

# Quality of TCR signaling determined by differential affinities of enhancers for the composite BATF–IRF4 transcription factor complex

Arifumi Iwata<sup>1</sup>, Vivek Durai<sup>1</sup>, Roxane Tussiwand<sup>2</sup>, Carlos G Briseño<sup>1</sup>, Xiaodi Wu<sup>1</sup>, Gary E Grajales-Reyes<sup>1</sup>, Takeshi Egawa<sup>1</sup>, Theresa L Murphy<sup>1</sup> & Kenneth M Murphy<sup>1,3</sup>

Variable strengths of signaling via the T cell antigen receptor (TCR) can produce divergent outcomes, but the mechanism of this remains obscure. The abundance of the transcription factor IRF4 increases with TCR signal strength, but how this would induce distinct types of responses is unclear. We compared the expression of genes in the T<sub>H</sub>2 subset of helper T cells to enhancer occupancy by the BATF–IRF4 transcription factor complex at varying strengths of TCR stimulation. Genes dependent on BATF–IRF4 clustered into groups with distinct TCR sensitivities. Enhancers exhibited a spectrum of occupancy by the BATF–IRF4 ternary complex that correlated with the sensitivity of gene expression to TCR signal strength. DNA sequences immediately flanking the previously defined AICE motif controlled the affinity of BATF–IRF4 for direct binding to DNA. Analysis by the chromatin immunoprecipitation–exonuclease (ChIP–exo) method allowed the identification of a previously unknown high-affinity AICE2 motif at a human single-nucleotide polymorphism (SNP) of the gene encoding the immunomodulatory receptor CTLA-4 that was associated with resistance to autoimmunity. Thus, the affinity of different enhancers for the BATF–IRF4 complex might underlie divergent signaling outcomes in response to various strengths of TCR signaling.

The strength of TCR signaling can influence thymocyte fate ‘choice’<sup>1</sup> and the effector outcome of T cells<sup>2</sup>, but how signal strength controls different gene programs has remained unclear<sup>3</sup>. Antigen dose can alter the T<sub>H</sub>1 cell–T<sub>H</sub>2 cell balance<sup>4,5</sup>, T<sub>H</sub>1 cell–follicular helper T cell (T<sub>FH</sub> cell) balance<sup>6</sup> and the production of interleukin 10 (IL-10) by T<sub>H</sub>1 cells<sup>7</sup>. The transcriptional repressor BCL-6 and transcription factor BLIMP-1, which support the development of T<sub>FH</sub> cells or T<sub>H</sub>1 cells, respectively, do show graded abundance at different TCR signal strengths, but this cannot explain all graded T cell responses<sup>2</sup>, and how differences in the strength of TCR signaling regulate the differential abundance of these factors is unknown.

The transcription factor IRF4 might mediate some aspects of variable TCR signaling<sup>3</sup>, including BLIMP-1 abundance<sup>8</sup>. The abundance of IRF4 increases in proportion to TCR signal strength and correlates with T cell population expansion and gene expression for metabolic and biosynthetic pathways<sup>8–10</sup>. IRF4 is required for the effector function of T cells<sup>11</sup> and for the development, class-switch recombination and plasma-cell differentiation of B cells<sup>12</sup>. IRF4 binds the DNA sequence GAAA but requires heterodimerization with other factors for high-affinity binding. In B cells and T cells, IRF4 forms a complex with a heterodimer composed of the transcription factor BATF (‘basic leucine zipper transcription factor, ATF-like’) and transcription factor Jun, which binds DNA at a specific sequence

motif: the AICE (‘activator protein 1 (AP-1)–interferon-regulatory factor (IRF) composite element’)<sup>13–16</sup>. IRF4 is also recruited to EICE motifs (‘E twenty-six (ETS)–IRF composite element’) through interactions with the ETS family members PU.1 and SpiB in B cells and dendritic cells (DCs), but not in T cells, due to the low abundance of ETS transcription factors<sup>11,17</sup>. In plasma cells, which have low expression of BATF, IRF4 has high expression and binds to interferon-sensitive response elements<sup>18</sup>.

The BATF subfamily of AP-1 factors includes BATF, BATF2 and BATF3, which all bind DNA as heterodimers with Jun factors<sup>17</sup>. BATF expression is restricted to the immune system and is required for the differentiation of T<sub>H</sub>9, T<sub>H</sub>17 and T<sub>FH</sub> cells<sup>17,19</sup> and for the differentiation and population expansion of effector CD8<sup>+</sup> T cells<sup>20</sup>. *Batf*<sup>-/-</sup> CD8<sup>+</sup> T cells produce less interferon- $\gamma$  (IFN- $\gamma$ ) than do wild-type cells, which indicates that BATF also regulates activation<sup>20</sup>. BATF is also required for the germinal-center reaction and for class-switch recombination in B cells<sup>21</sup>. BATF3 is expressed in DCs and is required for the development of CD24<sup>+</sup> DCs<sup>22</sup>. BATF and BATF3 are expressed in different cell types but can compensate for each other when expressed in the same cells<sup>13,17</sup>. BATFs enable IRF4- and IRF8-dependent transcription by binding cooperatively to two variants of AICEs: AICE1 (TTTCNNNNTGASTCA, where ‘N’ indicates any nucleotide, and ‘S’ indicates cytosine or guanine)

<sup>1</sup>Department of Pathology and Immunology, Washington University in St. Louis, School of Medicine, St. Louis, Missouri, USA. <sup>2</sup>Department of Biomedicine, University of Basel, Basel, Switzerland. <sup>3</sup>Howard Hughes Medical Institute, Washington University in St. Louis, School of Medicine, St. Louis, Missouri, USA. Correspondence should be addressed to K.M.M. (kmurphy@wustl.edu).

Received 4 August 2016; accepted 23 February 2017; published online 27 March 2017; doi:10.1038/ni.3714

and AICE2 (GAAATGASTCA)<sup>14,15,17</sup>. BATF and IRF4 are both induced within 4 h of TCR stimulation and thus might initiate the expression of many genes encoding products associated with activation and differentiation<sup>23</sup>.

The role of BATF in T<sub>H</sub>2 differentiation has remained unclear due to differing results<sup>13,24,25</sup>. BATF was shown not to be required for the development of T<sub>H</sub>2 cells<sup>13,21,25</sup>, but other studies have reported impaired T<sub>H</sub>2 development in distinct *Batf*<sup>-/-</sup> mice<sup>24</sup>. *Batf3* expression compensates for loss of BATF in T<sub>H</sub>2 development, maintaining the expression of IL-4 and IL-10 but not of CTLA-4, and *Batf1-Batf3*<sup>-/-</sup> (*Batf1-Batf3* DKO) T cells lack expression of IL-4, IL-10 and CTLA-4 (ref. 13). This suggests that genes that are targets of BATFs include some that are sensitive to compensation by endogenous BATF3, but that varying conditions of activation<sup>24</sup> might influence the amount BATF3 or the sensitivity of target genes to compensation by BATF3. In either case, the basis for such differential sensitivity has remained unclear.

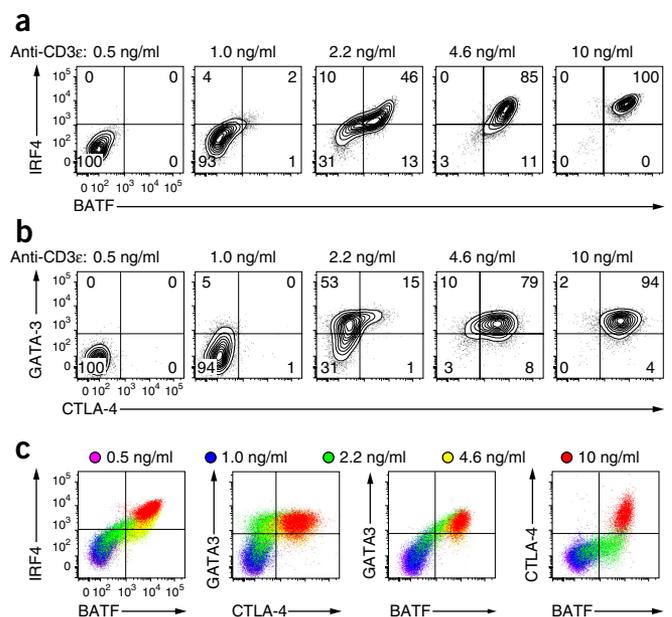
Here, we first documented clearly distinct sensitivities of several genes to compensation by BATF3 in *Batf*<sup>-/-</sup> T<sub>H</sub>2 cells. We found that enhancers that controlled BATF-dependent TCR-inducible genes responded to different levels of BATF-IRF4. For genes that were highly sensitive to low levels of total BATF, endogenous BATF3 was able to compensate for BATF in *Batf*<sup>-/-</sup> T<sub>H</sub>2 cells, while for genes whose expression required higher levels of BATF, it was unable to do so. By ChIP followed by deep sequencing (ChIP-seq) and ChIP-exo analysis, we found that the sensitivity of enhancers in these genes to BATF was regulated by sequences surrounding AICE motifs that influenced affinity for the BATF-IRF4 ternary complex. ChIP-exo analysis helped to identify a previously unknown AICE2 motif that conferred high affinity for BATF-IRF4 that might control a SNP in the *CTLA4* locus known to result in a lower incidence of autoimmune diseases<sup>26,27</sup>.

## RESULTS

### GATA-3 and CTLA-4 respond to distinct signal strengths

We assessed the expression of BATF, IRF4, GATA-3 and CTLA-4 in T<sub>H</sub>2 cells activated with a 'graded' level of TCR signaling. IRF4 was induced in a gradual and uniform manner in proportion to TCR signal strength (Fig. 1a), consistent with published reports<sup>8-10</sup>. BATF expression was also graded and had dose-dependent responses to TCR signaling similar to those of IRF4 (Fig. 1a). *Gata3* and *Ctla4* are known targets of IRF4 and BATF<sup>13,14,28</sup> but GATA-3 and CTLA-4 were induced at different strengths of TCR signaling (Fig. 1b). GATA-3 was induced at low signal strength, while CTLA-4 was induced at higher signal strength (Fig. 1c). During secondary stimulation, similar graded expression in response to TCR stimulation was observed for BATF and IRF4: GATA-3 expression remained high even at low TCR signal strength during secondary stimulation but was more sensitive to TCR signal strength than was CTLA-4 (Supplementary Fig. 1a-c). Likewise, BATF expression was induced in a graded manner in proportion to peptide dose, with induction of GATA-3 occurring at a low peptide dose and induction of CTLA-4 occurring only at a higher peptide dose (Supplementary Fig. 1d-f).

