction of the antagonistic regulation model and identify a dynamic set of BLIMP1 target elements linked to the presence of overlapping BLIMP1/IRF sites.

## DISCUSSION

In this study, we set out to address whether unbiased analysis of BLIMP1 occupancy across promoters of the human genome would support the functional relevance of overlap between BLIMP1 and IRF binding sites, and the existence of distinct target gene subsets. Our approach has been to define an extended but stringent set of targets and then use a rigorous analysis of motif representation. The results contribute in a number of ways to our understanding of the biology of BLIMP1.

### Multiple non-redundant matrices describe BLIMP1 specificity

We have used our *in vivo* target set to determine that the previously defined *in vitro* BLIMP1 binding motif provides a robust overall representation of binding specificity *in vivo* (26). However, our analysis of motif representation also indicates that the *in vitro*-defined BLIMP1 matrix does not fully capture binding specificity. *De novo* motif detection identifies a family of motifs, all of which show substantial similarity to the *in vitro* consensus. However, this family of motifs also identifies significant heterogeneity in motif selection. Two aspects of this result are the identification of motifs with a preferential 3′ rather than 5′ extension, and motifs allowing greater degeneracy at key positions. The significance of these findings is verified since these motifs act as genuine BLIMP1

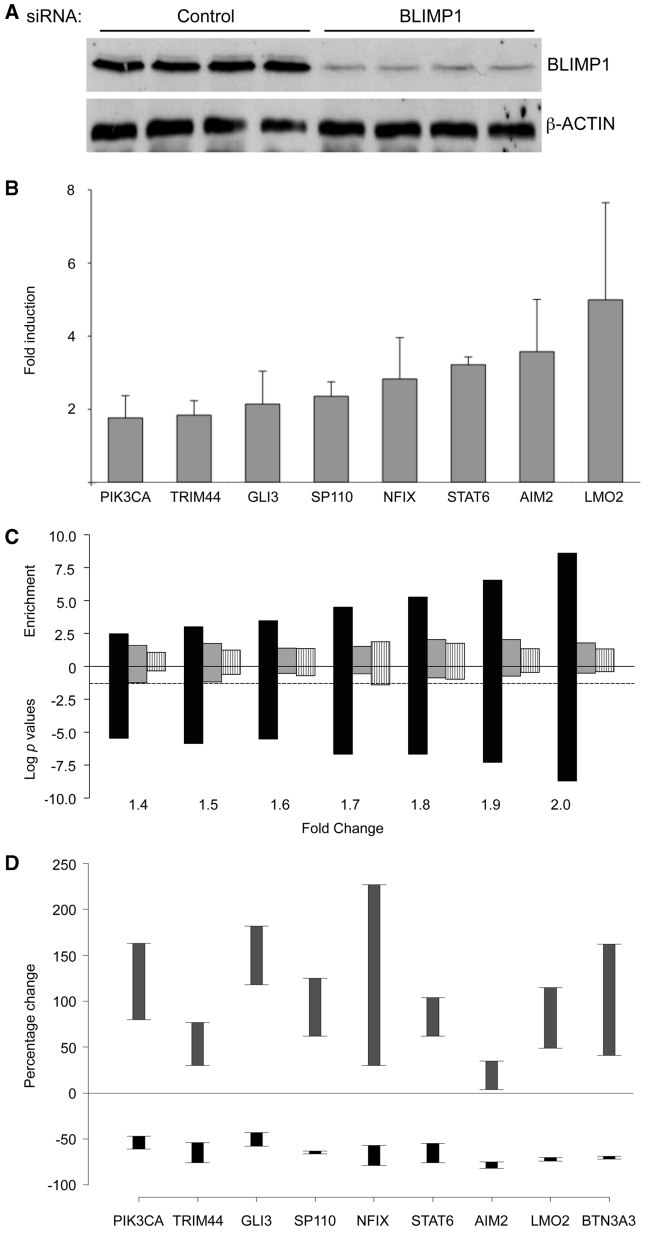


**Figure 7.** Occupancy and motif type define a distinct subset of dynamically regulated BLIMP1 targets. (**A**) siRNA knockdown of BLIMP1 in U266 myeloma cells. Quadruplicate samples of U266 cells were transfected with control siRNA or siRNA directed against *BLIMP1*. The protein levels of BLIMP1 and β-ACTIN present in cell lysates were determined 24 h post-transfection. (**B**) Alteration of BLIMP1 target gene expression in siRNA-treated U266 cells. Changes in expression of individual target genes were evaluated by quantitative RT-PCR. Shown are representative target genes with expression normalized to *GAPDH* and displayed as fold-change (*y*-axis) in expression relative to control siRNA-treated samples. (**C**) Global gene expression changes in siRNA-treated samples were evaluated on NimbleGen micro-arrays. BLIMP1 target genes were divided into three groups according to motif type: (i) those whose promoters contain a top 15% WWM-Q match with IRF overlap (black fill), (ii) those whose promoters contain a top 15% WWM-Q match without IRF overlap (grey fill) (iii) all other occupied promoters (vertical stripes). The enrichment of these target genes amongst all genes showing an increase in expression above fold-change cut-offs from 1.4- to 2-fold was evaluated using a hypergeometric test. The observed enrichment for each group amongst genes changing expression is shown in the upward bars, the significance of this enrichment is shown as the $\log_{10}$ of the *P*-value in downward bars. The dotted line represents $P \leq 0.05$. (D) Dynamically