3128–3142 *Nucleic Acids Research, 2015, Vol. 43, No. 6* *Published online 09 March 2015*
doi: 10.1093/nar/gkv182

Genome-wide CIITA-binding profile identifies sequence preferences that dictate function versus recruitment

Christopher D. Scharer, Nancy M. Choi, Benjamin G. Barwick, Parimal Majumder, Sarah Lohsen and Jeremy M. Boss*

Department of Microbiology and Immunology, Emory University, Atlanta, GA 30322, USA

Received April 11, 2014; Revised February 20, 2015; Accepted February 24, 2015

ABSTRACT

The class II transactivator (CIITA) is essential for the expression of major histocompatibility complex class II (MHC-II) genes; however, the role of CIITA in gene regulation outside of MHC-II biology is not fully understood. To comprehensively map CIITA-bound loci, ChIP-seq was performed in the human B lymphoblastoma cell line Raji. CIITA bound 480 sites, and was significantly enriched at active promoters and enhancers. The complexity of CIITA transcriptional regulation of target genes was analyzed using a combination of *CIITA*-null cells, including a novel cell line created using CRISPR/Cas9 tools. MHC-II genes and a few novel genes were regulated by CIITA; however, most other genes demonstrated either diminished or no changes in the absence of CIITA. Nearly all CIITA-bound sites were within regions containing accessible chromatin, and CIITA's presence at these sites was associated with increased histone H3K27 acetylation, suggesting that CIITA's role at these non-regulated loci may be to poise the region for subsequent regulation. Computational genome-wide modeling of the CIITA bound XY box motifs provided constraints for sequences associated with CIITA-mediated gene regulation versus binding. These data therefore define the CIITA regulome in B cells and establish sequence specificities that predict activity for an essential regulator of the adaptive immune response.

INTRODUCTION

Major histocompatibility class II (MHC-II) molecules present antigens acquired from the local immune environment to CD4+ T cells and are integral to the generation of a humoral immune response. MHC-II proteins are primarily expressed on the surface of professional antigen presenting cells, such as B cells, macrophages and dendritic cells, and can be induced in other cell types by cytokines such as interferon-γ. Loss of MHC-II expression results in acute immune dysfunction, as in bare lymphocyte syndrome (BLS), and aberrant expression is implicated in autoimmune diseases and cancer (1–3). The class II transactivator (CIITA) is the founding member of the NOD-like receptors (NLR). CIITA and the NLR protein NLRC5 (4) are distinct in their ability to function as transcriptional activators (5–7). CIITA is essential for MHC-II expression. This is highlighted in mouse models (8,9) and in BLS patients who carry null mutations in *CIITA*, and as a result lack MHC-II expression (10,11).

CIITA is recruited to MHC-II promoters through specific interactions with the transcription factors RFX, CREB and NF-Y (12–14). MHC-II proximal promoters contain a highly conserved *cis*-element consisting of W, X1, X2 and Y box elements (15). Cooperative binding between RFX (X1) (14), CREB (X2) (13) and NF-Y (Y) (16) is required for CIITA recruitment. The W box sequence is necessary for optimal MHC-II expression (17), however the exact mechanism by which it functions has yet to be determined (17). Precise spacing between MHC-II W, X and Y elements was shown to be essential for expression of the *HLA-DRA* MHC-II gene (18,19), suggesting that spatial alignment was important for assembly and function of the bound factors.

Outside of the classical MHC-II genes, CIITA also regulates non-classical genes such as *HLA-DM, HLA-DO* and *CD74* (invariant chain), which are associated with antigen processing and selection (20–22) and MHC-I (23). Bioinformatic searches using consensus WXY motifs yielded the presence of one promoter (*RAB4B*) and six intergenic CIITA-bound sites (24,25). Genome-wide ChIP-chip of all human promoters identified eight additional target promoters for CIITA in genes outside of the antigen presentation pathway (26). In dendritic cells *Plxna1*, which controls dendritic cell-T cell interactions, is positively regulated by CIITA (27). Additionally, repressive targets of CIITA have

*To whom correspondence should be addressed. Tel: +1 404 727 5973; Fax: +1 404 727 8250; Email: jmboss@emory.edu
Present address: Nancy M. Choi, Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, CA 92037, USA.

© The Author(s) 2015. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

been identified and include *IL4* (28), Fas ligand (29) and *COL1A2*, where CIITA was reported to act by squelching the activity of the transcriptional coactivator CBP (30). A transcriptional comparison of CIITA-expressing Raji cells and the CIITA-deficient RJ2.2.5 cells using cDNA microarrays identified 43 differentially expressed genes (31), implicating a broader role for CIITA in gene regulation outside of antigen presentation in B cells. In contrast, transgenic over expression and knockout mouse models of CIITA failed to validate previously identified non-MHC-II regulatory targets *in vivo* (32). Thus, while it is clear that MHC-II gene expression is fully dependent on CIITA for expression, the breadth of binding sites, repertoire of sequences that recruit CIITA, and relevance of target sites to transcriptional activity is not fully understood.

To comprehensively determine the sequences that recruit CIITA and assay the role of CIITA in gene regulation inside and outside of the MHC-II, CIITA-binding sites were mapped in Raji human B cells by ChIP-seq using three different antisera. CIITA bound extensively throughout the MHC-II locus and at over 400 sites outside of the MHC-II. To identify chromatin features that were dependent on CIITA, histone H3K27^{ac} and accessible chromatin features were mapped by ChIP-seq and the assay for transposase accessible chromatin (ATAC)-seq (33), respectively, in Raji and *CIITA*-null RJ2.2.5 cells. Subsets of CIITA-bound sites were validated and mRNA profiled at select CIITA-bound genes in both the RJ2.2.5 cell line, as well as a novel CRISPR/Cas9-derived *CIITA* mutant. While the MHC-II locus was dependent on CIITA, most genes outside of the antigen presentation pathway did not demonstrate a dependence on CIITA for transcription. Motif analysis identified a conserved XY box sequence in 99.8% of CIITA-bound sites. However, high variation was observed in both motif quality for X and Y box binding factors and sequence length, distinguishing the X and Y boxes for CIITA-bound sites mapping to genes outside of the antigen presentation pathway. Together, these data suggest a conserved role for CIITA in the antigen presentation pathway and defines sequence characteristics that separate CIITA recruitment from function.

MATERIALS AND METHODS

Cell lines

Raji, a human Burkitt's lymphoma cell line, was purchased from the American Type Tissue Collection (ATCC) (34). RJ2.2.5 cells are a *CIITA*-null cell line that express no surface MHC-II and were derived by γ-irradiation mutagenesis of Raji cells (10,35). Raji and RJ2.2.5 cells were cultured in RPMI containing 5% fetal bovine serum, 5% fetal calf serum, and 100 U/ml penicillin and streptomycin. For nucleofection experiments, 2×10^6 Raji or RJ2.2.5 cells were resuspended in 120 mM $NaPO_4H_2$ (pH. 7.2), 5 mM KCl, and 15 mM $MgCl_2$ and nucleofected using an Amaxa nucleofector, program M-013. The CIITA expression construct was previously described (36). 24 hrs after nucleofection, RJ2.2.5 cells were selected in 400 μg/mL hygromycin for 8 days, stained with anti-HLA-DR-PerCP and anti-HLA-DQ-FITC antibodies (BD Biosciences), and MHC-II positive cells were isolated by FACS. Isolated cells were lysed and total RNA purified for qPCR mRNA analysis.

Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) assays were performed as previously described (37,38). For ChIP-qPCR, 5 μg chromatin was immunoprecipitated with 1 μg of the indicated anti-CIITA antibody overnight at 4°C. Enrichment was determined using a 5-point genomic DNA standard curve and plotted as percentage of input chromatin or percentage of *HLA-DRA* enrichment. Standard error of the mean was used to represent experimental variation from three independent biological experiments. All ChIP primers used in this study are listed in Supplemental Table S2. For ChIP-seq, 30 μg chromatin was immunoprecipitated with 5 μg of anti-CIITA antibody 'B' generated in our lab (39), anti-H3K27^{ac} (Millipore, 07-360), or anti-H3K4^{me3} (Millipore, 07-473) antibody overnight at 4°C. For anti-CIITA antibodies from Rockland Immunochemicals Inc. (100-401-249) and Diagenode (C15410062), 10 μg of antibody was prebound to 30 μl of Protein A beads and incubated overnight at 4°C with 30 μg of chromatin. Antibody–chromatin complexes were captured with Protein A beads (Invitrogen), cross-links reversed and DNA purified. Five percent of each chromatin suspension was removed prior to the immunoprecipitation and used as the Input fraction.

Sequencing library preparation

Ten nanograms of DNA from the antibody B ChIP and input control were end repaired using 2 U T4 DNA polymerase, 0.5 U Klenow fragment of DNA Polymerase, and 5 U T4 DNA polynucleotide kinase for 30 min at 20°C in T4 DNA ligase buffer (NEB) supplemented with 0.4 mM dNTPs. End-repaired DNA was purified and dA-tailed using 5 U Klenow (3′→5′) exo- DNA polymerase for 30 min at 37°C in NEB buffer 2 supplemented with 0.2 mM dATP. dA-tailed DNA was purified and 250 μM annealed adaptors were ligated onto dA-tailed DNA using 2000 U quick ligase for 20 min at room temperature in quick ligase buffer. Adaptor ligated DNA was converted to double stranded DNA by PCR using 2 U of phusion polymerase in GC buffer supplemented with 0.4 mM dNTPs and 0.5 mM adaptor primers. PCR cycles consisted of denaturation at 98°C for 30 s, 5 cycles of 98°C for 10 s, 65°C for 30 s and 72°C for 30 s, followed by a final extension of 72°C for 5 min. PCR products were purified and double-stranded adaptor-ligated DNA was size selected on a 2% agarose gel. A 200–400 bp band was excised and DNA eluted using a Gel Extraction kit (Qiagen, Inc.). Size-selected DNA was PCR-amplified according to the above program for 8 additional cycles to generate sequencing libraries. All enzymes were purchased from New England Biolabs. All purifications were performed using a MiniElute PCR purification kit (Qiagen, Inc.). Sequencing adaptor and primer sequences are listed in Supplemental Table S2. Ten nanograms of H3K4^{me3} ChIP DNA was processed into a sequencing library using the NEXTflex ChIP-seq kit (Bioo Scientific, Cat.# 5143-01) according to the

manufacturer's instructions. For other ChIP-seq libraries, 1 ng of biological duplicate immunoprecipitations from Raji H3K27^{ac}, RJ2.2.5 H3K27^{ac}, CIITA Diagenode (D1 and D2), CIITA Rockland (R1 and R2) ChIP DNA were processed into a sequencing libraries using the KAPA Hyper Prep kit (Cat.# KK8500) according the manufacturers instructions. Sequencing libraries were analyzed for quality on an Agilent Bioanalyzer, quantitated by qPCR using a 6-point standard curve, and sequenced on an Illumina HiSeq2000 instrument.

ATAC-seq

Fifty thousand Raji and RJ2.2.5 cells were fluorescence-activated cell sorting (FACS) sorted into FACS sort buffer (1× phosphate buffered saline (PBS), 1 mM ethylenediaminetetraacetic acid (EDTA,) 1% bovine serum albumin (BSA), 0.25 μM filtered) and processed as described (33) using the Nextera DNA Sample and Indexing kits (Illumina Inc.). Following 10 cycles of PCR amplification, the adaptor ligated DNA was size selected to yield 200–1000 bp fragments. Libraries were quantitated by qPCR, analyzed for quality on an Agilent BioAnalyzer, and sequenced on one lane of an Illumina HiSeq2500 using 50 bp paired end version 4 chemistry.