The kinetics of gene expression showed patterns similar to those of the response to TCR dose. BATF and IRF4 were induced on day 1 and accumulated over the next 3 d (Supplementary Fig. 2a). In contrast, GATA-3 was expressed at nearly maximal levels on day 2, but CTLA-4 was not fully expressed until day 3, which correlated with high levels of BATF and IRF4 (Supplementary Fig. 2b-f). Thus, the expression of BATF and IRF4 showed a graded response to both the strength of TCR stimulation and length of activation.



**Figure 1** GATA-3 and CTLA-4 are differentially sensitive to graded expression of BATF and IRF4 in T<sub>H</sub>2 cells following increasing strength of TCR stimulation. (a) Flow cytometry analyzing the expression of IRF4 and BATF on day 4 of primary activation in wild-type CD4<sup>+</sup> T cells cultured under T<sub>H</sub>2 conditions (anti-IFN- $\gamma$ , anti-IL-12 and IL-4) with antibody to the co-receptor CD28 (anti-CD28) and various concentrations (above plots) of anti-CD3 $\epsilon$  crosslinked by plate-bound antibody to hamster immunoglobulin G (IgG). Numbers in quadrants indicate percent cells in each throughout. (b) Flow cytometry analyzing the expression of GATA-3 and CTLA-4 in wild-type CD4<sup>+</sup> T cells on day 4 of culture as in a. (c) Overlay of flow cytometry data from a and b showing expression of IRF4, BATF, GATA-3 and CTLA-4 at various doses (key) of plate bound anti-CD3 $\epsilon$ . Data are representative of two experiments with five biological replicates.

### Differential compensation of BATF targets in T<sub>H</sub>2 cells by *Batf3*

Expression of CTLA-4 in T<sub>H</sub>2 cells was partially dependent on BATF but was completely lacking in *Batf*<sup>-/-</sup>*Batf3*<sup>-/-</sup> (*Batf1-Batf3* DKO) T<sub>H</sub>2 cells, even at a high dose of antibody to the TCR invariant chain CD3 $\epsilon$  (anti-CD3 $\epsilon$ ) in a secondary stimulation, as reported<sup>13</sup> (Fig. 2a,b). In contrast, expression of GATA-3 in T<sub>H</sub>2 cells was BATF independent and was significantly lower in *Batf1-Batf3* DKO T<sub>H</sub>2 cells than in wild type or *Batf*<sup>-/-</sup> T<sub>H</sub>2 cells, similar to the expression of IL-10 (Fig. 2a,b). Thus, endogenous BATF3 was able to compensate for BATF for the expression of GATA-3 and IL-10 but not for that of CTLA-4. This suggested that BATF target genes in T<sub>H</sub>2 cells could have distinct sensitivities to the combined levels of BATF and BATF3.

BATF-dependent gene induction relies on interactions of IRF4 with several amino acid residues of BATF (His55, Lys63 and Glu77) with side chains that face out from the leucine zipper<sup>13</sup>. Substitution of those three BATF amino acids (H55Q, K63D and E77K (BATF-HKE)) eliminated the transcriptional activity of BATF, including IL-17 expression in T<sub>H</sub>17 cells, class-switch recombination in B cells and the development of CD24<sup>+</sup> DCs<sup>13</sup> (Supplementary Fig. 3a-c). We assessed the ability of BATF-HKE to restore the expression of GATA-3, IL-10 and CTLA-4 in *Batf1-Batf3* DKO T<sub>H</sub>2 cells. Retroviral expression of *Batf* or *Batf3* in *Batf1-Batf3* DKO T<sub>H</sub>2 cells fully restored the expression of GATA-3, IL-10 and CTLA-4 to levels similar to those in wild-type T<sub>H</sub>2 cells, as expected<sup>13</sup> (Fig. 2c). In contrast, expression of BATF-HKE in *Batf1-Batf3* DKO T<sub>H</sub>2 cells induced the expression of GATA-3 and IL-10 to intermediate levels and failed to induce CTLA-4 (Fig. 2c).

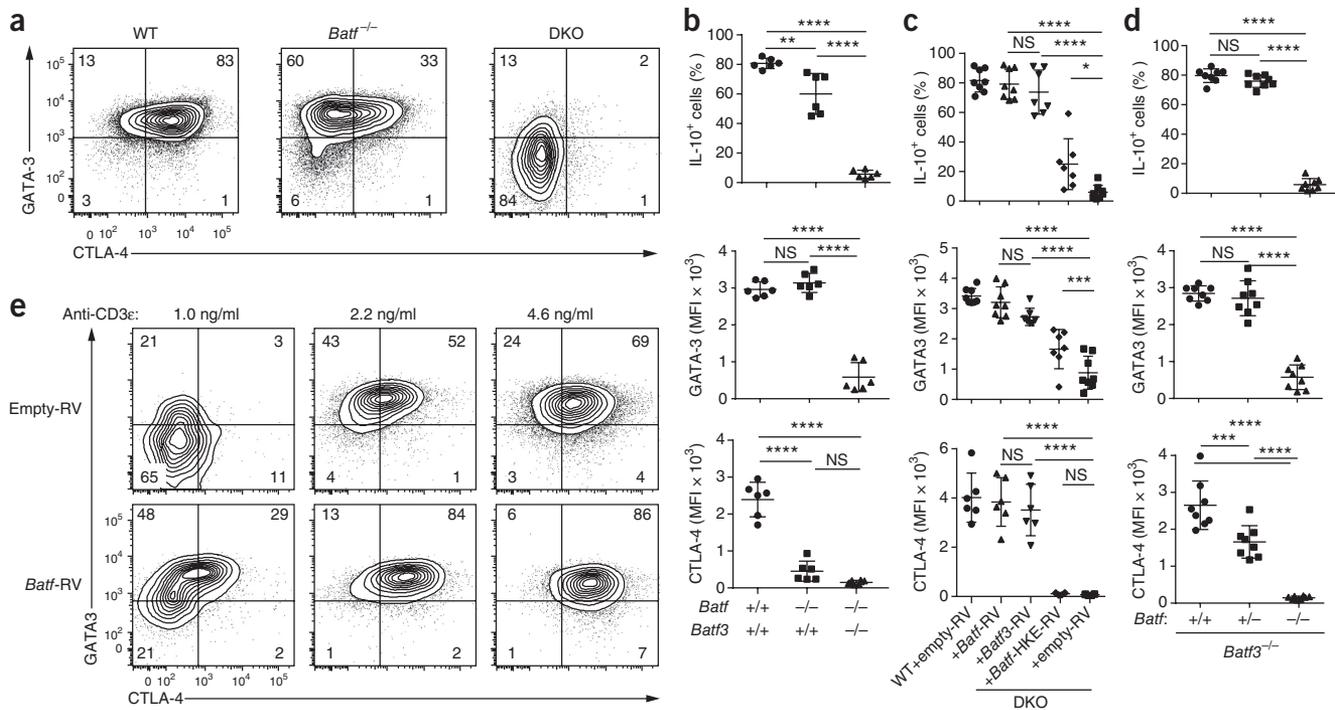
A similar pattern of dependence was observed when BATF levels were assessed across a spectrum in a comparison of cells with wild-type, heterozygous and knockout alleles encoding BATFs. The expression of GATA-3 and IL-10 in *Batf<sup>+/+</sup>Batf3<sup>-/-</sup>* T<sub>H2</sub> cells, which contain one functional *Batf* allele, was equal to that in *Batf<sup>+/+</sup>Batf3<sup>-/-</sup>* T<sub>H2</sub> cells, which contain two *Batf* alleles (Fig. 2d). In contrast, CTLA-4 expression was significantly lower in *Batf<sup>+/+</sup>Batf3<sup>-/-</sup>* T<sub>H2</sub> cells than in *Batf<sup>+/+</sup>Batf3<sup>-/-</sup>* T<sub>H2</sub> cells (Fig. 2d). Finally, retroviral expression of *Batf* in wild-type T<sub>H2</sub> cells shifted the response to TCR dose such that both GATA-3 and CTLA-4 were induced at a lower dose than required for their induction in cells treated with control vector (Fig. 2e). In summary, these results showed that GATA-3 and IL-10 were highly responsive to BATF, requiring only a small amount of total BATF and responding to the weak BATF–IRF4 interaction provided by BATF–HKE. In contrast, CTLA-4 was less sensitive to BATE, requiring a large amount of total BATF and responding only with the strong BATF–IRF4 interaction provided by wild-type BATF but not that provided by BATF–HKE.

### Varying sensitivity of BATF targets to TCR signal strength

We compared global gene expression in wild-type and *Batf1-Batf3* DKO T<sub>H2</sub> cells under weak TCR stimulation (10 ng/ml of

anti-CD3ε on day 2, or 2 ng/ml of anti-CD3ε on day 4) or under strong TCR stimulation (10 ng/ml of anti-CD3ε on day 4). *Batf1-Batf3* DKO T<sub>H2</sub> cells did not express *Ifng*, *Il4* or *Il17a* at any dose of anti-CD3ε (Supplementary Fig. 3d), which indicated that they did not take on a T<sub>H1</sub> or T<sub>H17</sub> phenotype under T<sub>H2</sub> conditions. In the two conditions of weak TCR stimulation, 30 or 68 genes had expression that was at least threefold higher in wild-type cells than in *Batf1-Batf3* DKO cells; with strong TCR stimulation, this increased to 207 genes (Fig. 3a). Spearman's rank-order correlation of gene expression was greatest for samples of similar strength of TCR stimulation, and there was strong positive correlation between the two weak-TCR-stimulation conditions (Fig. 3b). These results indicated that distinct sets of genes were induced by low and high levels of TCR stimulation.