regulated promoters show a shift in occupancy from BLIMP1 to IRFs following BLIMP1 knockdown. U266 cells were transfected with control siRNA or siRNA directed against BLIMP1. ChIP was performed with control antibody and antibodies specific to BLIMP1 or a mixture of antibodies recognizing IRF1 and IRF2. Enrichment of promoter fragments was determined for control and BLIMP1 siRNA transfected cells. The percentage change (y-axis) in enrichment for each promoter (BLIMP1-siRNA ChIP/Control siRNA ChIP) is displayed. Bars show the range of values observed for duplicate ChIP samples (IRF: grey bars, BLIMP1: black bars). Results are representative of three independent experiments.

binding sequences *in vitro*, and promoters containing such sites are occupied *in vivo*. Furthermore, it is of note that in the original random nucleotide selection data for human BLIMP1 (PRDI-BF1), three out of five clones had one or more adenosines 3′ of the central GTGAAAGTG (25). The motif identified by the alternate Oligo-analysis algorithm, also supports the significance of the 3′ extension. However, Oligo-analysis compiles the motifs into an extended symmetrical sequence. We directly tested whether merging shorter matrices into longer symmetrical motifs added to prediction efficiency, but there was no appreciable benefit over using pairs of independent matrices (data not shown).

One question emerging from the prediction of a family of similar motifs was whether these provided redundant information. It was not surprising to find that the matrices with 5′ or 3′ extensions failed to overlap in many cases and represented distinct binding preferences. However, the differences observed between GTG- and GNG-containing matrices (WWM-10 versus WWM-9 and WWM-13 versus WWM-11) demonstrated the degree to which relatively subtle variations in matrix composition substantially impact on motif identification in occupied promoters. Thus, over a range of motif thresholds none of the matrices tested could be viewed as redundant. Choosing a single consensus or PWM to represent BLIMP1 binding specificity would therefore ignore significant information. The primary explanation is likely to be the impact of interdependencies between binding site positions since these are not captured in PWMs, and can be a feature of C2H2 Zn-finger transcription factors (50,51). A number of groups have begun to address the utility of using mixtures of PWM to represent factor binding, and the importance of considering interdependencies for specific transcription factors (52–54). Recently, distinct instances in which *in vitro* binding data require multiple PWMs to represent specificity have been elegantly demonstrated for several murine transcription factors, but BLIMP1 was not amongst the DNA-binding domains tested (55). Our data demonstrate that multiple PWMs are required to efficiently describe BLIMP1 binding specificity, and in this study we have taken a practical approach adopting a quartet of matrices to represent variations in preferences. However, further studies with more advanced representations will be important in future.