Data analysis

Sequencing data is available under the accession GSE52941 from the GEO website (http://www.ncbi.nlm.nih.gov/geo/). Raw sequencing reads were mapped to the human genome (hg19) using Bowtie (40). Only uniquely mapped non-redundant reads were used in subsequent analyses. Significantly enriched peaks over input control were determined using HOMER software (41). All CIITA ChIP-seq, histone ChIP-seq and ATAC-seq peaks were identified with the HOMER package using the options '-style factor', '-style histone' and '-style dnase', respectively. A complete list of all peaks can be found with the raw data under the accession number GSE52491 at the GEO database and Supplemental Table S1. Biological duplicate data from Raji and RJ2.2.5 H3K27^{ac} were combined for the analyses presented here. All ENCODE data (42,43) used in this study are listed in Supplemental Table S2. Annotation, manipulation, and analyses of sequencing data were performed using HOMER software (41) and custom R and Perl scripts, which are available upon request. The XY box motif model was built and sequences searched using the MEME software suite (44) and the XY box motif model is provided in Supplemental Table S3. Significance of CIITA-binding site enrichment for GM12878 chromatin states was determined by permutation analysis. The locations of CIITA-binding sites were randomly shuffled 10 000 times and the overlap of annotated chromatin state calculated and compared to the observed overlap. The *P*-value was equal to the number of times the permuted data had an equal or greater overlap or equal or lesser overlap than the actual CIITA-binding sites, divided by the number of permutations, and again divided by 2.

Quantitative real time PCR

RNA was isolated from independent preparations of the indicated cell type using the RNeasy mini prep kit (Qiagen) and quantitative real-time RT-PCR (qRT-PCR) performed as previously described (45). Primers used in this study are listed in Supplemental Table S2. 18s rRNA was measured and used to normalize between samples. For each gene, qRT-PCR was performed at the same time for all cell lines examined, creating a single dataset for each gene. These data were then used as indicated in the generation of Figure 4, Table 1 and Supplemental Figure S1. Normalized data from three independent preparations were averaged and plotted as fold over the Raji cell samples. Student's *t*-tests were used to determine statistical significance between cell types and variation was represented by standard error of the mean (SEM).

XY box conservation analysis

Conservation of the XY box was determined using the 46-way placental mammal phastCons conservation score (46). phastCons scores for human genome build hg19 were downloaded from the UCSC Genome Browser (47) and combined into a bigWig file using Kent tools (48). An alignment of XY boxes was generated by GLAMSCAN software (49) and the phastCons score for each base was parsed from the bigWig file using the rtracklayer package (50) in R/Bioconductor. Results were plotted for each position of the alignment that contained a base in 40% or more of the sequences using custom R/Bioconductor scripts. All custom R/Bioconductor scripts are available upon request.

CRISPR/Cas9 deletion of CIITA exon 3

CRISPR/Cas9 guide sequences were identified and double-stranded guide oligos cloned into the pX330 wild-type Cas9 vector as previously described (51). Guide sequences were chosen to induce a 50 bp deletion spanning CIITA exon 3. Five nanograms each of pX330.CIITAex3.1 and pX330.CIITA.ex3.2 were nucleofected into Raji cells. 7-days post-nucleofection, limiting dilution single-cell cloning of the CRISPR/Cas9 CIITA pool was performed in 96-well plates using 3 dilutions of 3, 1 and 0.3 cells/well. Fourteen days later, wells with positive growth were stained with anti-HLA-DR-PE antibody (BD Biosciences) and screened for loss of HLA-DR surface expression by flow cytometry on an LSRII instrument. One single-cell clone was MHC-II negative (CIITAΔex3). Deletion of CIITA exon 3 was confirmed by PCR amplification across the deleted exon followed by TOPO TA cloning (Life Technologies) and sequencing of three independent colonies. All colonies contained the exact same 50 bp deletion at a position denoted by the PAM sequence. CRISPR/Cas9 guide oligos used for cloning and PCR primers spanning CIITA exon 3 are listed in Supplemental Table S2.

RESULTS

Genome-wide CIITA binding in human B cells

CIITA binds DNA through direct interactions with the transcription factors RFX, CREB1 and NF-Y (19,52), and

Table 1. Gene expression fold change of select genes with CIITA-binding sites

Category	Gene	XY box score[a]	XY box length[a]	Mean fold change from Raji[b]		
				RJ2.2.5	RJ-CIITA	CIITA Δex3
Group I: CIITA-regulated genes	*HLA-DRA*	39.7	48	0.0	0.7	0.0
	HLA-DRB1	35.5	47	0.0	0.8	0.0
	RAB4B	32.1	48	0.6	1.5	0.5
	CD74	31.6	48	0.2	0.5	0.3
	HLA-DOB	20.7	46	0.4	0.7	0.6
	RFX5	14.9	49	0.5	1.2	0.6
	SBK2	12.5	50	0.0	0.3	0.1
	PRR14	6.36	50	0.6	1.2	0.7
	MKNK2	5.86	63	0.5	0.8	0.4
Group II: Genes decreased and not rescued by CIITA	*MYBPC2*	30.2	48	0.2	0.3	0.2
	ZNF672	26.1	51	0.4	0.6	0.7
	ZFR2	19.5	64	0.1	0.2	0.1
	TTLL9	14.2	47	0.0	0.1	0.1
	ZFP82	10.2	88	0.6	0.7	0.8
	B4GALT3	6.52	41	0.4	0.5	0.6
	HLA-E	5.99	49	0.5	0.6	0.6
	PUM1	4.63	44	0.4	0.5	0.7
Group III: Genes unaffected or increased in the absence of CIITA	*MARK3*	19	52	2.6	2.9	3.7
	HAUS5	18.9	53	4.5	4.6	3.8
	PNRC2	12.7	68	1.3	1.4	1.0
	EBF1	12.1	46	0.8	0.8	1.0
	TXNDC15	3.81	43	0.8	0.8	1.0
	KAT7	3.16	45	0.9	1.1	0.9
	PSMB9	0.147	60	1.0	1.0	1.0
	ZNF157	-7.43	46	1.8	1.7	2.0
Group IV: Genes different in RJ2.2.5 and CIITAΔex3	*PARVG*	24.9	50	1.4	0.8	0.9
	HIPK3	13.8	70	0.9	0.8	1.0
	LOC285819	6.96	56	0.3	0.4	1.7
	MXD4	6.72	52	1.5	2.5	1.2
	KDM4A	3.68	58	0.7	1.0	1.4

[a]Determined using the PWM XY box model generated in Figure 6A.
[b]Determined by qRT-PCR and expressed as fold change from Raji. These data are generated from experiments presented in Figure 4 and Supplemental Figure S1.

is essential for the expression of MHC-II genes (12,53). Other genes for which CIITA was shown to bind were either negatively or only partially regulated (24,26). To identify all CIITA-binding sites using an unbiased approach and gain an understanding of how CIITA recruitment and transcriptional activity were related, chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-seq) was performed in the Raji human B cell line. Raji cells were chosen as a model as they can be coupled with the RJ2.2.5 cell line, which are *CIITA*-deficient and derived from Raji cells (10,35). ChIP-seq for CIITA was performed using three independent antisera and peaks classified based on overlap between the data sets. Peaks were categorized into those that overlapped in all five datasets ($n = 20$), three to four datasets ($n = 33$), and two datasets ($n = 427$). All 21 CIITA-binding sites previously located by a ChIP-chip approach in Raji cells were confirmed by this ChIP-seq approach (26). Fifteen of the 20 peaks that overlapped in all datasets were located in the MHC-II locus (Figure 1A and Supplemental Table S1), suggesting that these likely represent the most robust and fully occupied sites. Additionally, 12 sites were located in intergenic regions outside of core promoters at putative MHC-II enhancer elements and include the previously described XL4 and XL7 elements (54) and six sites identified using an *in silico* motif screen (25). The high number of intergenic CIITA-bound sites suggest the possibility that CIITA may function on non-promoter elements elsewhere in the genome.

Trimethylation of histone H3 on lysine 4 ($H3K4^{me3}$) is associated primarily with the transcription start sites (TSSs) of genes and acetylation of histone H3 on lysine 27 ($H3K27^{ac}$) is associated with active TSSs and enhancers (55). Accessibility of DNA to transcription factors is an important criterion in gene regulation and separates active from inactive chromatin. Thus, to characterize the CIITA-bound elements in Raji cells, TSSs and active chromatin were identified in parallel by mapping $H3K4^{me3}$ and $H3K27^{ac,}$ by ChIP-seq and the accessible chromatin land-