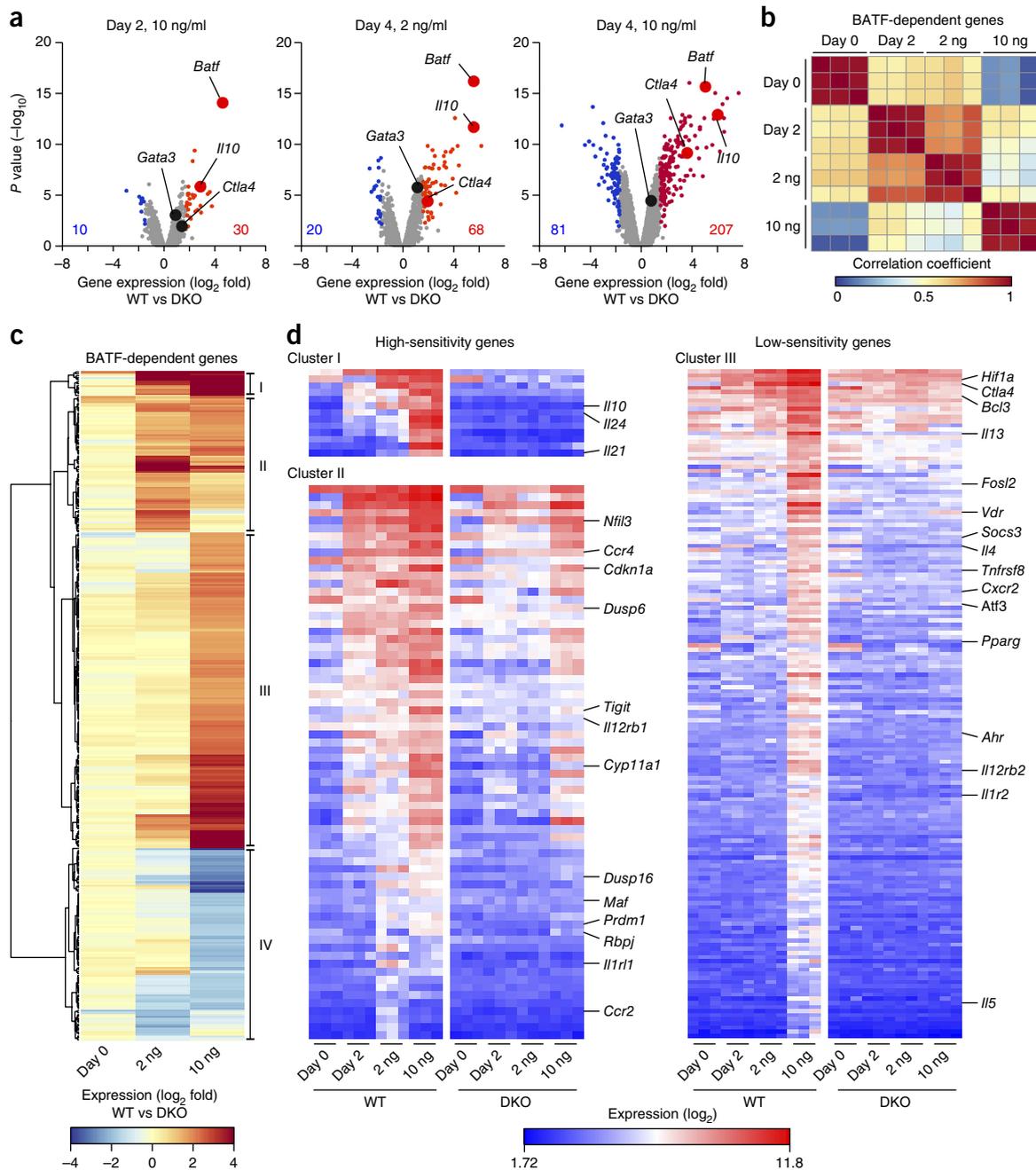
We carried out hierarchical clustering of genes based on the change in expression in wild-type T<sub>H2</sub> cells relative to that in *Batf1-Batf3* DKO T<sub>H2</sub> cells. Clusters I and II contained genes with large changes in expression in wild-type T<sub>H2</sub> cells relative to that in *Batf1-Batf3* DKO T<sub>H2</sub> cells with weak TCR stimulation, while cluster III contained genes that differed in expression only with strong TCR stimulation (Fig. 3c,d). Cluster IV contained genes that had higher



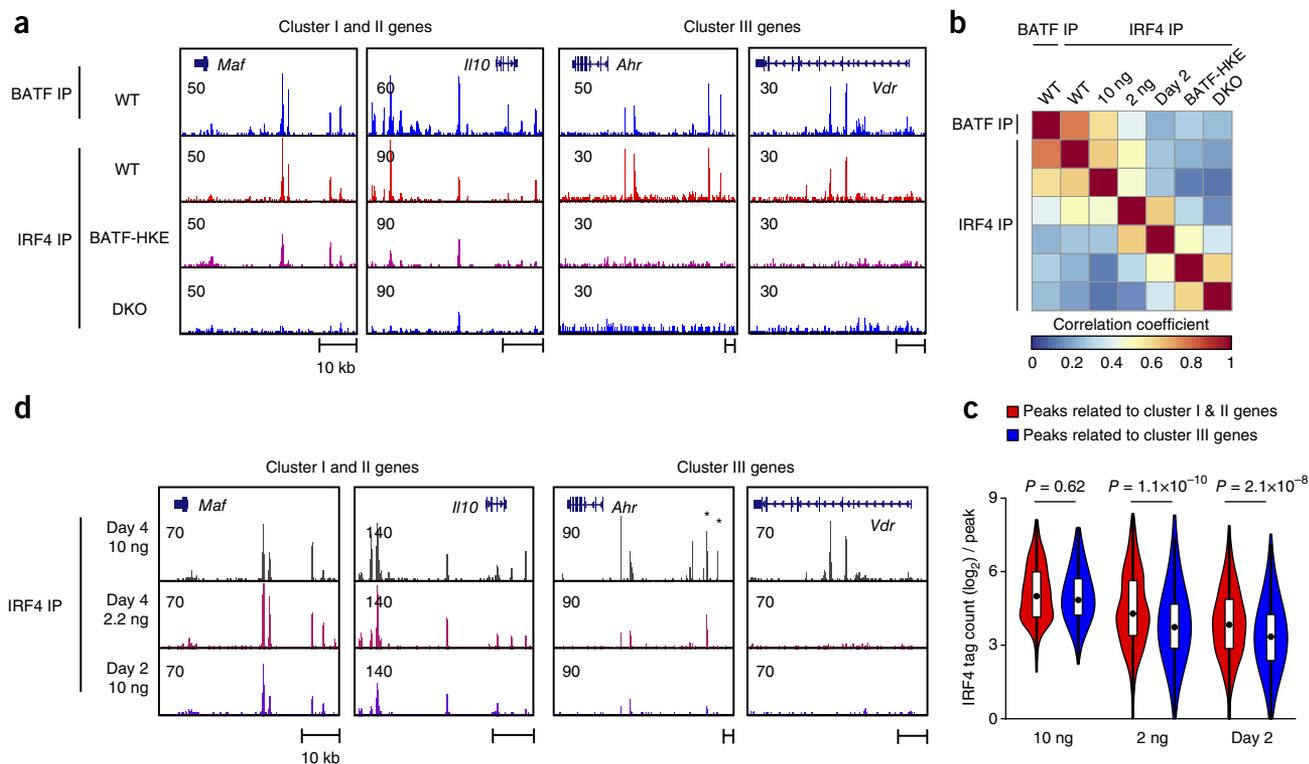
**Figure 2** GATA-3, IL-10 and CTLA-4 are differentially sensitive to the level of total BATF. **(a)** Flow cytometry analyzing the expression of GATA-3 and CTLA-4 in T<sub>H2</sub> cells from wild-type mice (WT), *Batf<sup>-/-</sup>* mice and *Batf1-Batf3* DKO mice (DKO) (above plots), assessed on day 4 after secondary stimulation with anti-CD3ε from ascites fluid (1:400 dilution) and anti-CD28 under T<sub>H2</sub> conditions (as in Fig. 1a). **(b)** Frequency of IL-10<sup>+</sup> cells (top) and mean fluorescence intensity (MFI) of GATA-3 (middle) and CTLA-4 (bottom) in cells obtained from *Batf<sup>+/+</sup>*, *Batf<sup>-/-</sup>* and *Batf1-Batf3* DKO mice (below plots) and cultured as in **a**. **(c)** Frequency of IL-10<sup>+</sup> cells and MFI of GATA-3 and CTLA-4 in T<sub>H2</sub> cells obtained from wild-type or *Batf1-Batf3* DKO mice and infected with empty retrovirus (empty-RV) or retrovirus expressing (-RV) *Batf*, *Batf3* or BATF-HKE (below plots), then analyzed on day 4 after secondary stimulation as in **a** and gated on human CD4 (as a marker of retroviral infection). **(d)** Frequency of IL-10<sup>+</sup> cells and MFI of GATA-3 and CTLA-4 in T<sub>H2</sub> cells from *Batf3<sup>-/-</sup>* mice that were *Batf<sup>+/+</sup>*, *Batf<sup>+/-</sup>* or *Batf<sup>-/-</sup>* (below plots), analyzed on day 4 after secondary stimulation as in **a**. **(e)** Flow cytometry analyzing the expression of GATA-3 and CTLA-4 on day 4 of primary activation of wild-type CD4<sup>+</sup> T cells cultured under T<sub>H2</sub> conditions (as in Fig. 1a) with anti-CD28 and various concentrations (above plots) of anti-CD3ε crosslinked by plate-bound anti-hamster IgG; cells were infected on day 1 with empty retrovirus or retrovirus expressing *Batf* (left margin). Each symbol (**b–d**) represents an individual mouse; small horizontal lines indicate the mean (± s.d.). NS, not significant ( $P > 0.05$ ); \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  and \*\*\*\* $P < 0.00001$  (one-way analysis of variance (ANOVA) with Tukey's (**b,d**) or Sidak's (**c**) multiple-comparison test). Data are representative of two experiments with  $n = 6$  mice per genotype (**a**) or two experiments (**e**) or are pooled from two experiments with  $n = 6$  mice per genotype (**b**), two (CTLA-4) or four (IL-10 and GATA3) experiments with  $n = 6$  mice (CTLA-4),  $n = 7$  mice (DKO + *Batf3*-RV and DKO + BATF-HKE-RV with IL-10 or GATA-3),  $n = 8$  mice (WT + empty-RV, DKO + empty-RV and DKO plus *Batf*-RV with IL-10 or GATA-3) (**c**), or three experiments with  $n = 8$  mice per genotype (**d**).

expression in activated *Batf1-Batf3* DKO  $T_H2$  cells than in wild-type  $T_H2$  cells, including *Foxp3* (**Supplementary Table 1**). Clusters I and II contained genes induced in wild-type  $T_H2$  cells by weak TCR stimulation, including *Il10*, *Ccr4*, *Maf* and *Prdm1* (**Supplementary**

**Table 1**). Cluster III contained genes that were induced only by strong TCR stimulation in wild-type  $T_H2$  cells, including *Ahr*, *Vdr*, *Ctla4*, *Hif1a* and *Socs3* (**Supplementary Table 1**). Within clusters, a spectrum of sensitivity to strength of activation was evident (**Fig. 3d**).