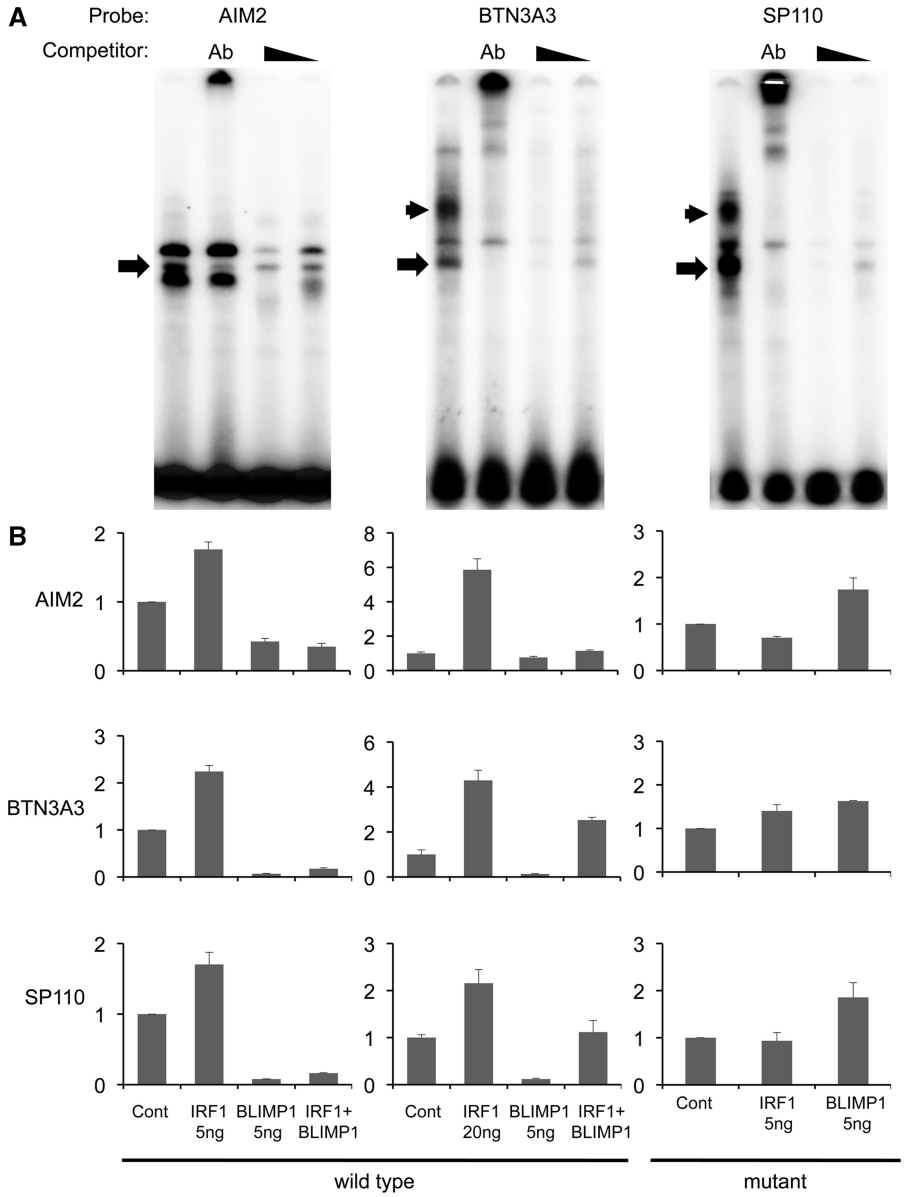


**Figure 8.** Overlapping BLIMP1/IRF binding sites bind IRF1 and mediate IRF1 and BLIMP1-dependent gene regulation. (**A**) EMSA was performed with nuclear extracts from *IRF1* transfected COS cells and oligonucleotides containing an overlapping BLIMP1/IRF site (*AIM2*, *BTN3A3* and *SP110*). IRF1 complexes were identified by antibody supershift (Ab), and specificity of interaction was confirmed by competition with 100- or 10-fold excess of unlabelled probe (wedge). IRF1 bound as a single molecule is indicated by a long arrow, whereas co-occupancy of the site by two molecules of IRF1 is indicated by short arrows. (**B**) HeLa cells were co-transfected with vectors carrying luciferase reporter constructs driven by wild-type promoters of the indicated genes or promoters where IRF/BLIMP1 overlap sites were mutated, together with empty vector (cont), two different amounts of IRF1 (5 ng or 20 ng), and/or BLIMP1 (5 ng) expression vectors. Luciferase activity was assayed 24 h post-transfection. Luciferase activity (*y*-axis) is displayed normalized to renilla and relative to the control transfected sample. Data are derived from triplicate samples and are displayed as the mean ± SD.

### Implications for DNA binding and methylation sensitivity

Methylation of the first three guanosine residues of the BLIMP1 consensus sequence within the positive regulatory domain I (PRDI) of the *IFNβ* promoter abrogates binding (24). These three guanosines are heavily weighted amongst the WWMs, and the importance of two of these positions is underlined by the effect of mutating G to C that we have previously demonstrated (28,29). The recognition of distinct matrix variants around these key contact points has interesting implications with regard to the nature of the BLIMP1 DNA interaction and the contribution of the five BLIMP1 Zn fingers to binding specificity. The contribution of BLIMP1 Zn fingers to DNA binding was tested using LacZ fusion proteins in the context of a concatamer of the *IFNβ* promoter PRDI, representing sequential BLIMP1 sites with 5′ and 3′ extensions (12). Zn fingers 1 and 2 were sufficient to confer specificity to this sequence, but

required additional appropriate N-terminal residues in order to mediate binding. Zn fingers 3–5 were not necessary to mediate recognition of the PRDI concatamer. However, all five Zn fingers together appear to show enhanced recognition in these assays and binding affinities of the various Zn-finger combinations were not directly assessed (12). The orientation of BLIMP1 Zn-fingers relative to the target site is not known, but the existence of motif variants with either 5′ or 3′ extensions suggests that both additional N- and C-terminal sequences may contribute to DNA recognition depending on circumstance. This would imply that C-terminal Zn fingers contribute to recognition of a subset of BLIMP1 target sites, most likely those with 3′ extensions.