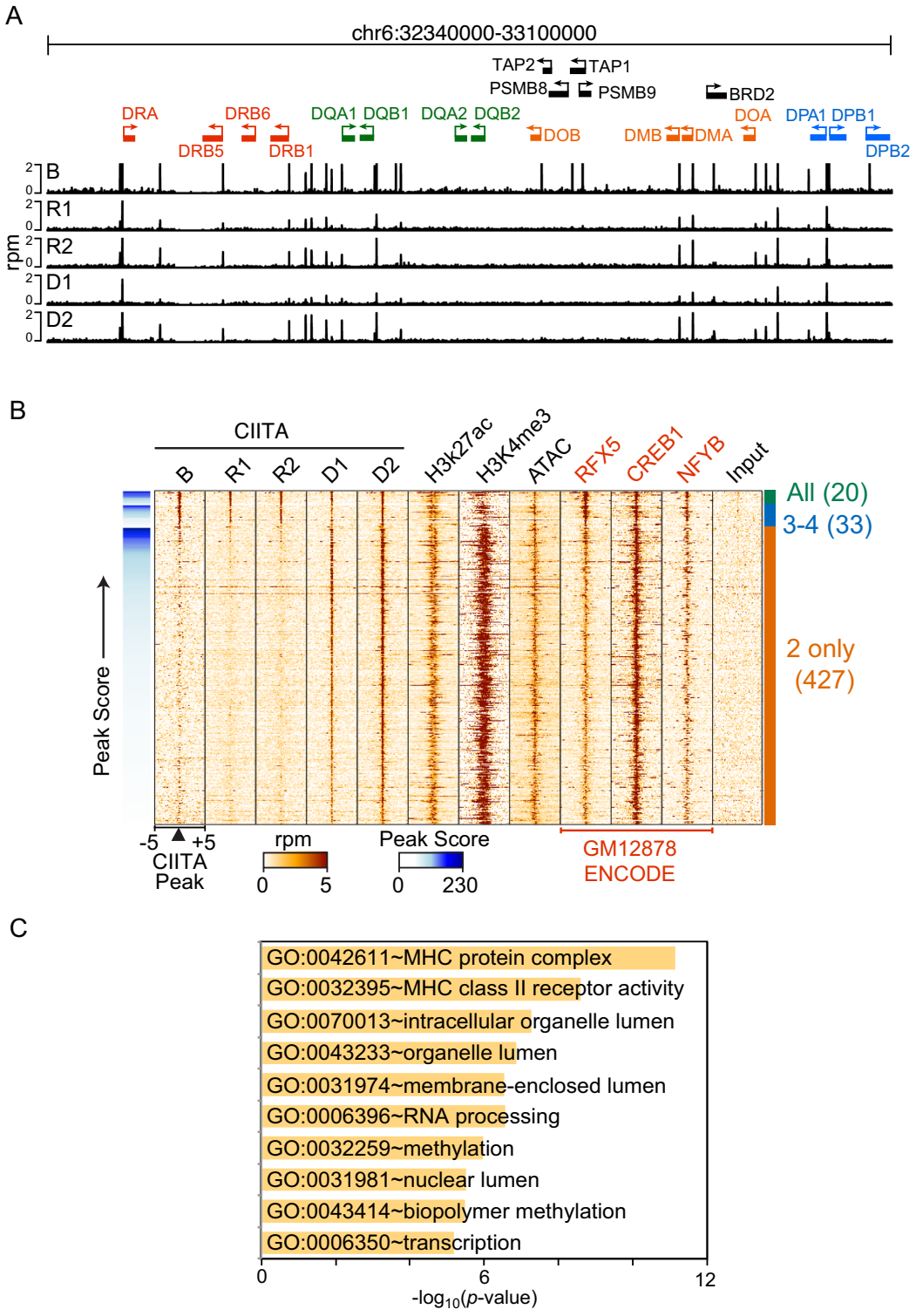


Figure 1. CIITA-binding sites cluster with active histone modifications. (**A**) A schematic of the 24 CIITA-bound sites within the human MHC-II region as defined by the ChIP-seq data set is shown. The classical MHC-II DR (red), DQ (green) and DP (blue) genes are highlighted with regard to the non-classical MHC-II DO and DM (orange) genes. CIITA read depth is plotted as reads per million (rpm). B, Boss Lab generated; R1 and R2, Rockland Immunochemical Inc., replicates 1 and 2; D1 and D2, Diagenode Inc., replicates 1 and 2. (**B**) Heat maps, representing ChIP-seq read depth for indicated factors/modifications in Raji and GM12878 (as indicated) cells by the ENCODE Consortium (42) at the 480 CIITA-binding sites are shown. Each row represents 5 kb surrounding a CIITA peak with read density normalized to rpm. CIITA peaks are sorted into three categories based on overlapping data sets and sorted by decreasing Peak Score (blue bar). (**C**) The top 10 significantly enriched GO ontologies of CIITA-bound genes are plotted. Enrichment is depicted as the $-\log_{10}$ of the *P*-value.

scape determined using ATAC-seq (33). All CIITA-bound loci were enriched in histone H3K27^{ac} and were accessible to transposase activity (ATAC-seq), indicating CIITA binds active and open chromatin (Figure 1B). The majority of peaks that overlapped in two data sets demonstrated enrichment for H3K4^{me3}, however a subset of peaks that were present in three or more data sets were only enriched for H3K27^{ac}, suggesting these sites represent potential enhancer elements. The ENCODE Consortium (42) ChIP-seq datasets were examined for the binding of the XY box specific transcription factors RFX5, a member of the RFX complex, CREB1 and NF-YB, a subunit of the NF-Y complex (Supplemental Table S2). Irrespective of genome location, all CIITA-binding sites were enriched for these three factors, supporting previously known requirements for CIITA recruitment to DNA (Figure 1B).

Gene Ontology (GO) analyses of the closest genes for which CIITA was bound identified numerous processes that may be regulated by CIITA (Figure 1C and Supplemental Table S4). Unsurprisingly, the most significant ontologies were associated with MHC protein complexes and their activity. Other top hits were associated with organelles and membrane lumens, which are also likely associated with genes in antigen processing pathways. An additional ontology indicated that a number of factors that impact gene transcription were regulated by CIITA. This included the previously known *RFX5*, *ZNF672* and *MYBPC2* (26), over 15 zinc finger DNA binding proteins and members of the TFIID initiation complex (*TAF4* and *TAF6*). This data suggests that the CIITA regulome is focused on directly regulating the MHC-II pathway, as well as other transcriptional regulators that may be important for B cell function.

The local chromatin state describes a unique combination of histone modifications and has been used to predict and annotate 15 functional chromatin states (55). The overlap of CIITA-binding sites and the annotated chromatin state data for the B cell line GM12878 were determined to further characterize promoter proximal and distal CIITA-bound loci. Ninety-five percent of all CIITA-binding sites mapped to promoters and enhancers (Figure 2A). This overrepresentation of CIITA-binding sites at promoters and enhancers was more than expected by random chance as determined by a permutation analysis (P-value <0.0002). Significant binding of CIITA at chromatin states indicative of repetitive and copy number variants (15_repetitive/CNV) was observed but represented only 0.8% of all sites. CIITA-binding sites were underrepresented at heterochromatic and repressed regions. These data therefore indicate that CIITA primarily localizes to active promoters and enhancer regions in human B cells.

The *HLA-DRA* gene promoter contains a CIITA-bound WXY box regulatory element within the proximal 100 bp of the TSS and is flanked by H3K4^{me3}-modified nucleosomes (56). To determine if the spatial proximity to the TSS was similar for other genes, a histogram summarizing the CIITA and H3K4^{me3} ChIP-seq signal was plotted for the 4 kb surrounding the TSS for 436 genes containing a promoter proximal CIITA-binding site. At this higher resolution, H3K4^{me3} demonstrated a bimodal signal with maximum enrichment located immediately downstream of the TSS that was preceded by a sharp decrease in enrichment characteristic of a nucleosome-free region (NFR) (57) (Figure 2B, top). In contrast, CIITA was enriched directly upstream of the TSS at the depletion in H3K4^{me3} signal. To determine if CIITA bound a similar NFR at promoter distal sites, a histogram of CIITA and H3K27^{ac} was plotted at the 126 sites annotated to overlap an enhancer. Similar to promoter proximal sites, H3K27^{ac} demonstrated a bimodal enrichment surrounding the maximum enrichment of CIITA (Figure 2B, bottom), indicating that CIITA preferentially binds the NFR at promoter proximal and distal genomic binding sites. Together, these data identify a large number of CIITA-bound regions that were located at enhancer elements distal to gene promoters.

Recently, large (5–20 kb), complex genomic regions associated with high levels of Mediator, histone acetylation and dense transcription factor binding were observed in multiple cell types and termed super enhancers (SE) or stretch enhancers (58,59). Using this annotation, CIITA-binding sites were found to be located in 90 of these SEs (Supplemental Table S1). One of these regions lies between the *HLA-DRB1* and *HLA-DQA1* genes, and including the proximal promoters of these genes contains six CIITA-binding sites (Figure 2C). Each CIITA peak is associated with a local peak of H3K27^{ac} and most show a higher than average level of chromatin accessibility. Another SE was called at the 5′ region of the *RFX5* gene, and is coincident with the CIITA-binding site. H3K27^{ac} surrounds the CIITA peak, which is also the most accessible region of the DNA within the gene. These data suggest that CIITA may contribute to the SE status of key B cell genes.

ChIP-qPCR validation of CIITA-bound loci

Cross-validation of CIITA target sites was performed on a range of sites chosen to encompass both novel and known loci, as well as promoter proximal (*HLA-DRA, RAB4B, CD74, HLA-E, PRR14 and GAPDH*), and promoter distal (*HLA-DQB1, -DQA1, -DPA1, SBK2, ZFR2 and TTLL9*) sites. All three antisera (Boss Lab, Rockland and Diagenode) were used. As expected, *HLA-DRA* showed the highest level of CIITA binding (Figure 3A). However, due to undefined differences in the CIITA-specific titer or affinity, the three antibodies showed differences in their ability to immunoprecipitate CIITA containing chromatin (Figure 3). All of the regions chosen showed significant increases in CIITA binding compared to the control *GAPDH* gene (Figure 3B) and when normalized to the *HLA-DRA* promoter CIITA site, all sites showed similar levels of CIITA enrichment (except *PRR14*) irrespective of antibody used (Figure 3C). ChIP-qPCR using the lab generated CIITA antibody for *HLA-DRA, HLA-E* and *CD74* using RJ2.2.5 cells, a B cell line derived from Raji that contains a compound heterozygous mutation in its *CIITA* genes (35,60), showed no detectable binding of CIITA to these genes (data not shown).

CIITA dependent and independent target genes

CIITA recruits multiple transcriptional coactivator complexes such as CBP/p300, GCN5 and PCAF to positively regulate transcription at MHC-II genes (56,61–63). To measure the influence of CIITA on gene transcription