**Figure 3** BATF-dependent genes display a spectrum of sensitivity to the strength of TCR signaling. **(a)** Gene expression (horizontal axis) in wild-type and *Batf1-Batf3* DKO  $T_H2$  cells activated under  $T_H2$ -inducing conditions (as in **Fig. 1a**) with various concentrations (above plots) of anti-CD3 $\epsilon$  and assessed on day 2 or 4 after primary activation (above plots), plotted against  $P$  values (vertical axes; corrected by the Benjamini-Hochberg procedure); colors indicate a  $P$  value of  $<0.01$  and expression more than threefold higher (red) or lower (blue) in wild-type  $T_H2$  cells than in *Batf1-Batf3* DKO  $T_H2$  cells (WT vs DKO). Numbers in plots indicate total genes with expression more than threefold higher (red; bottom right) or lower (blue; bottom left) in wild-type cells than in *Batf1-Batf3* cells. **(b)** Spearman's rank correlation coefficient (key) for gene expression ( $\log_2$  values) of BATF-dependent genes identified in **a** in unstimulated wild-type  $CD4^+$  T cells (Day 0) and wild-type  $CD4^+$  T cells activated as in **a** (above plot). **(c)** Hierarchical clustering (right margin) of BATF-dependent genes identified in **a** by change in expression ( $\log_2$  values; key) of unstimulated wild-type  $CD4^+$  T cells (Day 0) and wild-type  $CD4^+$  T cells activated for 4 d with various concentrations (below plot) of anti-CD3 $\epsilon$  (primary activation); left margin, clustering dendrogram. **(d)** Expression ( $\log_2$  values; key) of genes (right margin) induced by weak TCR stimulation (High-sensitivity genes; left), in clusters I and II (as in **c**), or strong TCR stimulation (Low-sensitivity genes; right), in cluster III (as in **c**), in wild-type and *Batf1-Batf3* DKO  $T_H2$  cells treated as in **a** (below plots). Data are pooled from two independent experiments with three samples (**a,c**) or are from two independent experiments with three samples (**b,d**).



**Figure 4** ChIP-seq analysis of the binding of BATF and IRF4 correlates with the sensitivity of target genes to TCR signaling. **(a)** ChIP-seq analysis (IP) of the binding of BATF (top) and IRF4 (below) to genomic regions associated with select genes from clusters I and II (left; *Maf* and *Il10*) or cluster III (right; *Ahr* and *Vdr*) in wild-type  $T_H2$  cells (BATF) and in wild-type and *Batf1-Batf3* DKO  $T_H2$  cells and in *Batf1-Batf3* DKO  $T_H2$  cells reconstituted with retroviral BATF-HKE (IRF4) (left margin), all stimulated with the phorbol ester PMA and ionomycin on day 4 of secondary stimulation; location of the gene body is adjacent to the gene symbols (top plots). **(b)** Spearman's rank correlation coefficient analysis of ChIP-seq tag counts ( $\log_2$  values) obtained from merged IRF4 peaks after primary stimulation at various doses of anti-TCR on days 2 and 4 (10 ng/ml on day 4, 2.2 ng/ml on day 4, and 10 ng/ml on day 2) and after secondary stimulation as in **a** (BATF IP WT, IRF4 IP WT, IRF4 IP BATF-HKE and IRF4 IP DKO). **(c)** IRF4 tag counts ( $\log_2$  values) per merged IRF4 peak 50 kb upstream or downstream of the transcription start site of genes in clusters I and II or cluster III (key) in the ChIP-seq analyses (experimental conditions, below plots), presented as 'violin plots'.  $P$  values (above plots), Mann-Whitney  $U$ -test with Bonferroni's correction. **(d)** ChIP-seq analysis of the binding of IRF4 to genes (as in **a**) in wild-type  $T_H2$  cells cultured for 2 or 4 d with anti-CD3e (10 ng/ml or 2.2 ng/ml) (left margin); asterisks (*Ahr* plot) indicate peaks at  $-108$  kb and  $-90$  kb. Data are from two independent experiments with one sample per condition.

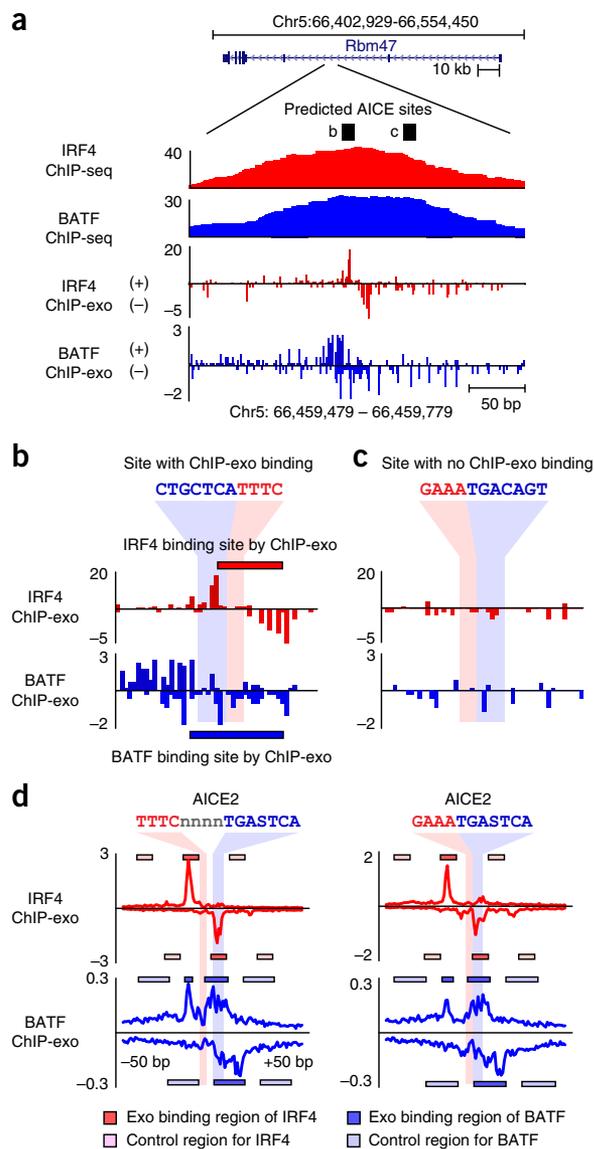
In summary, BATF-dependent genes induced in  $T_H2$  cells exhibited a wide range of sensitivity to TCR signal strengths.

#### Correlation of enhancer occupancy with TCR sensitivity

We performed ChIP-seq analysis of BATF in wild-type  $T_H2$  cells and of IRF4 in wild-type and *Batf1-Batf3* DKO  $T_H2$  cells. To mimic low BATF-IRF4 interaction, we did ChIP-seq analysis of IRF4 in *Batf1-Batf3* DKO  $T_H2$  cells stably reconstituted with BATF-HKE. Binding peaks for BATF and IRF4 were co-localized and had similar tag counts within selected genes from microarray clusters I and II, such as *Maf* and *Il10*, and within selected genes from cluster III, such as *Ahr* and *Vdr* (Fig. 4a). Such co-localization was consistent with the binding of BATF and IRF4 as a complex, as described for ChIP-seq analysis of BATF-IRF4 in  $T_H17$  cells<sup>13-16</sup>. IRF4 did not bind to those loci in *Batf1-Batf3* DKO  $T_H2$  cells, which showed that the binding of IRF4 to DNA was dependent on BATF in  $T_H2$  cells. Some IRF4 peaks were maintained in the presence of BATF-HKE at the *Maf* and *Il10* loci, which suggested that these were high-affinity IRF4-binding sites that did not require a strong BATF-IRF4 interaction. However, no IRF4 peaks were maintained with BATF-HKE in the *Ahr* and *Vdr* loci, which suggested that these binding sites were of low affinity and required a strong interaction between the BATF leucine zipper and IRF4.

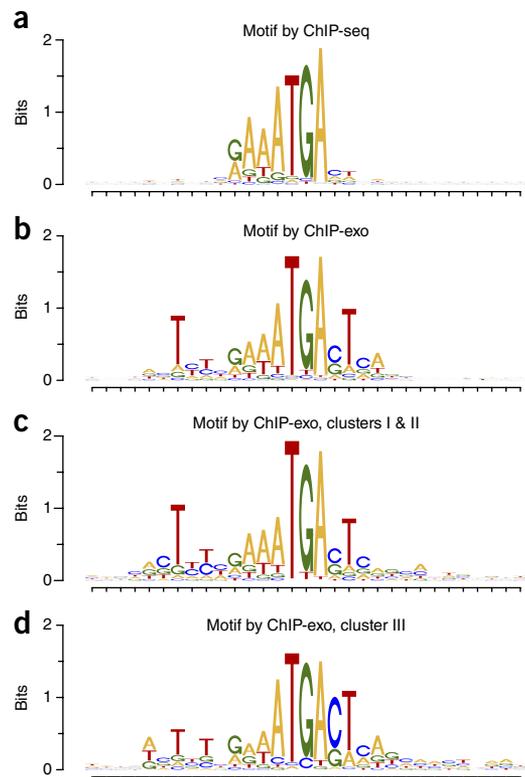
We performed additional ChIP-seq analysis of IRF4 using the same three TCR-stimulation conditions as those used for global

gene-expression analysis. We merged all binding peaks for IRF4 observed with any condition of primary stimulation and performed a Spearman's rank correlation analysis for the binding intensity of BATF and IRF4 (tag counts per peak) for all seven ChIP-seq experiments. The binding of BATF and that of IRF4 were highly correlated in wild-type cells ( $r = 0.77$ ) (Fig. 4b). The binding intensity of IRF4 was well correlated ( $r > 0.6$ ) in twice-activated wild-type  $T_H2$  cells and strong TCR-stimulation conditions, and in the two weakly stimulated samples (Fig. 4b). However, the binding intensities of IRF4 were more weakly correlated ( $r < 0.6$ ) for strong and weak TCR stimulation conditions (Fig. 4b). When experimental conditions were ordered by decreasing BATF expression, the Spearman's rank correlation coefficients for IRF4 tag counts were highest for samples of similar BATF expression (Fig. 4b). Globally, the binding of IRF4 to enhancers correlated with sensitivity to TCR stimulation. With strong TCR stimulation, the IRF4 tag counts of peaks in genes from clusters I and II (within 50 kb upstream or downstream of the transcription start site) were equal to those in genes from cluster III (Fig. 4c). However, with weak TCR stimulation, IRF4 tag counts of peaks in genes from cluster III were lower than those in genes from clusters I and II (Fig. 4c), in agreement with gene-specific differences in the binding of IRF4 to *Maf* and *Il10* (clusters I and II) and *Ahr* and *Vdr* (cluster III) (Fig. 4a). Specifically, IRF4-binding peaks near *Maf* and *Il10* were similar at strong and weak TCR stimulation (Fig. 4d). In contrast, the



**Figure 5** ChIP-exo reveals precise binding sites for BATF and IRF4 in the peaks with AICE motifs. **(a)** ChIP-seq analysis and ChIP-exo analysis of the binding of BATF and IRF4 (left margin) in wild-type  $T_H2$  cells; black bars indicate predicted AICE sites from *de novo* motif analysis; above plot (blue), genomic region containing *Rbm47*. Genomic coordinates refer to the NCBI37/mm9 NCBI assembly of the mouse genome. **(b,c)** Example sites from **a** with ChIP-exo tags **(b)** or no ChIP-exo tags **(c)**. **(d)** ChIP-exo tag counts (mean values) for the binding of BATF and IRF4 to consensus AICE1 and AICE2 motifs within IRF4 ChIP-seq peaks, showing flanking regions 50 bp upstream and downstream: red box, region with IRF4 ChIP-exo tags; light red box, control region for IRF4, blue box: region with BATF ChIP-exo tags, light blue box: control region for BATF (key). Data are merged from four independent experiments (BATF ChIP-exo) or two independent experiments (IRF4 ChIP).