The recognition of GNN triplets is a common feature amongst C2H2 Zn-finger domain proteins, and several examples recognizing GTG or GCG triplets have been identified (51). This is a common sequence element recognized at either end of the extended BLIMP1 consensus. The occurrence of a subset of sites with CpG residues suggested that like CTCF, which is another Zn-finger transcription factor with repressive activity, binding of BLIMP1 to some of its targets could be regulated by DNA methylation (46,56). Indeed, we show that the ability of BLIMP1 to bind to its target sequences is eliminated by CpG methylation. In the WWMs, a GCG is better tolerated at the 5′-end of the BLIMP1 binding site. Thus, as in the case of CTCF, specific interactions with one Zn finger may be the target of methylation-dependent regulation (46). BLIMP1 can initiate the process of epigenetic silencing which may eventually lead to DNA methylation (15). This suggests an interesting model in which BLIMP1 is actively excluded from a subset of its target sites, once silencing at the level of DNA methylation is established.

### Motif type defines a dynamic target set

The model of interaction between BLIMP1, IRFs and IFNs originates from the initial identification of BLIMP1 as a repressor of the *IFNβ* promoter where multiple IRFs act to control gene expression (24). IRF1 and IRF2 directly compete with BLIMP1 for regulation of IRF-Es and this is dependent on the precise sequence of the IRF-E, since two point mutations are sufficient to eliminate BLIMP1 binding and regulation (28,29). Such antagonistic regulation provides a molecular mechanism by which the cellular responses driven by BLIMP1 intersect with those driven by the cytokine milieu. Furthermore, other members of the IRF family play key roles in cellular responses to a diverse range of inflammatory signals, and these provide further potential for antagonistic interaction (57,58).

In this study, the unbiased analysis of promoter occupancy confirms the frequent overlap between BLIMP1 and IRF motifs within occupied promoters and demonstrates that this overlap is of general functional significance. Overlapping BLIMP1/IRF binding sites in occupied promoters are significantly enriched amongst conserved promoter elements, suggesting an evolutionary constraint on preservation of dual regulation. Genes whose promoters harbour occupied overlapping BLIMP1/IRF motifs are much more likely to show increases in expression following BLIMP1 knockdown in a cell type co-expressing BLIMP1 and IRFs (28). Moreover, it is the combination of occupancy and motif type rather than motif type per se which governs this response. In an experiment in which expression of a transcriptional repressor is knocked down, high affinity sites are expected to be subject to the most robust repression, and be amongst the least de-repressed of targets. Instead, we observe the opposite result and the most labile genes are associated with high-quality BLIMP1 sites (compare responses of group (i) and group (iii) in Figure 7C). This is not a feature of all genes associated with high-quality BLIMP1 sites but is restricted to the subset that shows overlap with an IRF consensus, and correlates with a shift in promoter occupancy from BLIMP1 to IRF following BLIMP1 knockdown. This response, therefore, confirms a prediction of the antagonistic regulation model, and identifies a set of BLIMP1 targets subject to dynamic repression. Genes included in this set are: transcriptional regulators *SP110, NFIX*, *CREB3*, *ATF3* and *GLI3*, and immune response regulators *AIM2*, *IFIT2*, *PSMB10* and *TAPBP*.

### Target genes and function

Finally, the target list itself provides important insights into novel genes and functions regulated directly by BLIMP1. These are likely to be of relevance beyond the B-cell lineage. Briefly, genes of particular note encode the transcription factors: FOXP1, necessary for early B-cell development and LMO2 a critical regulator of early haematopoietic differentiation both of which are also important prognostic markers in lymphoma (59–62); ATF3 that is involved in cellular stress responses and integration of signalling pathways (63); ATF6 and CREB3 that are transmembrane proteins cleaved to form active transcription factors during the unfolded protein response of the endoplasmic reticulum, a stress response essential to plasma cell differentiation (64); and the glucocorticoid receptor and nuclear transcription factor NR3C1 (65). Other genes of note encode components of signalling pathways such as: AIM2 that has recently emerged as a central regulator of the inflammasome in response to intracellular double-stranded DNA (66–68); PTPN6 (SHP1) a key negative regulator of B-cell signalling (69); and PIK3CA and DGKA that encode central components of lipid signalling pathways (70,71). A common theme shared between several targets is involvement in aspects of cellular stress responses. Given the potential role of IRFs either in the regulatory elements controlling these genes or the response to which they contribute (57,65,72), and the induction of BLIMP1 by diverse forms of cellular stress (24,40,73), we suggest these targets identify an intersection between two aspects of BLIMP1 function: stress-responsive induction and IRF antagonism.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