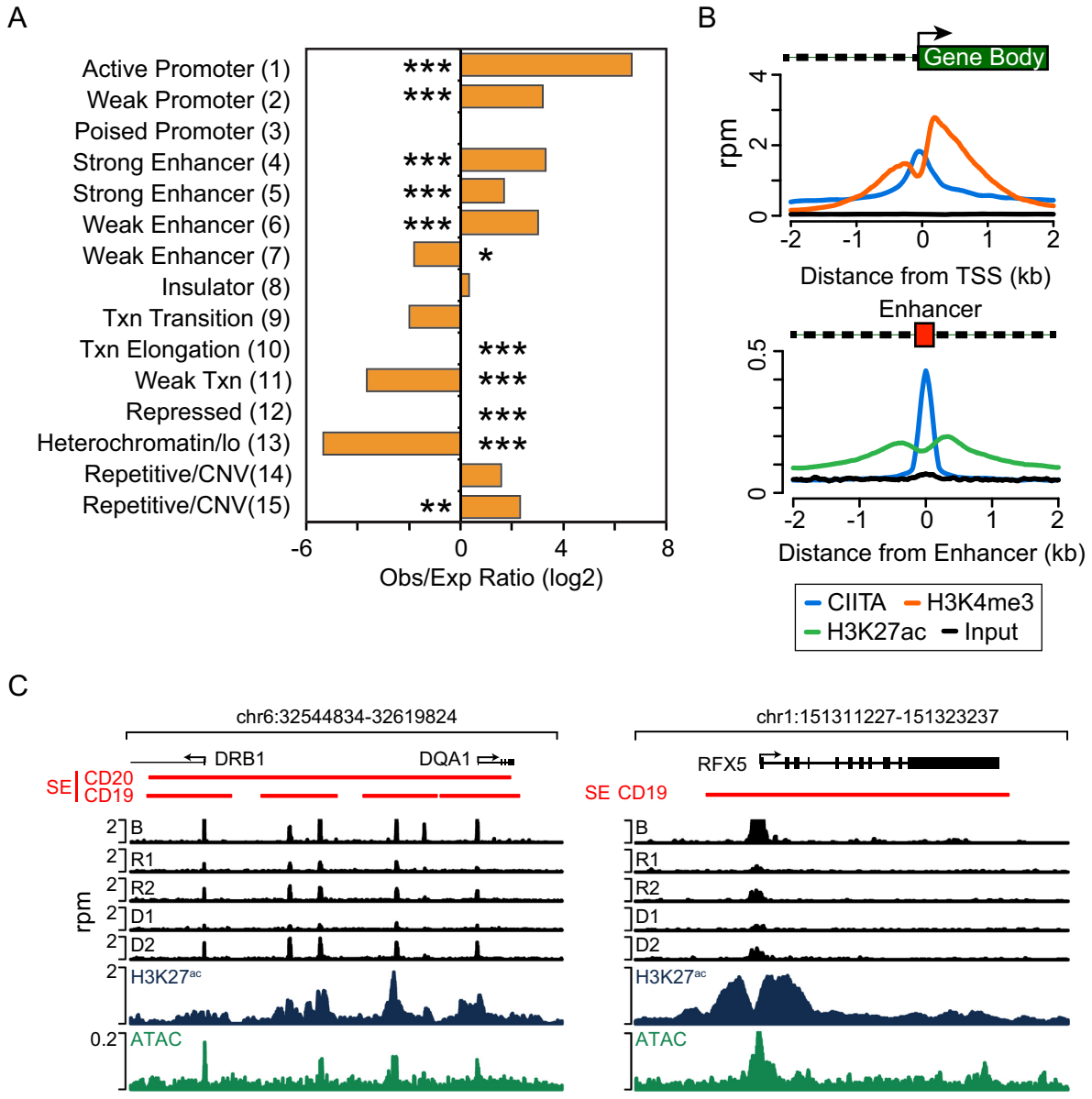


Figure 2. CIITA-binding sites are enriched at active promoters and enhancers in B cells. (**A**) Overlap of CIITA binding with the 15 functional annotated chromatin states (55) in GM12878 B cells. The observed CIITA overlap is plotted as the $\log_2$ of the ratio of observed overlap divided by the expected overlap and compared to the average overlap of 10 000 random genome-wide permutations of 480 regions. Negative ratios indicate depletion and positive ratios indicate enrichment. ***P-value < 0.0002; **P-value $= 0.001$; *P-value $= 0.0036$. (**B**) CIITA binds the nucleosome-free region at promoter proximal and enhancer regions. Histogram depicting CIITA, input and H3K4^{me3} or H3K27^{ac} read density in rpm for 4 kb surrounding either the TSS for all promoter proximal CIITA sites or all CIITA sites that overlap annotated enhancers defined in (A). Location of the TSS, gene body, promoter and enhancer are illustrated above each histogram. (**C**) Schematic of CIITA-bound regions that overlap Super Enhancers (SE) associated with the *HLA-DR*/DQ intergenic region of the MHC-II locus and the *RFX5* gene in human B cells. The enrichment of CIITA, H3K27^{ac} and ATAC-seq accessibility in Raji cells is plotted along with the location of each transcript. Predicted SE from CD19 or CD20 B cells (59) are indicated by red lines below the gene schematic.

at newly identified binding sites, the mRNA levels of target genes were measured in Raji and RJ2.2.5 cells. As expected RJ2.2.5 cells were negative for expression of the MHC-II genes *HLA-DRA* and *HLA-DRB1* (Figure 4). Consistent with previous results, known CIITA target genes *RFX5*, *RAB4B* and *CD74* exhibited a 20–70% reduction in transcript levels in the RJ2.2.5 cells (24,26,64). Transient mRNA levels of five novel CIITA target genes containing sites validated by ChIP-qPCR (Figure 3) were also measured. *HLA-E*, *SBK2, ZFR2* and *TTLL9* demonstrated a statistically significant loss of mRNA in RJ2.2.5 cells. *PRR14* trended lower as well (P-value $= 0.06$). The control gene *ACTB* was unchanged between Raji and RJ2.2.5 cells. To assign the decrease in transcription to CIITA, a cell line derived from RJ2.2.5 that was transfected with an episomal *CIITA*-expression vector and selected for MHC-II expression by FACS was analyzed (36). The *CIITA* complemented cells (RJ-CIITA) demonstrated wild-type expression levels

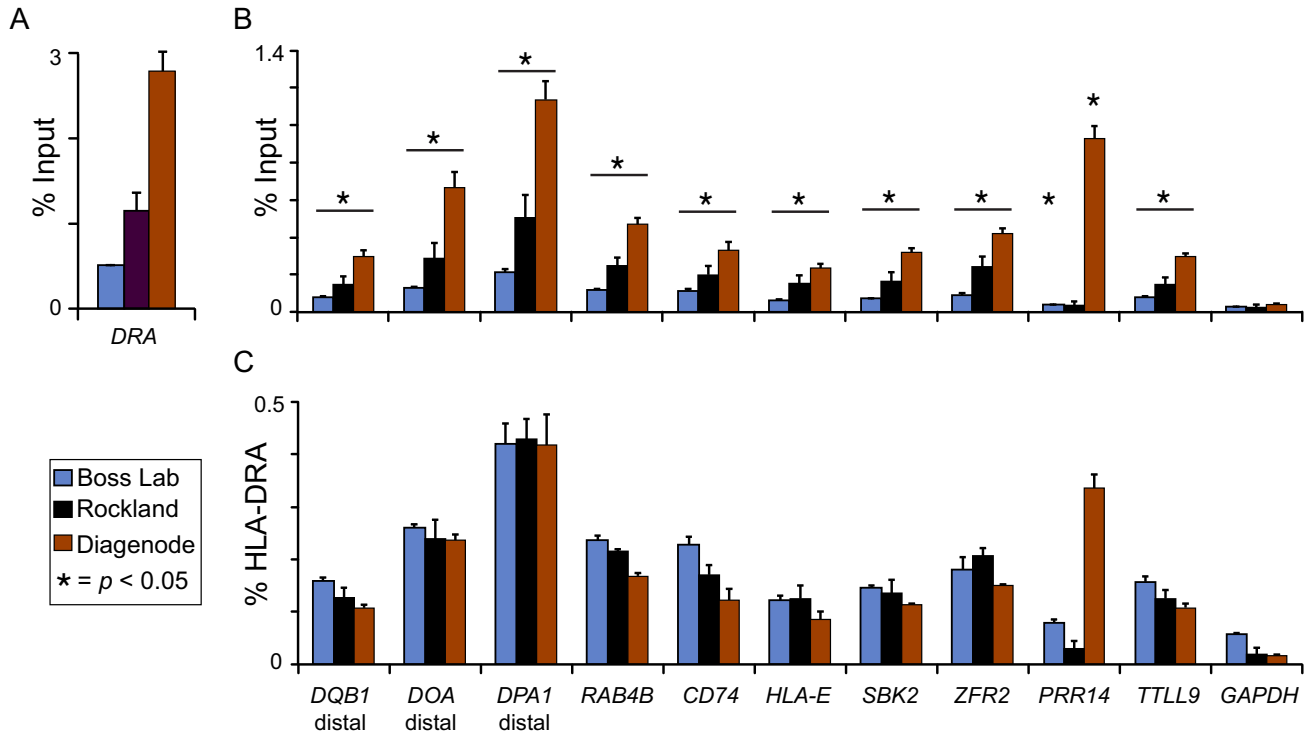


Figure 3. ChIP-qPCR validated CIITA occupancy of ChIP-seq peaks. ChIP-qPCR assays from Raji cells were performed using the Boss Lab generated, Rockland Immunochemicals, Inc., and Diagenode Inc. antisera specific to CIITA to cross validate CIITA-binding sites identified by ChIP-seq. (**A**) Analysis of the *HLA-DRA* promoter CIITA site. (**B**) Select CIITA-binding sites that included promoter proximal and distal were chosen and validated by ChIP-qPCR. (**C**) Data from (B) was plotted as percentage of *HLA-DRA* (A) in order to directly compare results from each antiserum. Data from three independent preparations of chromatin are expressed as percent of the input chromatin. Student's *t* tests were used to calculate significant differences over the *GAPDH* negative control. SEM was used to represent experimental variability. **P*-value <0.05.

of *CIITA* mRNA and restored expression of *HLA-DRA* and *HLA-DRB1* transcripts (Figure 4). mRNA levels of *RFX5*, *CD74*, *SKB2, PRR14* and *RAB4B* were significantly increased in RJ-CIITA cells and suggest that these genes are direct regulatory targets of CIITA. *HLA-E* and *ZFR2* expression trended upward in the complemented cells but this increase was not statistically significant (*P*-value = 0.1).

CRISPR/Cas9 deletion of CIITA in Raji cells

RJ2.2.5 cells were isolated from Raji cells by γ-irradiation mutagenesis (10,35). To control for off target mutations in RJ2.2.5, the CRISPR/Cas9 system (65) was used to ablate CIITA in Raji cells. Targeting oligos were designed to delete exon 3 and induce a frame shift in the *CIITA* transcript (Figure 5A). A clone (CIITAΔex3) with a 96% reduction in HLA-DR surface expression was isolated and characterized (Figure 5B). PCR amplification and sequencing across the targeted region revealed a homozygous deletion spanning exon 3 that precisely matches the intended break points designated by the guide RNA sequences used (Figure 5A and C). To assess the impact on gene expression of CIITA-target genes, RNA was isolated from CIITAΔex3 cells and genes previously analyzed in RJ2.2.5 cells were queried by qRT-PCR. Consistent with RJ2.2.5 cells, *RFX5*, *CD74*, *SBK2, ZFR2, RAB4B, TTLL9* demonstrated a 50–80% reduction in transcript levels in the CIITAΔex3 cells (Figure 4). *HLA-E* and *PRR14* showed lower levels, but these were not statistically significant (*P*-value = 0.08 and 0.09, respectively). These data confirm the results in RJ2.2.5 cells and identify subsets of CIITA target genes that clearly require CIITA for expression (*HLA-DRA*, *HLA-DRB1*) or are augmented by CIITA (*RAB4B*, *RFX5*, *SBK2*). The expression of *ZFR2* and *TTLL9* were robustly reduced in the absence of CIITA, but were not efficiently rescued; and thus, CIITA's role may be unique or indirect for these genes. *HLA-E* and *PRR14* may not be regulated directly by CIITA, despite evidence for its binding to their promoter regions. Thus, CIITA functions directly to regulate some genes and may play indirect roles for others.

CIITA-binding sites contain XY box sequences

The XY box contains binding sites for the DNA binding transcription factors RFX, CREB and NF-Y (13,14,16) and was present at all previously identified CIITA-binding sites. Previously, the nucleotide spacing and spatial alignment of the three elements was found to be critical for MHC-II gene expression (18,19). This suggests that spacing and/or alignment could potentially explain the differences in CIITA function when bound to the various target sites. To determine whether XY box motifs existed in the 480 CIITA-binding sites identified by ChIP-seq, the promoter XY boxes of 11 MHC-II genes were aligned to