IRF4-binding peaks near *Ahr* and *Vdr* were present with strong TCR stimulation but absent with weak TCR stimulation (**Fig. 4d**). The  $T_H2$  cell-related genes *Il4*, *Il5*, *Il13* and *Ctla4* were in cluster III and showed binding of IRF4 (by ChIP-seq analysis) consistent with low sensitivity to TCR stimulation (**Supplementary Fig. 4**). These results suggested that genes with enhancers that bound BATF–IRF4 with high affinity were induced at low TCR signal strength and tolerated weaker interactions between BATF and IRF4. In contrast, genes with



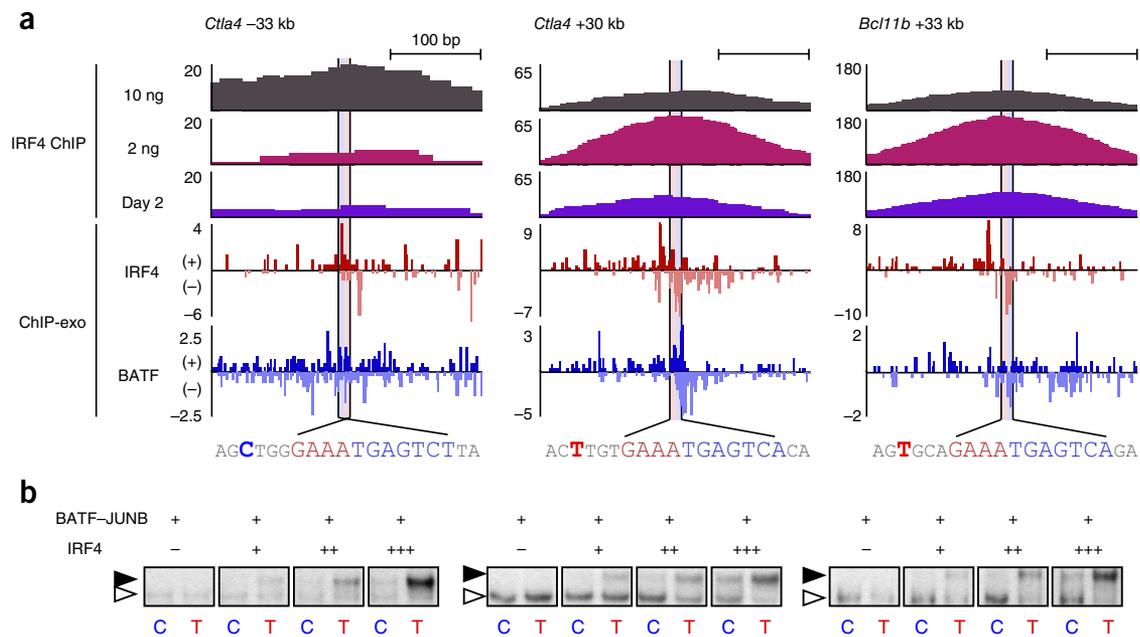
**Figure 6** ChIP-exo analysis reveals a previously unknown AICE2 motif. **(a)** AICE2 motif predicted by *de novo* motif analysis of IRF4 ChIP-seq. **(b)** AICE2 motif identified from sites with ChIP-exo binding. **(c)** AICE2 motif determined by ChIP-exo analysis in peaks 50 kb upstream or downstream of the transcription start sites of genes in clusters I and II. **(d)** AICE2 motif determined by ChIP-exo in peaks 50 kb upstream or downstream of the transcription start sites of genes in cluster III. Data are from the results in **Figures 4** and **5**.

enhancers that bound BATF–IRF4 with low affinity were induced at higher TCR signal strength and required full, strong interaction between the BATF leucine zipper and IRF4.

To determine whether compensation for BATF by BATF3 occurred through direct binding of BATF3, we performed ChIP-seq analysis of BATF3 and IRF4 in *Batf*<sup>-/-</sup> cells using conditions of strong TCR stimulation. BATF3- and IRF4-binding sites in *Batf*<sup>-/-</sup> cells coincided with BATF- and IRF4-binding sites in wild-type cells for genes in clusters I–II and cluster III, as well as in  $T_H2$  cell-related genes (**Supplementary Fig. 4c,d**). Thus, BATF3 was sufficient to allow binding of IRF4 to many sites on genes in clusters I and II and some sites on genes in cluster III in *Batf*<sup>-/-</sup> cells, in agreement with compensation for the expression of GATA-3 and IL-10 by BATF3 in *Batf*<sup>-/-</sup>  $T_H2$  cells. Some binding peaks for BATF–IRF4 in the *Ctla4* locus in wild-type cells were not bound by BATF3 and IRF4 in *Batf*<sup>-/-</sup> cells (**Supplementary Fig. 4d**), which might explain the failure of BATF3 to compensate for BATF in inducing the expression of CTLA-4. In summary, some BATF-dependent genes with enhancers that bound BATF–IRF4 with high affinity were expressed in *Batf*<sup>-/-</sup>  $T_H2$  cells because they bound BATF3 even though it had lower expression than BATF.

#### High-sensitivity AICE motif revealed by ChIP-exo analysis

Some IRF4 peaks (for example, an *Ahr* peak 108 kb upstream of the transcription start site (‘–108 kb peak’)) were present only with strong TCR stimulation, while others (for example, *Ahr* –90 kb peak) were also present with weak TCR stimulation (**Fig. 4d**). To identify



**Figure 7** DNA sequences flanking identical AICE motifs regulate the DNA-binding affinity of BATF-IRF4 and enhancer activity. **(a)** ChIP-seq analysis of the binding of IRF4 to genes in mouse  $T_H2$  cells cultured in primary activation conditions (left margin) and ChIP-exo analysis of the binding of IRF4 and BATF, showing selected AICE2-containing peaks containing an upstream cytosine (bold blue C; *Ctla4* -33 kb) or an upstream thymine (bold red T; *Ctla4* +30 kb and *Bcl11b* +33 kb). **(b)** EMSA of nuclear extracts of HEK293 FT human embryonic kidney cells expressing BATF and JUNB (BATF-JUNB) and various amounts of IRF4 (above blots), assessed with probes (full sequence, **Supplementary Table 2**) based on the AICE2 motifs in **a** containing the native sequence, a cytosine-to-thymine mutation (*Ctla4* -33 kb) or thymine-to-cytosine mutations (*Ctla4* +30 kb and *Bcl11b* +33 kb) at the -4 bp position (letters under plot indicate the presence of cytosine (blue C) or thymine (red T)); arrowheads (left margins) indicate the BATF-JUNB DNA complex (open) or BATF-JUNB-IRF4 DNA complex (filled). Data are from the results in **Figures 4** and **5** (a) or are representative of two experiments (b).

sequences that might distinguish high-affinity peaks from low-affinity peaks, we first classified peaks as being of high, intermediate or low affinity on the basis of the number of conditions of TCR stimulation in which they occurred (**Supplementary Fig. 5a**). Genes from clusters I and II contained more high-affinity peaks and fewer low-affinity peaks than did genes from cluster III (**Supplementary Fig. 5b**). However, *de novo* motif analysis of IRF4 peaks of all categories identified enrichment for AP-1, AICE1, AICE2 and ETS motifs but found no significant differences between categories in their sequence motifs (**Supplementary Fig. 5c**).

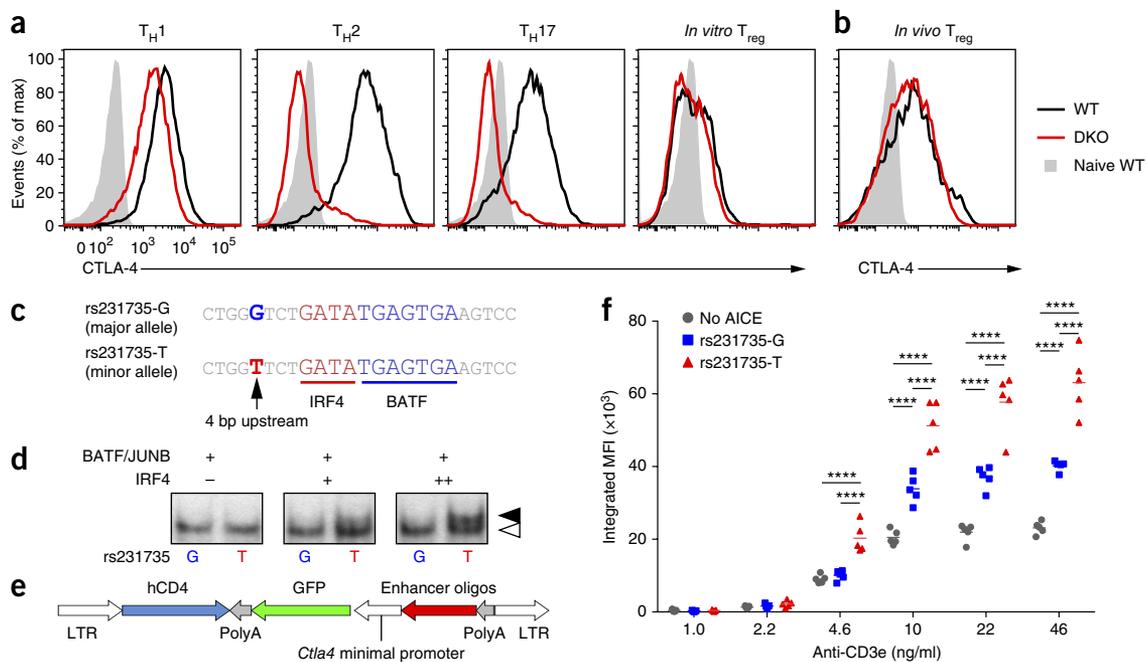
We sought to determine whether nucleotides flanking an AICE motif could influence the binding affinity of BATF-IRF4. To assess this, we used four pairs of electrophoretic mobility-shift assay (EMSA) probes with identical AICE sequences but different flanking regions (**Supplementary Fig. 6a-d**). We generated AICE probes that had identical core AICE sequences from four high-affinity peaks (AICE1, *Enpp6* -45 kb and *Bcor* +65 kb; and AICE2, *Prdm1* +14 kb and *Ccr4* +8 kb) and from intermediate-affinity or low-affinity peaks (AICE1, *Ptchd3* -26 kb and *Mzt1* +230 kb; and AICE2, *Ctla4* -33 kb and *Snrpe* +38 kb). These probes differed in the flanking genomic sequences surrounding identical AICE motifs. Probes from all high-affinity peaks produced a strong BATF-IRF4 complex by EMSA, while probes from intermediate- or low-affinity peaks produced a substantially weaker complex (**Supplementary Fig. 6e,f**). Thus, the affinity of an AICE motif for BATF-IRF4 was affected by DNA sequences outside the core AICE motif, but *de novo* motif analysis of ChIP-seq data was unable to resolve the motif further.