UK Biotechnology and Biological Sciences Research Council (BB/E000746/1); Leukaemia & Lymphoma Research (07/001); Yorkshire Cancer Research (EK/RC); Cancer Research UK (C78445/A10066 to R.M.T.). Funding for open access charge: Cancer Research UK.

*Conflict of interest statement*. None declared.

## REFERENCES

1. Bikoff,E.K., Morgan,M.A. and Robertson,E.J. (2009) An expanding job description for Blimp-1/PRDM1. *Curr. Opin. Genet. Dev*, **19**, 379–385.
2. Shapiro-Shelef,M., Lin,K.I., McHeyzer-Williams,L.J., Liao,J., McHeyzer-Williams,M.G. and Calame,K. (2003) Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity*, **19**, 607–620.
3. Shaffer,A.L., Lin,K.I., Kuo,T.C., Yu,X., Hurt,E.M., Rosenwald,A., Giltnane,J.M., Yang,L., Zhao,H., Calame,K. *et al.* (2002) Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity*, **17**, 51–62.
4. Kallies,A., Hasbold,J., Tarlinton,D.M., Dietrich,W., Corcoran,L.M., Hodgkin,P.D. and Nutt,S.L. (2004) Plasma cell ontogeny defined by quantitative changes in blimp-1 expression. *J. Exp. Med*, **200**, 967–977.
5. Kallies,A., Hawkins,E.D., Belz,G.T., Metcalf,D., Hommel,M., Corcoran,L.M., Hodgkin,P.D. and Nutt,S.L. (2006) Transcriptional repressor Blimp-1 is essential for T cell homeostasis and self-tolerance. *Nat. Immunol.*, **7**, 466–474.
6. Martins,G.A., Cimmino,L., Shapiro-Shelef,M., Szabolcs,M., Herron,A., Magnusdottir,E. and Calame,K. (2006) Transcriptional repressor Blimp-1 regulates T cell homeostasis and function. *Nat. Immunol.*, **7**, 457–465.
7. Shin,H., Blackburn,S.D., Intlekofer,A.M., Kao,C., Angelosanto,J.M., Reiner,S.L. and Wherry,E.J. (2009) A role for the transcriptional repressor Blimp-1 in CD8(+) T cell exhaustion during chronic viral infection. *Immunity*, **31**, 309–320.
8. Kallies,A., Xin,A., Belz,G.T. and Nutt,S.L. (2009) Blimp-1 transcription factor is required for the differentiation of effector CD8(+) T cells and memory responses. *Immunity.*, **31**, 283–295.
9. Rutishauser,R.L., Martins,G.A., Kalachikov,S., Chandele,A., Parish,I.A., Meffre,E., Jacob,J., Calame,K. and Kaech,S.M. (2009) Transcriptional repressor Blimp-1 promotes CD8(+) T cell terminal differentiation and represses the acquisition of central memory T cell properties. *Immunity.*, **31**, 296–308.
10. Johnston,R.J., Poholek,A.C., Ditoro,D., Yusuf,I., Eto,D., Barnett,B., Dent,A.L., Craft,J. and Crotty,S. (2009) Bcl6 and Blimp-1 are reciprocal and antagonistic regulators of T follicular helper cell differentiation. *Science*, **325**, 1006–1010.
11. Fazilleau,N., McHeyzer-Williams,L.J., Rosen,H. and McHeyzer-Williams,M.G. (2009) The function of follicular helper T cells is regulated by the strength of T cell antigen receptor binding. *Nat. Immunol.*, **10**, 375–384.
12. Keller,A.D. and Maniatis,T. (1992) Only two of the five zinc fingers of the eukaryotic transcriptional repressor PRDI-BF1 are required for sequence-specific DNA binding. *Mol. Cell Biol.*, **12**, 1940–1949.
13. Ren,B., Chee,K.J., Kim,T.H. and Maniatis,T. (1999) PRDI-BF1/Blimp-1 repression is mediated by corepressors of the Groucho family of proteins. *Genes Dev.*, **13**, 125–137.
14. Yu,J., Angelin-Duclos,C., Greenwood,J., Liao,J. and Calame,K. (2000) Transcriptional repression by blimp-1 (PRDI-BF1) involves recruitment of histone deacetylase. *Mol. Cell Biol.*, **20**, 2592–2603.
15. Gyory,I., Wu,J., Fejer,G., Seto,E. and Wright,K.L. (2004) PRDI-BF1 recruits the histone H3 methyltransferase G9a in transcriptional silencing. *Nat. Immunol.*, **5**, 299–308.
16. Ancelin,K., Lange,U.C., Hajkova,P., Schneider,R., Bannister,A.J., Kouzarides,T. and Surani,M.A. (2006) Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nat. Cell Biol.*, **8**, 623–630.