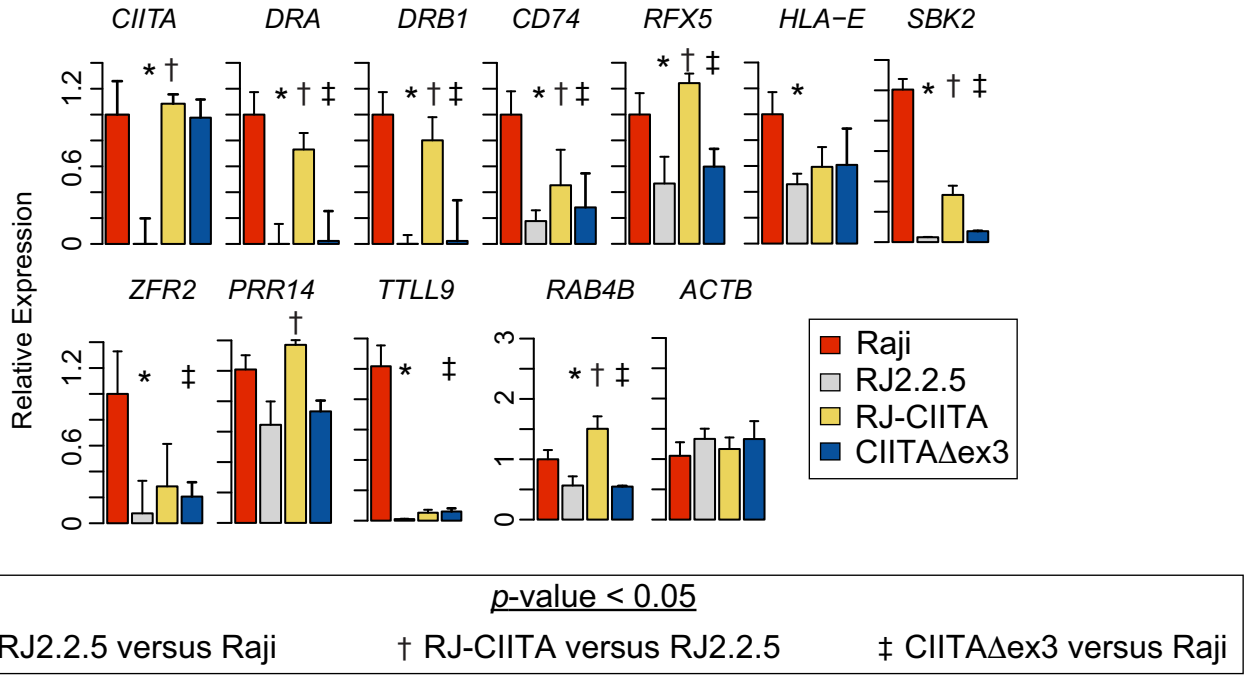


Figure 4. CIITA dependent and independent target genes. Total RNA from three independent isolations of Raji, RJ2.2.5, CIITA complemented RJ2.2.5 cells (RJ-CIITA) and CIITAΔex3 cells (see Figure 5) was extracted and the expression of genes with validated CIITA-binding sites analyzed by qRT-PCR. mRNA levels were normalized to 18s rRNA levels and plotted with respect to the expression in Raji cells. SEM was used to represent experimental variability from three independent experiments. Student's *t* test was used to calculate significant differential expression between cells. *P*-values are indicated.

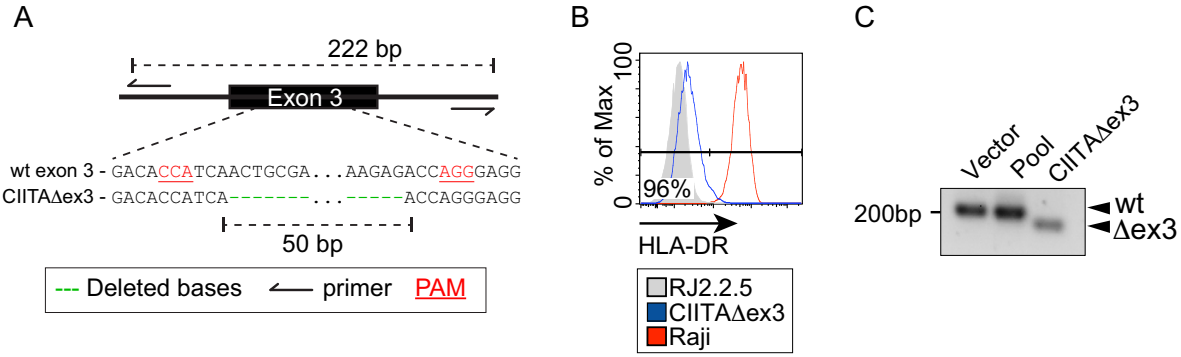


Figure 5. CRISPR/Cas9 Deletion of CIITA. (**A**) Schematic of CIITA exon 3 illustrating the locations of PCR primers used to determine deletion. Enlarged region depicts an alignment of the CRISPR/Cas9-mediated 50 bp deleted sequences in the CIITAΔex3 cells compared to the wild type. Deleted bases are represented by green dashes and the targeted PAM sequence is underlined in red. (**B**) Flow cytometry analysis of HLA-DR surface expression in RJ2.2.5, Raji, and CIITAΔex3 cells. HLA-DR expression is plotted as a histogram and the percentage of CIITAΔex3 HLA-DR negative cells indicated. (**C**) Agarose gel showing PCR amplification across CIITA exon 3 of genomic DNA purified from vector control, CIITA-CRISPR nucleofected (Pool), and CIITAΔex3 cell lines. The size of the wild-type and CRISPR deleted exon 3 are depicted by arrows. The deleted region was not observed in the pool due to the low percentage of cells/alleles that were targeted.

build a preliminary sequence model. GLAMSCAN software (49), which can account for variable spacing between two motifs, was used to search for XY motifs in the CIITA-binding peaks. Newly revealed XY sequences were incorporated into the alignment and the analysis repeated through two additional iterations to build a position-weight matrix (PWM) sequence model that could account for non-MHC-II CIITA-binding sites. A 25 bp X box motif and 14 bp Y box motif, separated by a spacer sequence that varied from 0 to 42 bp were identified in 99.8% (479) of all CIITA-binding sites (Figure 6A). The two X box motifs, X1 and X2, matched consensus-binding sequences for the RFX and CREB family of transcription factors while the Y box motif contained a canonical NF-Y binding motif (66).

As an alternative approach, all 480 CIITA-binding sites were analyzed for the presence of RFX5, CREB1 and NF-YB motifs using HOMER software (41). The occurrence of each motif in a 200 bp window centered at each of the 480 CIITA-binding sites was plotted. All three motifs were oriented in the same direction with the RFX sites occur-

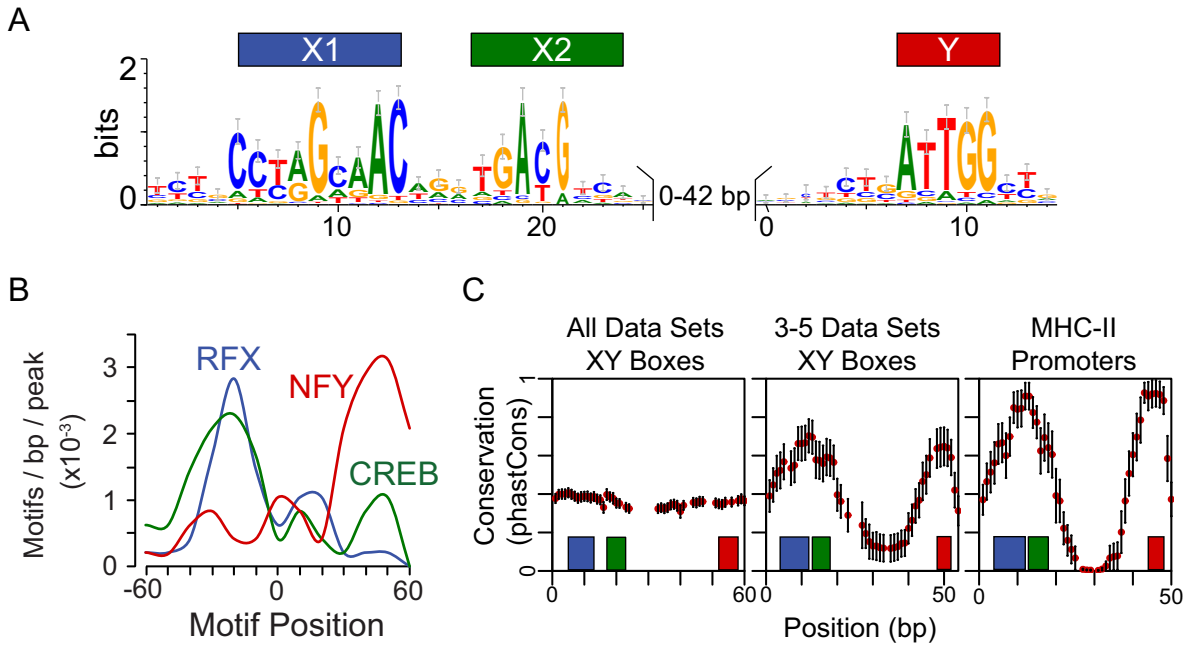


Figure 6. CIITA peaks contain an evolutionarily conserved XY box. (**A**) A position-weight matrix (PWM) that highlighted the sequence variation at 479 CIITA-bound XY boxes was identified using GLAMSCAN software (49). The location of the RFX (X1), CREB (X2) and NF-Y (Y) sites and the location of the variable spacing are indicated. (**B**) RFX5, CREB1 and NF-YB binding motifs were enriched in CIITA peaks. The position of each motif in all CIITA-binding sites was determined using HOMER software (41) and plotted as the average motif per bp per peak. (**C**) 479 CIITA-binding sites were searched using the PWM identified in A and alignments generated for 376 sites (left), 53 sites that overlapped in 3–5 datasets (middle) and the 11 MHC-II promoters (right). The 46-way placental mammal phastCons scores were computed for each base within the three alignments. The average conservation of each position was plotted (red dots) and the standard deviation indicated (black bars). Only those positions present in >40% of sequences contributed to the analysis. Location of X1 (blue), X2 (green) and Y (red) boxes are indicated beneath each plot.

ring upstream and NF-Y downstream (Figure 6B). The CREB motif was found to overlap the RFX site the most, but was also found downstream of it and overlapping the NF-Y motif as well. Importantly, the spatial alignment between the two motif models was consistent (Figure 6A and B), and matched previously identified constraints (18), providing demonstrable evidence that CIITA binding requires RFX, CREB and NF-Y DNA binding proteins, even at non-MHC-II target sites.

Evolutionary conservation of XY box spacing

Regulatory elements often demonstrate evolutionary sequence conservation. To determine if the XY boxes in CIITA-binding sites were conserved across placental mammals, 376 of the 479 XY boxes identified by the PWM model that could be aligned were analyzed for conservation from 46 placental mammals using phastCons scores (46,67) (Figure 6C, left). These sequences showed 40% conservation across the entire length of the XY motif but had significant variability in spacing. Applying this analysis to all of the sites that overlapped in three or more of the CIITA ChIP-seq datasets demonstrated greater than 60% conservation at the RFX, CREB and NF-Y sites (Figure 6C, middle). This alignment showed a higher concordance in motif length. In contrast, alignment of all human MHC-II promoter XY boxes revealed a 52 bp motif with consistent spacing between X and Y box sequences. Sequence conservation of all placental MHC-II promoters was strikingly higher than non-MHC XY box motifs, with the RFX, CREB, and NF-Y core binding sequences displaying 70–90% conservation (Figure 6C, right). Interestingly, reduced conservation of the spacer sequences between the X and Y box was observed, suggesting that selective pressure has retained the spacing between the elements but not the intervening sequence itself.