Often, more than one AICE motif was found within one IRF4 ChIP-seq peak (**Fig. 5a** and **Supplementary Fig. 7a,b**). Conceivably, a peak could arise if one AICE were actually a true binding site for BATF-IRF4, with the other(s) being included on the basis of its (their) proximity to that site. To optimize motif analysis to distinguish the

sequence requirements of high-affinity AICE motifs versus those of low-affinity AICE motifs, we needed to identify which motif(s) within a peak was (were) functional for binding BATF-IRF4. Thus, we performed ChIP-exo sequencing<sup>29</sup> in wild-type  $T_H2$  cells. As an example, a peak within intron 2 of *Rbm47* had two AICE motif sites (**Fig. 5a**). However, ChIP-exo revealed that only one site was actually occupied (**Fig. 5a-c**). In other peaks, both AICE motifs identified from motif analysis by ChIP-seq were also found to bind by ChIP-exo analysis (**Supplementary Fig. 7a-c**).

To determine which predicted motifs bound IRF4 and BATF, by ChIP-exo analysis, we analyzed the mean ChIP-exo tag counts within 50 bp upstream or downstream of all AICE1, AICE2 and ETS motifs predicted by *de novo* motif analysis or identified as consensus motifs. ETS motif sites within IRF4 peaks did not show enrichment for binding of BATF and IRF4, as assessed by ChIP-exo analysis (**Supplementary Fig. 8a,b**); this suggested that BATF and IRF4 did not bind directly to ETS motifs. However, in AICE1 and AICE2 sites, a position just upstream of the consensus binding motif for IRF (TTTC in AICE1, or GAAA in AICE2) showed significant enrichment for IRF4 ChIP-exo tag counts, and sites upstream of the consensus binding motif for AP-1 (TGASTCA), as well as those upstream of the consensus binding motif for IRF, showed enrichment for BATF ChIP-exo tag counts (**Fig. 5d** and **Supplementary Fig. 8a**). Next, we filtered AICE1 and AICE2 motifs, either predicted by *de novo* motif analysis or by consensus motifs, to identify sites that had significantly higher ChIP-exo tag counts upstream of IRF4- and BATF-binding sites than on flanking control bases. This analysis defined 1,960 sites for AICE1 and 1,506 sites for AICE2 (data not shown).

The AICE1 motif identified from the sites with ChIP-exo binding was nearly identical to the AICE1 motif determined by *de novo* motif analysis via ChIP-seq (**Supplementary Fig. 8c,d**). In contrast, the



**Figure 8** A SNP in human *CTLA4* affects the DNA-binding affinity of BATF–IRF4 and enhancer activity. **(a)** Flow cytometry analyzing CTLA-4 expression in naive wild-type T cells ( $CD3^+CD4^+CD8^-CD25^-CD44^-$ ) (Naive WT) or in T cells from wild-type mice or *Batf1-Batf3* DKO mice (key) after culture for 4 d on crosslinked anti-CD3e and anti-CD28 under  $T_H1$  conditions (anti-IL-4, IL-12 and IFN- $\gamma$ ),  $T_H2$  conditions (anti-IL-12, anti-IFN- $\gamma$  and IL-4),  $T_H17$  conditions (anti-IFN- $\gamma$ , anti-IL-12, anti-IL-4, IL-6, TGF- $\beta$  and IL-1 $\beta$ ) or  $T_{reg}$  cell conditions (anti-IFN- $\gamma$ , anti-IL-12, anti-IL-4, TGF- $\beta$ ) (above plots). **(b)** Flow cytometry analyzing CTLA-4 expression in  $CD3^+CD4^+$  Foxp3 $^+$  T cells in mesenteric lymph nodes MLNs from wild-type or *Batf1-Batf3* DKO mice, or naive T cells from wild-type mice as in **(a)** (key). **(c)** Sequences of the human SNP rs231735 (*CTLA4* –38kb); below (underline), binding sites for IRF4 and BATF. **(d)** EMSA of nuclear extracts of HEK293 FT cells expressing BATF and JUNB plus various amounts of IRF4 (above blots), assessed with probes (full sequence, **Supplementary Table 2**) based on rs231735-G and rs231735-T (below plots) (arrowheads (right margin), as in **Fig. 7b**). **(e)** Structure of reporter retrovirus for **f**. **(f)** Reporter activity in  $T_H2$  cells expressing retroviral reporter construct containing no AICE, rs231735-G or rs231735-T (key) and incubated with various concentrations of anti-CD3e (horizontal axis), presented as integrated MFI. \*\*\*\* $P < 0.0001$  (two-way ANOVA with Tukey's multiple comparison). Data are representative of two experiments (**a**), one experiment with two biological replicates (**b**) or two experiments (**d**) or are pooled from two experiments with  $n = 5$  mice (**f**).

AICE2 motif identified from sites with ChIP-exo binding showed enrichment for thymine at a position 4 bp upstream of the consensus binding motif for IRF relative to the frequency of thymine at this position in the AICE2 motif identified by ChIP-seq (**Fig. 6a,b**). Genes from clusters I and II showed enrichment for this thymine-containing AICE2 motif relative to its frequency in genes from cluster III, consistent with its association with high-affinity peaks (**Fig. 6c,d**). These results suggested that a thymine located at –4 bp in the AICE2 motif might increase the affinity of the binding of BATF–IRF4 to DNA and might ‘tune’ the sensitivity of target genes for activation by BATF–IRF4.

#### Identification of a high-affinity AICE consensus motif

We tested the functional effect of a thymine at –4 bp in an AICE2 by analyzing three EMSA probes derived from peaks in *Ctla4* and *Bcl11b* that bound IRF4 by ChIP-exo analysis (**Fig. 7a**). A probe derived from the low-affinity peak in the –33 kb region of *Ctla4* was an AICE2 motif with cytosine at the –4 bp position in the native genome; the native probe did not form a BATF–IRF4 complex in EMSA even at high concentrations of IRF4 (**Fig. 7b**). However, changing the nucleotide at –4 bp from cytosine to thymine led to the formation of a strong BATF–IRF4 complex (**Fig. 7b**). Two other probes derived from high-affinity peaks in the +30 kb region of *Ctla4* and the +33 kb region of *Bcl11b* contained thymine at the –4 bp position in the native genome (**Fig. 7a**). Both probes formed a strong BATF–IRF4 complex by EMSA (**Fig. 7b**). However, changing the thymine at –4 bp to cytosine led

to loss of the complex at all concentrations of IRF4 (**Fig. 7b**). These results showed that the affinity of AICE2 motifs could be ‘tuned’ by the DNA sequence at the –4 bp position, which lies outside the previously recognized AICE consensus motif.

Notably, an example of the high-affinity AICE2 sequence is present within a SNP in the human *CTLA4* locus, in which substitution of thymine for guanine at –38 kb (rs231735) has been reported as a protective SNP for rheumatoid arthritis<sup>26</sup> and granulomatosis with polyangiitis<sup>27</sup>. Since these are  $T_H17$  cell-related diseases<sup>30,31</sup>, we analyzed CTLA-4 expression in mouse  $T_H17$  cells and found that it was also dependent on BATF in these cells (**Fig. 8a,b**). The *CTLA4* SNP was located –4 bp relative to an AICE2 motif (**Fig. 8c**). Conceivably, this SNP might increase the sensitivity of CTLA-4 expression to TCR stimulation by increasing the binding of BATF–IRF4. An EMSA probe based on the major guanine-containing allele of rs231735 (rs231735-G) did not form a BATF–IRF4 complex, even at high concentrations of IRF4 (**Fig. 8d**). However, a probe based on the thymine-containing allele of rs231735 (rs231735-T) formed a BATF–IRF4 complex (**Fig. 8d**), similar to the thymine-containing AICE2 probes above (**Fig. 7b**).

We tested that SNP in a functional assay *in vivo* in T cells. We used a retrovirus-based reporter<sup>32</sup> containing the *Ctla4* minimal promoter with or without insertion of an upstream 36-bp region with rs231735-G or rs231735-T (**Fig. 8e**). We assessed the activity of these reporters after stable integration into  $T_H2$  cells by activation with increasing doses of anti-CD3e. The promoter lacking either region

showed basal levels of stimulation (Fig. 8f). Promoter activity was increased somewhat by the rs231735-G enhancer but was significantly increased further at all doses of anti-CD3 $\epsilon$  by the rs231735-T enhancer. Notably, with the intermediate level of TCR stimulation (4.6 ng/ml of anti-CD3 $\epsilon$ ), activity was observed only with rs231735-T. These results highlighted the importance of a SNP for the binding of a transcription factor to a site and for optimal, finely tuned gene expression.