17. Tunyaplin,C., Shapiro,M.A. and Calame,K.L. (2000) Characterization of the B lymphocyte-induced maturation protein-1 (Blimp-1) gene, mRNA isoforms and basal promoter. *Nucleic Acids Res.*, **28**, 4846–4855.
18. Martins,G. and Calame,K. (2008) Regulation and functions of Blimp-1 in T and B lymphocytes. *Annu. Rev. Immunol.*, **26**, 133–169.
19. Lin,K.I., Angelin-Duclos,C., Kuo,T.C. and Calame,K. (2002) Blimp-1-dependent repression of Pax-5 is required for differentiation of B cells to immunoglobulin M-secreting plasma cells. *Mol. Cell Biol.*, **22**, 4771–4780.
20. Lin,Y., Wong,K. and Calame,K. (1997) Repression of c-myc transcription by Blimp-1, an inducer of terminal B cell differentiation. *Science*, **276**, 596–599.
21. Piskurich,J.F., Lin,K.I., Lin,Y., Wang,Y., Ting,J.P. and Calame,K. (2000) BLIMP-I mediates extinction of major histocompatibility class II transactivator expression in plasma cells. *Nat. Immunol.*, **1**, 526–532.
22. Martins,G.A., Cimmino,L., Liao,J., Magnusdottir,E. and Calame,K. (2008) Blimp-1 directly represses Il2 and the Il2 activator Fos, attenuating T cell proliferation and survival. *J. Exp. Med.*, **205**, 1959–1965.
23. Cimmino,L., Martins,G.A., Liao,J., Magnusdottir,E., Grunig,G., Perez,R.K. and Calame,K.L. (2008) Blimp-1 attenuates Th1 differentiation by repression of ifng, tbx21, and bcl6 gene expression. *J. Immunol.*, **181**, 2338–2347.
24. Keller,A.D. and Maniatis,T. (1991) Identification and characterization of a novel repressor of beta-interferon gene expression. *Genes Dev.*, **5**, 868–879.
25. Keller,A.D. and Maniatis,T. (1991) Selection of sequences recognized by a DNA binding protein using a preparative southwestern blot. *Nucleic Acids Res.*, **19**, 4675–4680.
26. Kuo,T.C. and Calame,K.L. (2004) B lymphocyte-induced maturation protein (Blimp)-1, IFN regulatory factor (IRF)-1, and IRF-2 can bind to the same regulatory sites. *J. Immunol.*, **173**, 5556–5563.
27. Fujii,Y., Shimizu,T., Kusumoto,M., Kyogoku,Y., Taniguchi,T. and Hakoshima,T. (1999) Crystal structure of an IRF-DNA complex reveals novel DNA recognition and cooperative binding to a tandem repeat of core sequences. *EMBO J.*, **18**, 5028–5041.
28. Tooze,R.M., Stephenson,S. and Doody,G.M. (2006) Repression of IFN-gamma induction of class II transactivator: a role for PRDM1/Blimp-1 in regulation of cytokine signaling. *J. Immunol.*, **177**, 4584–4593.
29. Doody,G.M., Stephenson,S., McManamy,C. and Tooze,R.M. (2007) PRDM1/BLIMP-1 modulates IFN-gamma-dependent control of the MHC class I antigen-processing and peptide-loading pathway. *J. Immunol.*, **179**, 7614–7623.
30. Tamura,T., Yanai,H., Savitsky,D. and Taniguchi,T. (2008) The IRF family transcription factors in immunity and oncogenesis. *Annu. Rev. Immunol.*, **26**, 535–584.
31. Liu,C.L., Schreiber,S.L. and Bernstein,B.E. (2003) Development and validation of a T7 based linear amplification for genomic DNA. *BMC Genomics*, **4**, 19.
32. Song,J.S., Johnson,W.E., Zhu,X., Zhang,X., Li,W., Manrai,A.K., Liu,J.S., Chen,R. and Liu,X.S. (2007) Model-based analysis of two-color arrays (MA2C). *Genome Biol.*, **8**, R178.
33. Pavesi,G., Mereghetti,P., Mauri,G. and Pesole,G. (2004) Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res.*, **32**, W199–W203.
34. van Helden,J., Andre,B. and Collado-Vides,J. (1998) Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J. Mol. Biol.*, **281**, 827–842.
35. Workman,C.T., Yin,Y., Corcoran,D.L., Ideker,T., Stormo,G.D. and Benos,P.V. (2005) enoLOGOS: a versatile web tool for