XY box sequence length and composition distinguishes binding from regulatory function

The PWM motif model identified in CIITA peaks contained a variable spacing of up to 42 bp between X and Y boxes. This large variation contradicts the previously characterized spatial separations required for CIITA to regulate an *HLA-DRA* XY box reporter construct (18,19), and may explain why loss of CIITA affects some target genes but not others. To understand the relationship between spacing and functional CIITA-binding sites, the expression of genes representing nearly all spacer lengths was determined in Raji, RJ2.2.5, RJ-CIITA and CIITAΔex3 cells (Table 1 and Supplemental Figure S1). From this analysis, genes were placed into four groups (Table 1): CIITA-regulated genes (Group I); genes that lost expression in the absence of CIITA but were not rescued by CIITA complementation (Group II); genes unaffected or increased in expression by loss of CIITA (Group III) and genes whose expression were not decreased in both RJ2.2.5 and CIITAΔex3 cells (Group IV). Of the genes regulated by CIITA (Group I), nearly all demon-

strated a periodicity between X and Y box elements equivalent to complete turns of the double helix with spacing ranging between 46 and 50 bp, with one gene, *MKNK2* displaying a 63-bp motif. XY box motif scores as determined using the PWM model showed that all of the XY boxes in this category were high. This indicates that these sites were among the strongest targets within the analysis. Genes decreased but not rescued by CIITA complementation had spacing that ranged from 41 to 88 bp with no clear periodicity pattern. Although some of the genes in this category had high XY box motif scores or periodic spacing, they did not have both. Similarly, genes unaffected or those that showed increases in expression in the presence of CIITA also displayed no clear spacing pattern and had motif lengths ranging from 43 to 68 bp. Even though some of these genes, such as *MARK3, HAUS5*, *PNRC2 and EBF1* had motif scores at the low end of group I genes, spacing between the X and Y boxes was inconsistent with Group I genes, suggesting that these motifs may be out of phase. Lastly, Group IV had genes that were bound by CIITA, but their expression did not decrease in the absence of CIITA expression in the two independent CIITA mutant lines. Thus, this analysis suggests that both spacing and motif score are important for CIITA regulated gene expression.

To examine how motif score and XY box length correlate, all 479 PWM CIITA-binding sites were analyzed (Figure 7A). Human MHC-II genes contained the highest scoring motifs with a total length ranging from 47 to 49 bp. CIITA sites outside of the MHC-II locus identified previously by ChIP-chip (26) also fell within a narrow window consistent with a defined spatial requirement of XY boxes (18). With the exception of *MKNK2*, all of Group I genes had spacing that matched the MHC-II phasing/periodicity previously described for the *HLA-DRA* XY box (18). Thus, the CIITA regulated genes display distinct spacing/phasing that separates this group of genes from those that recruit CIITA.

To examine the influence of CIITA binding on chromatin, ATAC-seq and H3K27^{ac} ChIP-seq were performed on RJ2.2.5 cells and compared to the data collected for Raji cells (Figure 7B). Changes in accessibility at the 479 CIITA-bound peaks were plotted with respect to XY motif score. In the absence of CIITA, small but significant losses in accessibility were observed. Additionally, a robust correlation between the XY motif score and histone H3K27^{ac} between the two cell lines was observed. Thus, CIITA binding is associated with small changes in accessibility and larger changes in histone acetylation. The latter may be due to CIITA's intrinsic acetyltransferase activity (68) or through the recruitment of CBP, p300 and other HAT containing complexes (15).

CIITA binds a limited subset of RFX, CREB and NF-Y sites

Analysis of RFX5, CREB1 and NF-YB ENCODE ChIP-seq datasets revealed that these factors were each bound at more than 20 000 sites across the genome in GM12878 B cells. This raised the possibility that only a fraction of these sites co-localized. To address this issue, the co-occurrence of RFX5, CREB1 and NF-YB, binding sites within a 100 bp window in the GM12878 cell line was computed and determined to occur at 4975 locations (Figure 7C). To further define the relationship between the three factors, each of the 4975 co-occurring sites was examined using the XY box PWM model and the relationship between XY box length and score plotted. The vast majority of sites exhibited negative motif scores, indicating poor matches to the XY model sequence (Figure 7D). XY box motif length demonstrated enrichment between 40 and 50 bp. XY boxes that mapped to MHC-II promoters were annotated and accounted for all of the top 12 sites identified and all motifs scoring over 20. Therefore, despite the presence of thousands of RFX5, CREB1 and NF-YB binding sites in the human genome, only a small fraction contained high quality motifs with the right spacing to interact with CIITA.

DISCUSSION

CIITA was originally discovered as an essential regulator of MHC-II gene expression (10). Outside of the antigen presentation pathway, CIITA has been shown to regulate MHC-I genes in response to interferon-γ (23) and *Plxna1* in dendritic cells (27). To define a global view on CIITA binding, ChIP-seq in the Raji human B cell line was performed with three different antibodies: two raised to the whole CIITA protein (Boss Lab and Diagenode antisera) and the third to a CIITA peptide (Rockland). By combining data in which two or more ChIP-seq experiments detected the same site, 480 total CIITA-bound loci were identified, including all 27 of the previously reported binding sites in B cells (3,25,26). By applying more stringent selection, 53 and 20 of the above loci could be found in three to four or all datasets, respectively. While this work was in progress, an independent CIITA ChIP-seq on human B cells was published (69). In that study, only the Diagenode antiserum was used and 843 CIITA loci were identified, which had significant overlap with the data presented herein. Together, these data demonstrate that there are many more CIITA-binding sites within the genome and implicates a larger role for CIITA in B cell biology than previously anticipated.

The most universal role that CIITA may have is through its ability to recruit factors that modify chromatin. CIITA recruits a host of histone-modifying complexes that catalyze active histone marks at the *HLA-DRA* gene promoter (37,56). The ability of CIITA to do the same at other regions may depend on where it is bound. Nucleosomes surrounding TSS are highly enriched in histone H3K4 methylation and H3K27 acetylation modifications (55) and form a NFR where core transcription initiation factors bind (57,70). At promoter regions, CIITA was bound in the NFR immediately upstream of the TSS, and all of these regions bound were enriched for active histone modifications. CIITA-binding sites also occurred at intergenic loci, including annotated enhancers. CIITA binding to these sites may facilitate binding of other transcription factors that function in that gene's specific transcriptional program.

In the MHC-II locus, 12 intergenic CIITA-binding sites were identified (Figure 1). Evidence that these intergenic sites could function as enhancers was provided by the ATAC-seq dataset and ChIP-seq data demonstrating an enrichment for histone marks that occur at active enhancers, such as H3K27^{ac} (55,71). Four of these CIITA sites were located in the intergenic sequences between *HLA-DRB1* and

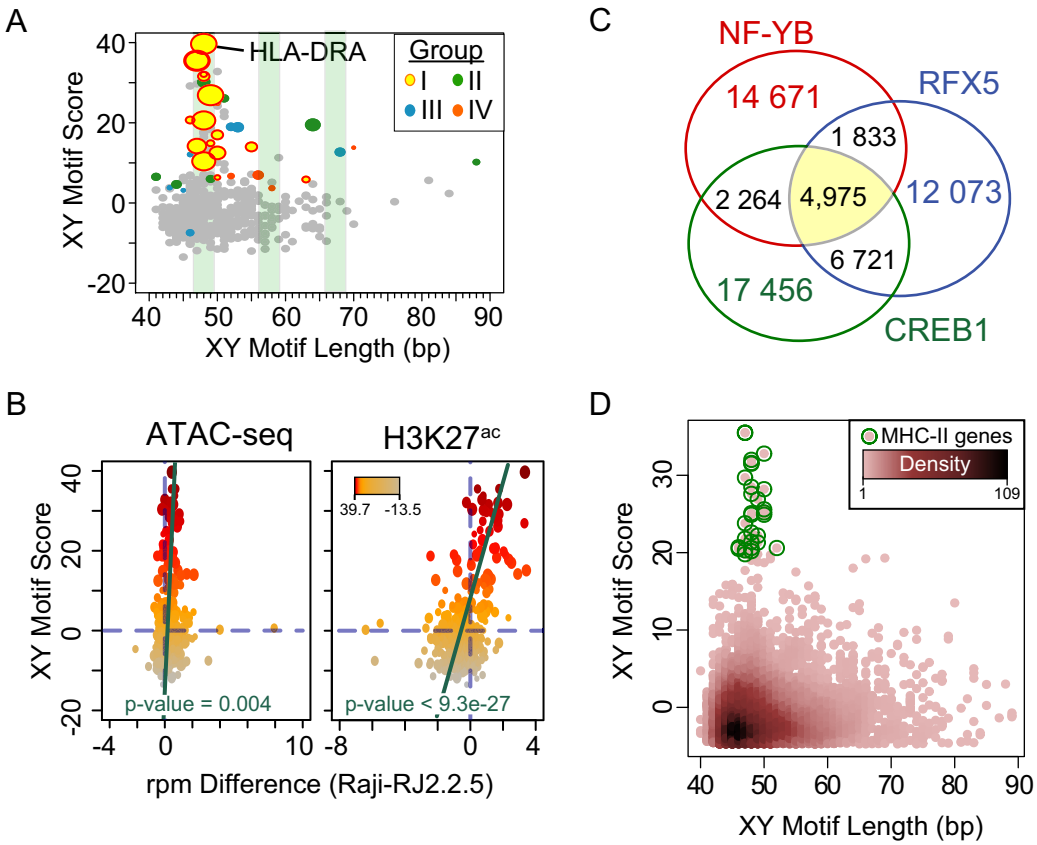


Figure 7. XY box sequence length and motif score distinguish binding from regulatory function. (**A**) The relationship between the XY box motif score and motif length was plotted for 479 sequences in CIITA-binding sites that matched the XY box PWM model generated in Figure 6A. Gene expression data was integrated and color-coded according to the grouping in Table 1. The size of each circle is proportional to the fold change in RJ2.2.5 from Raji. Ten base pair phasing of XY box lengths that correlate with MHC-II promoters are shaded in green. (**B**) Scatter plots of XY box motif score and change in ATAC-seq and H3K27^{ac} reads at 479 CIITA-binding sites. Read differences were determined by subtracting the average rpm in 1 kb windows surrounding each CIITA peak in Raji from RJ2.2.5 cells. The color density of each point corresponds to motif score. The linear regression of the correlation of rpm differences with XY box score is plotted in green with the *P*-value denoting the significance of the correlation. (**C**) Venn diagram representing the number of RFX5, CREB1, and NF-YB bound sites identified by ChIP-seq from GM12878 cells from the ENCODE Consortium (42) that co-occurred in a 100 bp window. (**D**) Density scatter plot of XY box motif score versus motif length for sites identified in (B) that matched the XY box model generated in Figure 6A. Sites that mapped to an MHC-II gene are indicated by green open circles.

HLA-DQA1. This region is of particular interest for two reasons. The first is that it contains an MHC-II insulator element (XL9) that binds the insulator factor CTCF and forms direct interactions with the *HLA-DRB1* and *HLA-DQA1* promoters (72,73). The second is that more recently this entire region was designated as a SE in B cells due to the extremely high levels of active histone marks and the overall size of the domain (58). ATAC-seq and H3K27^{ac} ChIP-seq data presented here demonstrated concordance of accessible and active chromatin state with CIITA binding in this region. How or whether this region regulates MHC-II genes is unknown.

CIITA mutation and complementation experiments produced a range of expression changes for the genes most closely linked to CIITA-binding sites, allowing the placement of the genes into four groups that reflected CIITA dependent expression differences. Although only a select group of genes were analyzed, the finding of genes not regulated by presence/absence of CIITA was unexpected. To understand this, computational tools were used and found significant diversity between both the quality of and the distances between the RFX, CREB and NF-Y binding sites. From these data a PWM of a functional CIITA site was derived. All of the top-scoring sites were present at MHC-II genes. Genes at sites that were not regulated by CIITA failed to have the ideal distance between the elements or exhibited poor RFX, CREB or NF-Y motif scores. It is possible that CIITA binding to a non-ideal site would simply lead to altered histone modifications without further interactions with gene promoters. This could allow CIITA to function in future programming or differentiation responses at these loci without directly affecting the expression of the target gene in the cells analyzed here.