## DISCUSSION

Our results have resolved a discrepancy related to the role of BATF in T<sub>H</sub>2 cell development<sup>24,25</sup>. A requirement for BATF in the development of T<sub>H</sub>17 and T<sub>FH</sub> cells but not in the development of T<sub>H</sub>1 or T<sub>H</sub>2 cells has been reported in studies of *Batf*<sup>-/-</sup> mice in which exons 1 and 2 were deleted<sup>21,25</sup>, but another study of *Batf*<sup>-/-</sup> mice in which exon 3 was deleted has reported an additional role for BATF in the development of T<sub>H</sub>2 cells<sup>24</sup>. This discrepancy could have resulted from differences in genetic backgrounds, since the former studies used 129SvEv mice<sup>21,25</sup>, while the latter study used C57BL/6 mice<sup>24</sup>. However, T<sub>H</sub>2 cell development is abolished in *Batf*<sup>-/-</sup> mice in which *Batf3* is also deleted<sup>13</sup>. That result shows that T<sub>H</sub>2 development depends on activity of the BATF family and suggests that the discrepancy could have resulted from differences between the studies in the amount of compensation provided by BATF3. Here, we confirmed that interpretation by directly showing that BATF3 induced T<sub>H</sub>2 cell development when expressed in *Batf1-Batf3* DKO T cells. Furthermore, *Gata3* was a BATF-dependent gene induced by low TCR signal strength and was fully induced at a level of 50% of BATF on a *Batf3*<sup>-/-</sup> background, and BATF-HKE compensated for its expression. For those reasons, GATA-3 appears to be selectively compensated for by low levels of BATF3 expressed in T cells. Thus, differences in the strength of TCR stimulation and the genetic backgrounds used could have contributed to the apparent discrepancy in T<sub>H</sub>2 cell development in *Batf*<sup>-/-</sup> mice<sup>24,25</sup>. However, we have not addressed whether those findings also explain published claims about antigen dose and T<sub>H</sub>1 cell–T<sub>H</sub>2 cell balance<sup>4,5</sup> or T<sub>H</sub>1 cell–T<sub>FH</sub> cell balance<sup>6</sup>.

Second, our results addressed the effect of graded IRF4 expression on the activation of CD8<sup>+</sup> T cells<sup>9,33</sup>. The population expansion of CD8<sup>+</sup> T cells requires IRF4, which is induced in a graded manner in response to different strengths of TCR signaling<sup>9</sup>. We found that BATF was also induced in a graded manner coordinately with IRF4. At increasing levels of BATF–IRF4 expression, we identified a hierarchy of induced genes, with genes in clusters I and II representing highly sensitive responder targets and genes in cluster III being low-sensitivity targets. We also identified (by ChIP-seq analysis) a hierarchy of IRF4-binding sites that depended on levels of BATF–IRF4 expression. Notably, global levels of IRF4 binding were different for these clusters at low levels of signaling but not at high levels of signaling, or at early times after activation. Further, the genes in cluster I and II contained more high-affinity IRF4-binding sites than did genes in cluster III. A published study has shown that histone H3 acetylated at Lys27 (H3K27ac) at BATF-binding sites<sup>15</sup> increases as the strength of TCR signaling is increased<sup>34</sup>. Our results extended that correlation by showing that the binding of IRF4 to BATF-binding sites also increased in proportion to the strength of TCR stimulation. However, we found that IRF4 binding did not increase uniformly across the genome but occurred ‘preferentially’ at genes with high sensitivity to BATF. In summary, we have shown that analog expression of BATF and IRF4 increasing in proportion to the strength of TCR stimulation induced a hierarchy of gene expression based on differing affinities of enhancer binding sites for the BATF–IRF4 complex.

Third, our results demonstrated how enhancers with AICE motifs were able to respond to different levels of BATF–IRF4. In essence, we found that flanking sequences surrounding a recognized AICE motif strongly influenced affinity for the BATF–IRF4 complex in a chromatin-independent manner. Varying the affinity of enhancers for transcription factors is a recognized mechanism for generating graded responses to varying strength of signaling, as, for example, in morphogen-dependent expression of target genes controlled by motifs with different affinities for Dorsal<sup>35,36</sup>. Similarly, we identified enhancers with AICE1 or AICE2 motifs of identical core sequence that had variable *in vivo* affinity for binding BATF–IRF4, by ChIP-seq analysis. When we compared these regions by EMSA, we found that flanking sequences determined overall binding in EMSA in the same pattern as that observed *in vivo* by ChIP-seq. For example, *Prdm1* and *Ctla4* each had an AICE2 motif identified by ChIP-seq with an identical core sequence: GAAATGAGTCT. The site in *Prdm1* was of high affinity, on the basis of occupancy at low TCR signal strength, while the site in *Ctla4* was occupied only at high TCR signal strength. Notably, the EMSA complex formed by the AICE2 region of *Prdm1* was much stronger than that formed by the region of *Ctla4*, despite their identical core motifs; this suggested that *in vitro* binding might reflect *in vivo* affinity. As BATF and IRF4 bind to an AICE motif that functions in T<sub>H</sub>17 cell development<sup>16</sup>, the ability of flanking sequences to modulate the affinity of an AICE motif for BATF–IRF4 provides a mechanism for controlling the sensitivity of target genes to TCR signal strength. Base pairs flanking transcription-factor-binding sites are known to regulate the binding of transcription factors through DNA structure and guanine-cytosine content<sup>37</sup>. Dysregulation of the immune system can be observed in humans with heterozygous germline mutations of *CTLA4* (ref. 38), and the human *CTLA4* SNP examined in our study is associated with protection for autoimmunity<sup>39</sup>. The thymine located at –4 bp in the AICE2 motif of this SNP increased *in vitro* binding in EMSA and increased enhancer activity *in vivo*, so perhaps an increase in CTLA-4 expression could act to repress autoimmunity<sup>40</sup>, although further work is clearly needed.

## METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

## ACKNOWLEDGMENTS

We thank the Alvin J. Siteman Cancer Center at Washington University School of Medicine for use of the Center for Biomedical Informatics and Multiplex Gene Analysis Geneschip Core Facility. Supported by the Howard Hughes Medical Institute (K.M.M.) and the US National Institutes of Health (RO1 AI097244-01A1 to T.E., F30DK108498 to V.D., 1F31CA189491-01 to G.E.G.-R.).

## AUTHOR CONTRIBUTIONS

A.I., T.L.M. and K.M.M. designed the study, and wrote the manuscript with contributions from all authors; A.I. and C.G.B. performed experiments related to cell sorting, culture and flow cytometry with advice from V.D., R.T., G.E.G.-R. and T.E.; A.I. and C.G.B. performed microarray experiments with advice from X.W. and T.L.M.; and A.I. performed and analyzed ChIP-Seq experiments with advice from V.D. and T.E.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Li, M.O. & Rudensky, A.Y. T cell receptor signalling in the control of regulatory T cell differentiation and function. *Nat. Rev. Immunol.* **16**, 220–233 (2016).
2. Tubo, N.J. & Jenkins, M.K. TCR signal quantity and quality in CD4<sup>+</sup> T cell differentiation. *Trends Immunol.* **35**, 591–596 (2014).
3. Mayya, V. & Dustin, M.L. What scales the T cell response? *Trends Immunol.* **37**, 513–522 (2016).
4. Constant, S. *et al.* Extent of T cell receptor ligation can determine the functional differentiation of naive CD4<sup>+</sup> T cells. *J. Exp. Med.* **182**, 1591–1596 (1995).
5. Hosken, N.A. *et al.* The effect of antigen dose on CD4<sup>+</sup> T helper cell phenotype development in a T cell receptor- $\alpha\beta$ -transgenic model. *J. Exp. Med.* **182**, 1579–1584 (1995).
6. Tubo, N.J. *et al.* Single naive CD4<sup>+</sup> T cells from a diverse repertoire produce different effector cell types during infection. *Cell* **153**, 785–796 (2013).
7. Saravia, M. *et al.* Interleukin-10 production by Th1 cells requires interleukin-12-induced STAT4 transcription factor and ERK MAP kinase activation by high antigen dose. *Immunity* **31**, 209–219 (2009).
8. Man, K. *et al.* The transcription factor IRF4 is essential for TCR affinity-mediated metabolic programming and clonal expansion of T cells. *Nat. Immunol.* **14**, 1155–1165 (2013).
9. Nayar, R. *et al.* Graded levels of IRF4 regulate CD8<sup>+</sup> T cell differentiation and expansion, but not attrition, in response to acute virus infection. *J. Immunol.* **192**, 5881–5893 (2014).
10. Yao, S. *et al.* Interferon regulatory factor 4 sustains CD8<sup>+</sup> T cell expansion and effector differentiation. *Immunity* **39**, 833–845 (2013).
11. Huber, M. & Lohoff, M. IRF4 at the crossroads of effector T-cell fate decision. *Eur. J. Immunol.* **44**, 1886–1895 (2014).
12. De Silva, N.S. *et al.* The diverse roles of IRF4 in late germinal center B-cell differentiation. *Immunol. Rev.* **247**, 73–92 (2012).
13. Tussiwand, R. *et al.* Compensatory dendritic cell development mediated by BATF-IRF interactions. *Nature* **490**, 502–507 (2012).
14. Glasmacher, E. *et al.* A genomic regulatory element that directs assembly and function of immune-specific AP-1-IRF complexes. *Science* **338**, 975–980 (2012).
15. Li, P. *et al.* BATF-JUN is critical for IRF4-mediated transcription in T cells. *Nature* **490**, 543–546 (2012).
16. Ciofani, M. *et al.* A validated regulatory network for th17 cell specification. *Cell* **151**, 289–303 (2012).
17. Murphy, T.L., Tussiwand, R. & Murphy, K.M. Specificity through cooperation: BATF-IRF interactions control immune-regulatory networks. *Nat. Rev. Immunol.* **13**, 499–509 (2013).
18. Ochiai, K. *et al.* Transcriptional regulation of germinal center B and plasma cell fates by dynamical control of IRF4. *Immunity* **38**, 918–929 (2013).
19. Jabeen, R. *et al.* Th9 cell development requires a BATF-regulated transcriptional network. *J. Clin. Invest.* **123**, 4641–4653 (2013).
20. Kurachi, M. *et al.* The transcription factor BATF operates as an essential differentiation checkpoint in early effector CD8<sup>+</sup> T cells. *Nat. Immunol.* **15**, 373–383 (2014).
21. Ise, W. *et al.* The transcription factor BATF controls the global regulators of class-switch recombination in both B cells and T cells. *Nat. Immunol.* **12**, 536–543 (2011).
22. Hildner, K. *et al.* Batf3 deficiency reveals a critical role for CD8 $\alpha^+$  dendritic cells in cytotoxic T cell immunity. *Science* **322**, 1097–1100 (2008).
23. Yosef, N. *et al.* Dynamic regulatory network controlling T<sub>H</sub>17 cell differentiation. *Nature* **496**, 461–468 (2013).
24. Betz, B.C. *et al.* Batf coordinates multiple aspects of B and T cell function required for normal antibody responses. *J. Exp. Med.* **207**, 933–942 (2010).
25. Schraml, B.U. *et al.* The AP-1 transcription factor Batf controls T<sub>H</sub>17 differentiation. *Nature* **460**, 405–409 (2009).
26. Gregersen, P.K. *et al.* REL, encoding a member of the NF- $\kappa$ B family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat. Genet.* **41**, 820–823 (2009).
27. Chung, S.A. *et al.* Meta-analysis of genetic polymorphisms in granulomatosis with polyangiitis (Wegener's) reveals shared susceptibility loci with rheumatoid arthritis. *Arthritis Rheum.* **64**, 3463–3471 (2012).
28. Lohoff, M. *et al.* Dysregulated T helper cell differentiation in the absence of interferon regulatory factor 4. *Proc. Natl. Acad. Sci. USA* **99**, 11808–11812 (2002).
29. Rhee, H.S. & Pugh, B.F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408–1419 (2011).
30. McInnes, I.B. & Schett, G. The pathogenesis of rheumatoid arthritis. *N. Engl. J. Med.* **365**, 2205–2219 (2011).
31. Chen, M. & Kallenberg, C.G. ANCA-associated vasculitides—advances in pathogenesis and treatment. *Nat. Rev. Rheumatol.* **6**, 653–664 (2010).
32. Grajales-Reyes, G.E. *et al.* Batf3 maintains autoactivation of Irf8 for commitment of a CD8 $\alpha^+$  conventional DC clonogenic progenitor. *Nat. Immunol.* **16**, 708–717 (2015).
33. Nayar, R. *et al.* IRF4 regulates the ratio of T-Bet to Eomesodermin in CD8<sup>+</sup> T cells responding to persistent LCMV infection. *PLoS One* **10**, e0144826 (2015).
34. Allison, K.A. *et al.* Affinity and dose of TCR engagement yield proportional enhancer and gene activity in CD4<sup>+</sup> T cells. *eLife* **5**, e10134 (2016).
35. Papatsenko, D. & Levine, M. Quantitative analysis of binding motifs mediating diverse spatial readouts of the Dorsal gradient in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. USA* **102**, 4966–4971 (2005).
36. Ashe, H.L. & Briscoe, J. The interpretation of morphogen gradients. *Development* **133**, 385–394 (2006).
37. Levo, M. & Segal, E. In pursuit of design principles of regulatory sequences. *Nat. Rev. Genet.* **15**, 453–468 (2014).
38. Kuehn, H.S. *et al.* Immune dysregulation in human subjects with heterozygous germline mutations in CTLA4. *Science* **345**, 1623–1627 (2014).
39. Romo-Tena, J., Gomez-Martin, D. & Alcocer-Varela, J. CTLA-4 and autoimmunity: new insights into the dual regulator of tolerance. *Autoimmun. Rev.* **12**, 1171–1176 (2013).
40. Lenardo, M., Lo, B. & Lucas, C.L. Genomics of immune diseases and new therapies. *Annu. Rev. Immunol.* **34**, 121–149 (2016).