energy normalized sequence logos. *Nucleic Acids Res.*, **33**, W389–W392.
36. Mahony,S. and Benos,P.V. (2007) STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic Acids Res.*, **35**, W253–W258.
37. Thomas-Chollier,M., Sand,O., Turatsinze,J.V., Janky,R., Defrance,M., Vervisch,E., Brohee,S. and van Helden,J. (2008) RSAT: regulatory sequence analysis tools. *Nucleic Acids Res.*, **36**, W119–W127.
38. Irizarry,R.A., Bolstad,B.M., Collin,F., Cope,L.M., Hobbs,B. and Speed,T.P. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.*, **31**, e15.
39. Smyth,G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, Article3.
40. Doody,G.M., Stephenson,S. and Tooze,R.M. (2006) BLIMP-1 is a target of cellular stress and downstream of the unfolded protein response. *Eur. J. Immunol.*, **36**, 1572–1582.
41. Acosta-Alvear,D., Zhou,Y., Blais,A., Tsikitis,M., Lents,N.H., Arias,C., Lennon,C.J., Kluger,Y. and Dynlacht,B.D. (2007) XBP1 controls diverse cell type- and condition-specific transcriptional regulatory networks. *Mol. Cell*, **27**, 53–66.
42. Tompa,M., Li,N., Bailey,T.L., Church,G.M., De Moor,B., Eskin,E., Favorov,A.V., Frith,M.C., Fu,Y., Kent,W.J. *et al.* (2005) Assessing computational tools for the discovery of transcription factor binding sites. *Nat. Biotechnol.*, **23**, 137–144.
43. van Helden,J., Rios,A.F. and Collado-Vides,J. (2000) Discovering regulatory elements in non-coding sequences by analysis of spaced dyads. *Nucleic Acids Res.*, **28**, 1808–1818.
44. Tanaka,N., Kawakami,T. and Taniguchi,T. (1993) Recognition DNA sequences of interferon regulatory factor 1 (IRF-1) and IRF-2, regulators of cell growth and the interferon system. *Mol. Cell Biol.*, **13**, 4531–4538.
45. Saxonov,S., Berg,P. and Brutlag,D.L. (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl Acad. Sci. USA*, **103**, 1412–1417.
46. Renda,M., Baglivo,I., Burgess-Beusse,B., Esposito,S., Fattorusso,R., Felsenfeld,G. and Pedone,P.V. (2007) Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *J. Biol. Chem.*, **282**, 33336–33345.
47. Escalante,C.R., Yie,J., Thanos,D. and Aggarwal,A.K. (1998) Structure of IRF-1 with bound DNA reveals determinants of interferon regulation. *Nature*, **391**, 103–106.
48. Hibino,T., Loza-Coll,M., Messier,C., Majeske,A.J., Cohen,A.H., Terwilliger,D.P., Buckley,K.M., Brockton,V., Nair,S.V., Berney,K. *et al.* (2006) The immune gene repertoire encoded in the purple sea urchin genome. *Dev. Biol.*, **300**, 349–365.
49. Livi,C.B. and Davidson,E.H. (2006) Expression and function of blimp1/krox, an alternatively transcribed regulatory gene of the sea urchin endomesoderm network. *Dev. Biol.*, **293**, 513–525.
50. Liu,J. and Stormo,G.D. (2008) Context-dependent DNA recognition code for C2H2 zinc-finger transcription factors. *Bioinformatics*, **24**, 1850–1857.
51. Wolfe,S.A., Nekludova,L. and Pabo,C.O. (2000) DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 183–212.
52. Hannenhalli,S. and Wang,L.S. (2005) Enhanced position weight matrices using mixture models. *Bioinformatics*, **21(Suppl 1)**, i204–i212.
53. Georgi,B. and Schliep,A. (2006) Context-specific independence mixture modeling for positional weight matrices. *Bioinformatics*, **22**, e166–e173.
54. Tomovic,A. and Oakeley,E.J. (2007) Position dependencies in transcription factor binding sites. *Bioinformatics*, **23**, 933–941.