A computational model of the XY box DNA motif was established that facilitated the identification of XY box sequences in 99.8% of CIITA-binding sites. A model that incorporated the W box element failed to account for CIITA binding outside of MHC-II promoters. This is consistent with a previous report that failed to identify a W box (also referred to as S box) motif at all CIITA sites identi-

fied by ChIP-chip (26). Deletion of the W box sequence or spatially altering the distance between W and X box negatively impacts expression of an *HLA-DRA* reporter construct (19), and mutations in the W box decrease CIITA binding (17). Previously, *in silico* genome-wide searches for CIITA-binding sites using a WXY motif aligned from the MHC-II promoters identified *RAB4B* and 6 intergenic enhancer sites (24,25). The identification of W box independent sites for CIITA suggest these sequences are only required for MHC-II gene regulation and may explain why W box inclusive studies did not identify the breadth of CIITA-binding sites described here.

Of the three DNA-binding factors that are required for CIITA recruitment to target loci, the most variation was observed in the CREB-binding site (X2 box). Mapping the location of each factor relative to the peak center revealed sharp enrichment for both the RFX5 and NF-YB motif. In contrast, the CREB1 motif exhibited a broader distribution. Additionally, conservation analysis of either all XY box sequences or only those present in MHC-II promoters demonstrated that the CREB site was the least conserved compared to the RFX and NF-Y sequences. Cooperativity between the XY box binding factors may be a critical factor in the recruitment of CREB to low affinity sites. In the absence of RFX or NF-Y, the ternary (RFX-CREB-NF-Y) complex does not assemble at MHC-II promoters (74–76). RFX alone is able to bind the X box; however, in the presence of CREB, stability is enhanced 12-fold *in vitro* (77,78). These data may explain why the RFX and NF-Y binding sites are more conserved and show less variation than the CREB binding sequence.

The current data support two models of CIITA recruitment to DNA. In the enhanceosome model (79), all of the DNA binding factors (RFX, CREB and NF-Y) and their sites are arranged in an orientation and phase specific pattern with respect to a gene. Such a model allows the DNA to correctly assemble the factors into a scaffold that can recruit non-DNA binding coactivators such as CIITA. This is most likely the case for MHC-II promoters where the conservation across all placental mammals supports not only the orientation and phasing of the XY motifs, but also the spacing between the motifs (Figure 7C). Group I genes may also fall within this model as they are regulated by CIITA and have both high XY motif scores and similar spacing to MHC-II genes. A transcription factor collective model (79), which would allow one or more factor binding motifs to be missing, would allow greater flexibility in motif composition and/or spacing, as seen with the CREB motif when analyzed at all CIITA sites (Figure 6B). The collective model therefore could occur for members of groups II-IV, where recruitment of CIITA would occur but lack the necessary complement of interactions to confer regulatory activity on that gene. In this latter model, the ability of CIITA to influence histone modifications, such as H3K27ac may be an important but not essential event in the regulation of the target gene.

The data presented here reveal the complete CIITA-binding repertoire in human B cells and define sequence recognition requirements that correlate with CIITA activity at bound sites. Owing to the large correlation of non-coding MHC-II locus sequences with human disease (80), the extensive intergenic CIITA-bound regions in the MHC-II locus are particularly interesting and may serve as sites that explain non-coding disease associations. Integrating the role of CIITA with other MHC-II sequence specific regulators, such as insulators (72,73,81) and SEs will be critical for a complete mechanistic understanding of MHC-II regulation and a potential role of these sequences at enhancer regions during B cell differentiation and immune responses.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the members of the Boss lab for their helpful critiques and suggestions, as well as the Southern California Genotyping Consortium and New York University Genomics Core for DNA sequencing, and the Emory Integrated Genomics Core and Emory Flow Cytometry Core.

FUNDING

National Institutes of Health [RO1 GM47310 to J.M.B., F31 AI112261-01 to B.G.B. and T32 support for S.L. and B.G.B. GM008490]. Funding for open access charge: National Institutes of Health [RO1 GM47310 to J.M.B., F31 AI112261-01 to B.G.B. and T32 support for S.L. and B.G.B. GM008490].

Conflict of interest statement. Dr Boss has a licensing agreement with Rockland Immunochemicals for antibodies generated as a collaborative agreement with the company. He receives royalties for those antibodies. None of the antibodies used in this work are part of that licensing agreement.

REFERENCES

1. Kanazawa,S., Ota,S., Sekine,C., Tada,T., Otsuka,T., Okamoto,T., Sonderstrup,G. and Peterlin,B.M. (2006) Aberrant MHC class II expression in mouse joints leads to arthritis with extraarticular manifestations similar to rheumatoid arthritis. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 14465–14470.
2. Reith,W. and Mach,B. (2001) The bare lymphocyte syndrome and the regulation of MHC expression. *Annu. Rev. Immunol.*, **19**, 331–373.
3. Ting,J.P. and Trowsdale,J. (2002) Genetic control of MHC class II expression. *Cell*, **109**(Suppl.), S21–S33.
4. Robbins,G.R., Truax,A.D., Davis,B.K., Zhang,L., Brickey,W.J. and Ting,J.P. (2012) Regulation of class I major histocompatibility complex (MHC) by nucleotide-binding domain, leucine-rich repeat-containing (NLR) proteins. *J. Biol. Chem.*, **287**, 24294–24303.
5. Riley,J.L., Westerheide,S.D., Price,J.A., Brown,J.A. and Boss,J.M. (1995) Activation of class II MHC genes requires both the X box region and the class II transactivator (CIITA). *Immunity*, **2**, 533–543.
6. Zhou,H. and Glimcher,L.H. (1995) Human MHC class II gene transcription directed by the carboxyl terminus of CIITA, one of the defective genes in type II MHC combined immune deficiency. *Immunity*, **2**, 545–553.
7. Harton,J.A., Linhoff,M.W., Zhang,J. and Ting,J.P. (2002) Cutting edge: CATERPILLER: a large family of mammalian genes containing CARD, pyrin, nucleotide-binding, and leucine-rich repeat domains. *J. Immunol.*, **169**, 4088–4093.
8. Williams,G.S., Malin,M., Vremec,D., Chang,C.H., Boyd,R., Benoist,C. and Mathis,D. (1998) Mice lacking the transcription factor CIITA–a second look. *Int. Immunol.*, **10**, 1957–1967.
9. Chang,C.H., Guerder,S., Hong,S.C., van Ewijk,W. and Flavell,R.A. (1996) Mice lacking the MHC class II transactivator (CIITA) show tissue-specific impairment of MHC class II expression. *Immunity*, **4**, 167–178.

10. Steimle,V., Otten,L.A., Zufferey,M. and Mach,B. (1993) Complementation cloning of an MHC class II transactivator mutated in hereditary MHC class II deficiency (or bare lymphocyte syndrome). *Cell*, **75**, 135–146.
11. Benichou,B. and Strominger,J.L. (1991) Class II-antigen-negative patient and mutant B-cell lines represent at least three, and probably four, distinct genetic defects defined by complementation analysis. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 4285–4288.
12. Dorn,A., Durand,B., Marfing,C., Le Meur,M., Benoist,C. and Mathis,D. (1987) Conserved major histocompatibility complex class II boxes–X and Y–are transcriptional control elements and specifically bind nuclear proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 6249–6253.
13. Moreno,C.S., Beresford,G.W., Louis-Plence,P., Morris,A.C. and Boss,J.M. (1999) CREB regulates MHC class II expression in a CIITA-dependent manner. *Immunity*, **10**, 143–151.
14. Reith,W., Satola,S., Sanchez,C.H., Amaldi,I., Lisowska-Grospierre,B., Griscelli,C., Hadam,M.R. and Mach,B. (1988) Congenital immunodeficiency with a regulatory defect in MHC class II gene expression lacks a specific HLA-DR promoter binding protein, RF-X. *Cell*, **53**, 897–906.
15. Choi,N.M., Majumder,P. and Boss,J.M. (2011) Regulation of major histocompatibility complex class II genes. *Curr. Opin. Immunol.*, **23**, 81–87.
16. Dorn,A., Bollekens,J., Staub,A., Benoist,C. and Mathis,D. (1987) A multiplicity of CCAAT box-binding proteins. *Cell*, **50**, 863–872.
17. Muhlethaler-Mottet,A., Krawczyk,M., Masternak,K., Spilianakis,C., Kretsovali,A., Papamatheakis,J. and Reith,W. (2004) The S box of major histocompatibility complex class II promoters is a key determinant for recruitment of the transcriptional co-activator CIITA. *J. Biol. Chem.*, **279**, 40529–40535.
18. Vilen,B.J., Cogswell,J.P. and Ting,J.P. (1991) Stereospecific alignment of the X and Y elements is required for major histocompatibility complex class II DRA promoter function. *Mol. Cell. Biol.*, **11**, 2406–2415.
19. Zhu,X.S., Linhoff,M.W., Li,G., Chin,K.C., Maity,S.N. and Ting,J.P. (2000) Transcriptional scaffold: CIITA interacts with NF-Y, RFX, and CREB to cause stereospecific regulation of the class II major histocompatibility complex promoter. *Mol. Cell. Biol.*, **20**, 6051–6061.
20. Kern,I., Steimle,V., Siegrist,C.A. and Mach,B. (1995) The two novel MHC class II transactivators RFX5 and CIITA both control expression of HLA-DM genes. *Int. Immunol.*, **7**, 1295–1299.
21. Westerheide,S.D., Louis-Plence,P., Ping,D., He,X.F. and Boss,J.M. (1997) HLA-DMA and HLA-DMB gene expression functions through the conserved S-X-Y region. *J. Immunol.*, **158**, 4812–4821.
22. Taxman,D.J., Cressman,D.E. and Ting,J.P. (2000) Identification of class II transcriptional activator-induced genes by representational difference analysis: discoordinate regulation of the DN alpha/DO beta heterodimer. *J. Immunol.*, **165**, 1410–1416.
23. Martin,B.K., Chin,K.C., Olsen,J.C., Skinner,C.A., Dey,A., Ozato,K. and Ting,J.P. (1997) Induction of MHC class I expression by the MHC class II transactivator CIITA. *Immunity*, **6**, 591–600.
24. Krawczyk,M., Leimgruber,E., Seguin-Estevez,Q., Dunand-Sauthier,I., Barras,E. and Reith,W. (2007) Expression of RAB4B, a protein governing endocytic recycling, is co-regulated with MHC class II genes. *Nucleic Acids Res.*, **35**, 595–605.
25. Krawczyk,M., Peyraud,N., Rybtsova,N., Masternak,K., Bucher,P., Barras,E. and Reith,W. (2004) Long distance control of MHC class II expression by multiple distal enhancers regulated by regulatory factor X complex and CIITA. *J. Immunol.*, **173**, 6200–6210.
26. Krawczyk,M., Seguin-Estevez,Q., Leimgruber,E., Sperisen,P., Schmid,C., Bucher,P. and Reith,W. (2008) Identification of CIITA regulated genetic module dedicated for antigen presentation. *PLoS Genet.*, **4**, e1000058.
27. Wong,A.W., Brickey,W.J., Taxman,D.J., van Deventer,H.W., Reed,W., Gao,J.X., Zheng,P., Liu,Y., Li,P., Blum,J.S. *et al.* (2003) CIITA-regulated plexin-A1 affects T-cell-dendritic cell interactions. *Nat. Immunol.*, **4**, 891–898.
28. Zhou,X., Jiang,Y., Lu,L., Ding,Q., Jiao,Z., Zhou,Y., Xin,L. and Chou,K.Y. (2007) MHC class II transactivator represses human IL-4 gene transcription by interruption of promoter binding with CBP/p300, STAT6 and NFAT1 via histone hypoacetylation. *Immunology*, **122**, 476–485.
29. Gourley,T.S. and Chang,C.H. (2001) Cutting edge: the class II transactivator prevents activation-induced cell death by inhibiting Fas ligand gene expression. *J. Immunol.*, **166**, 2917–2921.
30. Xu,Y., Wang,L., Buttice,G., Sengupta,P.K. and Smith,B.D. (2004) Major histocompatibility class II transactivator (CIITA) mediates repression of collagen (COL1A2) transcription by interferon gamma (IFN-gamma). *J. Biol. Chem.*, **279**, 41319–41332.
31. Nagarajan,U.M., Bushey,A. and Boss,J.M. (2002) Modulation of gene expression by the MHC class II transactivator. *J. Immunol.*, **169**, 5078–5088.
32. Otten,L.A., Leibundgut-Landmann,S., Huarte,J., Kos-Braun,I.C., Lavanchy,C., Barras,E., Borisch,B., Steimle,V., Acha-Orbea,H. and Reith,W. (2006) Revisiting the specificity of the MHC class II transactivator CIITA in vivo. *Eur. J. Immunol.*, **36**, 1548–1558.
33. Buenrostro,J.D., Giresi,P.G., Zaba,L.C., Chang,H.Y. and Greenleaf,W.J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*, **10**, 1213–1218.
34. Epstein,M.A. and Barr,Y.M. (1965) Characteristics and mode of growth of tissue culture strain (Eb1) of human lymphoblasts from Burkitt's lymphoma. *J. Natl. Cancer Inst.*, **34**, 231–240.
35. Accolla,R.S. (1983) Human B cell variants immunoselected against a single Ia antigen subset have lost expression of several Ia antigen subsets. *J. Exp. Med.*, **157**, 1053–1058.
36. Majumder,P., Scharer,C.D., Choi,N.M. and Boss,J.M. (2014) B cell differentiation is associated with reprogramming the CCCTC binding factor-dependent chromatin architecture of the murine MHC class II locus. *J. Immunol.*, **192**, 3925–3935.
37. Beresford,G.W. and Boss,J.M. (2001) CIITA coordinates multiple histone acetylation modifications at the HLA-DRA promoter. *Nat. Immunol.*, **2**, 652–657.
38. Scharer,C.D., McCabe,C.D., Ali-Seyed,M., Berger,M.F., Bulyk,M.L. and Moreno,C.S. (2009) Genome-wide promoter analysis of the SOX4 transcriptional network in prostate cancer cells. *Cancer Res.*, **69**, 709–717.
39. Brown,J.A., Rogers,E.M. and Boss,J.M. (1998) The MHC class II transactivator (CIITA) requires conserved leucine charged domains for interactions with the conserved W box promoter element. *Nucleic Acids Res.*, **26**, 4128–4136.
40. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
41. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
42. Consortium,E.P. (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.*, **9**, e1001046.
43. Consortium,E.P., Dunham,I., Kundaje,A., Aldred,S.F., Collins,P.J., Davis,C.A., Doyle,F., Epstein,C.B., Frietze,S., Harrow,J. *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
44. Bailey,T.L., Boden,M., Buske,F.A., Frith,M., Grant,C.E., Clementi,L., Ren,J., Li,W.W. and Noble,W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
45. Yoon,H.S., Scharer,C.D., Majumder,P., Davis,C.W., Butler,R., Zinzow-Kramer,W., Skountzou,I., Koutsonanos,D.G., Ahmed,R. and Boss,J.M. (2012) ZBTB32 is an early repressor of the CIITA and MHC class II gene expression during B cell differentiation to plasma cells. *J. Immunol.*, **189**, 2393–2403.
46. Siepel,A., Bejerano,G., Pedersen,J.S., Hinrichs,A.S., Hou,M., Rosenbloom,K., Clawson,H., Spieth,J., Hillier,L.W., Richards,S. *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, **15**, 1034–1050.
47. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler,D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
48. Kent,W.J., Zweig,A.S., Barber,G., Hinrichs,A.S. and Karolchik,D. (2010) BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*, **26**, 2204–2207.

49. Frith,M.C., Saunders,N.F., Kobe,B. and Bailey,T.L. (2008) Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput. Biol.*, **4**, e1000071.
50. Lawrence,M., Gentleman,R. and Carey,V. (2009) rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics*, **25**, 1841–1842.
51. Ran,F.A., Hsu,P.D., Wright,J., Agarwala,V., Scott,D.A. and Zhang,F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.
52. DeSandro,A.M., Nagarajan,U.M. and Boss,J.M. (2000) Associations and interactions between bare lymphocyte syndrome factors. *Mol. Cell. Biol.*, **20**, 6587–6599.
53. Boss,J.M. and Strominger,J.L. (1986) Regulation of a transfected human class II major histocompatibility complex gene in human fibroblasts. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 9139–9143.
54. Gomez,J.A., Majumder,P., Nagarajan,U.M. and Boss,J.M. (2005) X box-like sequences in the MHC class II region maintain regulatory function. *J. Immunol.*, **175**, 1030–1040.
55. Ernst,J., Kheradpour,P., Mikkelsen,T.S., Shoresh,N., Ward,L.D., Epstein,C.B., Zhang,X., Wang,L., Issner,R., Coyne,M. *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, **473**, 43–49.
56. Choi,N.M. and Boss,J.M. (2012) Multiple histone methyl and acetyltransferase complex components bind the HLA-DRA gene. *PLoS One*, **7**, e37554.
57. Jiang,C. and Pugh,B.F. (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.*, **10**, 161–172.
58. Whyte,W.A., Orlando,D.A., Hnisz,D., Abraham,B.J., Lin,C.Y., Kagey,M.H., Rahl,P.B., Lee,T.I. and Young,R.A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**, 307–319.
59. Hnisz,D., Abraham,B.J., Lee,T.I., Lau,A., Saint-Andre,V., Sigova,A.A., Hoke,H.A. and Young,R.A. (2013) Super-enhancers in the control of cell identity and disease. *Cell*, **155**, 934–947.
60. Brown,J.A., He,X.F., Westerheide,S.D. and Boss,J.M. (1996) Characterization of the expressed CIITA allele in the class II MHC transcriptional mutant RJ2.2.5. *Immunogenetics*, **43**, 88–91.
61. Kretsovali,A., Agalioti,T., Spilianakis,C., Tzortzakaki,E., Merika,M. and Papamatheakis,J. (1998) Involvement of CREB binding protein in expression of major histocompatibility complex class II genes via interaction with the class II transactivator. *Mol. Cell. Biol.*, **18**, 6777–6783.
62. Fontes,J.D., Kanazawa,S., Jean,D. and Peterlin,B.M. (1999) Interactions between the class II transactivator and CREB binding protein increase transcription of major histocompatibility complex class II genes. *Mol. Cell. Biol.*, **19**, 941–947.
63. Spilianakis,C., Papamatheakis,J. and Kretsovali,A. (2000) Acetylation by PCAF enhances CIITA nuclear accumulation and transactivation of major histocompatibility complex class II genes. *Mol. Cell. Biol.*, **20**, 8489–8498.
64. Masternak,K., Muhlethaler-Mottet,A., Villard,J., Zufferey,M., Steimle,V. and Reith,W. (2000) CIITA is a transcriptional coactivator that is recruited to MHC class II promoters by multiple synergistic interactions with an enhanceosome complex. *Genes Dev.*, **14**, 1156–1166.
65. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
66. Mathelier,A., Zhao,X., Zhang,A.W., Parcy,F., Worsley-Hunt,R., Arenillas,D.J., Buchman,S., Chen,C.Y., Chou,A., Ienasescu,H. *et al.* (2014) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **42**, D142–D147.
67. Pollard,K.S., Hubisz,M.J., Rosenbloom,K.R. and Siepel,A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.*, **20**, 110–121.
68. Raval,A., Howcroft,T.K., Weissman,J.D., Kirshner,S., Zhu,X.S., Yokoyama,K., Ting,J. and Singer,D.S. (2001) Transcriptional coactivator, CIITA, is an acetyltransferase that bypasses a promoter requirement for TAF(II)250. *Mol. Cell*, **7**, 105–115.
69. Wong,D., Lee,W., Humburg,P., Makino,S., Lau,E., Naranbhai,V., Fairfax,B.P., Chan,K., Plant,K. and Knight,J.C. (2014) Genomic mapping of the MHC transactivator CIITA using an integrated ChIP-seq and genetical genomics approach. *Genome Biol.*, **15**, 494.
70. Mavrich,T.N., Ioshikhes,I.P., Venters,B.J., Jiang,C., Tomsho,L.P., Qi,J., Schuster,S.C., Albert,I. and Pugh,B.F. (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.*, **18**, 1073–1083.
71. Creyghton,M.P., Cheng,A.W., Welstead,G.G., Kooistra,T., Carey,B.W., Steine,E.J., Hanna,J., Lodato,M.A., Frampton,G.M., Sharp,P.A. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 21931–21936.
72. Majumder,P. and Boss,J.M. (2010) CTCF controls expression and chromatin architecture of the human major histocompatibility complex class II locus. *Mol. Cell. Biol.*, **30**, 4211–4223.
73. Majumder,P., Gomez,J.A., Chadwick,B.P. and Boss,J.M. (2008) The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J. Exp. Med.*, **205**, 785–798.
74. Kara,C.J. and Glimcher,L.H. (1991) In vivo footprinting of MHC class II genes: bare promoters in the bare lymphocyte syndrome. *Science*, **252**, 709–712.
75. Reith,W., Siegrist,C.A., Durand,B., Barras,E. and Mach,B. (1994) Function of major histocompatibility complex class II promoters requires cooperative binding between factors RFX and NF-Y. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 554–558.
76. Wright,K.L., Vilen,B.J., Itoh-Lindstrom,Y., Moore,T.L., Li,G., Criscitiello,M., Cogswell,P., Clarke,J.B. and Ting,J.P. (1994) CCAAT box binding protein NF-Y facilitates in vivo recruitment of upstream DNA binding transcription factors. *EMBO J.*, **13**, 4042–4053.
77. Moreno,C.S., Emery,P., West,J.E., Durand,B., Reith,W., Mach,B. and Boss,J.M. (1995) Purified X2 binding protein (X2BP) cooperatively binds the class II MHC X box region in the presence of purified RFX, the X box factor deficient in the bare lymphocyte syndrome. *J. Immunol.*, **155**, 4313–4321.
78. Louis-Plence,P., Moreno,C.S. and Boss,J.M. (1997) Formation of a regulatory factor X/X2 box-binding protein/nuclear factor-Y multiprotein complex on the conserved regulatory regions of HLA class II genes. *J. Immunol.*, **159**, 3899–3909.
79. Spitz,F. and Furlong,E.E. (2012) Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.*, **13**, 613–626.
80. Hindorff,L.A., Sethupathy,P., Junkins,H.A., Ramos,E.M., Mehta,J.P., Collins,F.S. and Manolio,T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 9362–9367.
81. Majumder,P. and Boss,J.M. (2011) Cohesin regulates MHC class II genes through interactions with MHC class II insulators. *J. Immunol.*, **187**, 4236–4244.