55. Badis,G., Berger,M.F., Philippakis,A.A., Talukder,S., Gehrke,A.R., Jaeger,S.A., Chan,E.T., Metzler,G., Vedenko,A., Chen,X. *et al.* (2009) Diversity and complexity in DNA recognition by transcription factors. *Science*, **324**, 1720–1723.
56. Phillips,J.E. and Corces,V.G. (2009) CTCF: master weaver of the genome. *Cell*, **137**, 1194–1211.
57. Honda,K. and Taniguchi,T. (2006) IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. *Nat. Rev. Immunol.*, **6**, 644–658.
58. Severa,M., Coccia,E.M. and Fitzgerald,K.A. (2006) Toll-like receptor-dependent and -independent viperin gene expression and counter-regulation by PRDI-binding factor-1/BLIMP1. *J. Biol. Chem.*, **281**, 26188–26195.
59. Hu,H., Wang,B., Borde,M., Nardone,J., Maika,S., Allred,L., Tucker,P.W. and Rao,A. (2006) Foxp1 is an essential transcriptional regulator of B cell development. *Nat. Immunol.*, **7**, 819–826.
60. Nam,C.H. and Rabbitts,T.H. (2006) The role of LMO2 in development and in T cell leukemia after chromosomal translocation or retroviral insertion. *Mol. Ther.*, **13**, 15–25.
61. Natkunam,Y., Farinha,P., Hsi,E.D., Hans,C.P., Tibshirani,R., Sehn,L.H., Connors,J.M., Gratzinger,D., Rosado,M., Zhao,S. *et al.* (2008) LMO2 protein expression predicts survival in patients with diffuse large B-cell lymphoma treated with anthracycline-based chemotherapy with and without rituximab. *J. Clin. Oncol.*, **26**, 447–454.
62. Barrans,S.L., Fenton,J.A., Banham,A., Owen,R.G. and Jack,A.S. (2004) Strong expression of FOXP1 identifies a distinct subset of diffuse large B-cell lymphoma (DLBCL) patients with poor outcome. *Blood*, **104**, 2933–2935.
63. Gilchrist,M., Thorsson,V., Li,B., Rust,A.G., Korb,M., Roach,J.C., Kennedy,K., Hai,T., Bolouri,H. and Aderem,A. (2006) Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4. *Nature*, **441**, 173–178.
64. Todd,D.J., Lee,A.H. and Glimcher,L.H. (2008) The endoplasmic reticulum stress response in immunity and autoimmunity. *Nat. Rev. Immunol.*, **8**, 663–674.
65. Glass,C.K. and Ogawa,S. (2006) Combinatorial roles of nuclear receptors in inflammation and immunity. *Nat. Rev. Immunol.*, **6**, 44–55.
66. Burckstummer,T., Baumann,C., Bluml,S., Dixit,E., Durnberger,G., Jahn,H., Planyavsky,M., Bilban,M., Colinge,J., Bennett,K.L. *et al.* (2009) An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome. *Nat. Immunol.*, **10**, 266–272.
67. Fernandes-Alnemri,T., Yu,J.W., Datta,P., Wu,J. and Alnemri,E.S. (2009) AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature*, **458**, 509–513.
68. Hornung,V., Ablasser,A., Charrel-Dennis,M., Bauernfeind,F., Horvath,G., Caffrey,D.R., Latz,E. and Fitzgerald,K.A. (2009) AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature*, **458**, 514–518.
69. Doody,G.M., Justement,L.B., Delibrias,C.C., Matthews,R.J., Lin,J., Thomas,M.L. and Fearon,D.T. (1995) A role in B cell activation for CD22 and the protein tyrosine phosphatase SHP. *Science*, **269**, 242–244.
70. Bader,A.G., Kang,S., Zhao,L. and Vogt,P.K. (2005) Oncogenic PI3K deregulates transcription and translation. *Nat. Rev. Cancer*, **5**, 921–929.
71. Sakane,F., Imai,S., Kai,M., Yasuda,S. and Kanoh,H. (2007) Diacylglycerol kinases: why so many of them? *Biochim. Biophys. Acta*, **1771**, 793–806.
72. Takeshita,F. and Ishii,K.J. (2008) Intracellular DNA sensors in immunity. *Curr. Opin. Immunol.*, **20**, 383–388.
73. Barnes,N.A., Stephenson,S.J., Tooze,R.M. and Doody,G.M. (2009) Amino acid deprivation links BLIMP-1 to the immunomodulatory enzyme indoleamine 2,3-dioxygenase. *J. Immunol.*, **183**, 5768–5777.