## ONLINE METHODS

**Mice.** Wild-type, *Batf*<sup>-/-</sup> and *Batf*<sup>-/-</sup> *Batf3*<sup>-/-</sup> mice<sup>22,25</sup> on the 129S6/SvEvTac background, DO11.10 and BALB/c mice were maintained in a specific-pathogen-free animal facility following institutional guidelines and with protocols approved by the Animal Studies Committee at Washington University in St. Louis. All experiments were performed with sex-matched mice 6–12 weeks of age without randomization or blinding.

**Antibodies and flow cytometry.** Cells were stained at 4 °C in the presence of Fc Block (2.4G2; BioXcell) in flow cytometry buffer (0.5% BSA in PBS). All flow cytometry antibodies were used at a dilution of 1:200. The following antibodies were used: PE-conjugated anti-CTLA-4 (UC10-4F10-11), PE-Cy7-conjugated anti-CD25 (PC81) and biotin-conjugated anti-CD8b (53-5.8) (all from Becton Dickinson (BD)); Pacific blue-conjugated anti-CD4 (RM4-5), PerCP/Cy5.5-conjugated anti-Thy1.1 (OX-7), Brilliant Violet421-conjugated anti-human CD4 (OKT4), APC-conjugated anti-human CD4 (RPA-T4) and biotin-conjugated anti-CD45R/B220 (RA3-6B2) (from BioLegend); APC-conjugated IL-10 (JESS-16E3), efluor660-conjugated anti-GATA-3 (TWAJ), PE-conjugated anti-IRF4 (3E4) and biotin-conjugated anti-CD49b (DX5) (all from eBioscience); FITC conjugated anti-CD4 (GK1.5) (from Tonbo Biosciences); and R-PE-conjugated anti-human CD4 (S3.5) (from Invitrogen). For IL-10 and CTLA-4 staining, cells were fixed in 2% paraformaldehyde for 15 min at RT and permeabilized with 0.5% Saponin before staining. For BATF, GATA-3 and IRF4, cells were fixed and permeabilized with Foxp3 Staining Buffer Set (eBioscience) following the manufacturer's instructions. Cells were analyzed on a FACSCanto II or FACSARIA Fusion and data were analyzed with FlowJo software (TreeStar).

**Isolation and culture of CD4<sup>+</sup> T cells.** All cells are cultured in IMDM supplemented with 10% FCS, 2 mM glutamine, 100 IU/ml penicillin, 100 µg/ml streptomycin, 1 mM sodium pyruvate, non-essential amino acids and 55 µM β-mercaptoethanol. Spleen and lymph node cells were harvested, treated with ACK lysis buffer and passed through a 70-µm nylon filter.

For time-course and TCR-dose-titration experiments (flow cytometry, microarray and ChIP) naive CD4<sup>+</sup> T cells were sorted as CD4<sup>+</sup>CD25<sup>-</sup>CD44<sup>+</sup> cells using spleen and lymph node cells after negative selection with biotinylated antibodies to B220, DX5 and CD8 (identified above), streptavidin-nanobeads and MojoSort Magnetic Cell Separation system (BioLegend). Naive CD4<sup>+</sup> T cells were activated with soluble anti-CD3ε (145-2C11, BioXCell) at 1, 2.2, 4.6, 10, 22 and 46 ng/ml cross-linked by plate-bound anti-hamster IgG (MP biomedical cat.# 55397 1:50) and anti-CD28 (37.51, BioXCell) (4 µg/ml) under T<sub>H</sub>2 conditions (anti-IFN-γ 10 µg/ml (XMG1.2, BioXCell), anti-IL-12 10 µg/ml (Tosh; prepared in-house), IL-4 10 ng/ml (Peprotech)) and were analyzed on at various times after primary activation as indicated in the figure legends.

For peptide-dose titration, naive KJ126<sup>+</sup> cells from DO11.10 mice were purified and activated with MACS-purified CD11c<sup>+</sup> dendritic cells from spleens of BALB/c mice in the presence of ovalbumin peptide (amino acids 323–339) under T<sub>H</sub>2 conditions. Cell populations were expanded threefold on day 3 and were analyzed by flow cytometry on day 4.

For experiments using secondary stimulations, CD4<sup>+</sup> T cells were isolated using Dynabeads FlowComp Mouse CD4 kit (Invitrogen) and were activated on plates coated with anti-CD3ε (500A2 ascites diluted 1:400 to achieve maximal activation) and anti-CD28 (37.51, 4 µg/ml, BioXCell) under T<sub>H</sub>2 conditions. On day 3, cells were diluted threefold in fresh media on uncoated plates. On day 7, cells were re-stimulated under the same conditions or were re-stimulated using soluble anti-CD3ε (145-2C11, 1 µg/ml, BioXCell) at various concentrations cross linked by plate-bound anti-hamster IgG (MP biomedical, cat.# 55397 1:50) and anti-CD28 (37.51, BioXCell) (4 µg/ml) under T<sub>H</sub>2 conditions. On day 4 of the second stimulation, cells were analyzed by flow cytometry for expression of GATA-3 and CTLA-4, or were activated by PMA and ionomycin in the presence of brefeldin A for 5 h for analysis of IL-10 expression, or for 2 h for ChIP.

**EMSA.** Oligonucleotide pairs were annealed to generate probes that were labeled with <sup>32</sup>P-dCTP using Klenow polymerase (Supplementary Table 2). HEK293FT cells were transiently transfected with retroviral vectors for *Batf*, *JunB* or *Irf4* using TransIT-LT1. After 48 h, cells were lysed with buffer

(10 mM HEPES-KOH, pH 7.9, 1.5 mM MgCl<sub>2</sub> and 10 mM KCl) containing 0.2% NP40 and protease inhibitors. Nuclei were pelleted, resuspended in buffer C (20 mM HEPES-KOH, pH 7.9, 420 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA and 25% glycerol), and centrifuged to obtain nuclear extracts<sup>41</sup>. EMSA was essentially as described<sup>42</sup> using combinations of nuclear extracts from cells transfected with *Batf*, *JunB* and *Irf4* (up to 1.5 µg total), 0.25 µg poly dI-dC (Sigma) and <sup>32</sup>P-labeled probes in 10 µl binding reactions for 20 min at 4 °C. Reactions were separated on 4–7%T 3.3%C polyacrylamide mini-gels in 0.4× TBE for 50 min at 250 V and 4 °C and were analyzed by autoradiography.

**Retroviral analysis.** Retroviral vectors were transfected into Plat-E cells with TransIT-LT1 (Mirus Bio) and viral supernatants were collected 2 d later. On day 1 after activation of CD4<sup>+</sup> T cells, culture medium was replaced with supernatants of transfected packaging cells containing 6 µg/ml polybrene. Cells were transduced by centrifugation at room temperature for 90 min at 1,170g. Viral supernatant was replaced by T<sub>H</sub>2 culture medium.

Enhancer elements were cloned into a retroviral reporter with additional insertion of polyA sequence upstream of enhancer sites (hCD4 pA GFP RV)<sup>32</sup> (Supplementary Table 3). For analysis, we used integrated MFI, which combines the metrics of frequency and MFI as a measure of total functional response<sup>43,44</sup>.

For ChIP-seq analysis of retrovirally transduced cells, infected cells were sorted on day 6 after primary activation, re-stimulated on day 7 and harvested on day 11 following PMA-ionomycin activation for 2 h.

**Expression microarray analysis.** Total RNA was extracted using RNAqueous-Micro Kit (Ambion) and was amplified with the Ovation Pico wild-typeA System (NuGEN) and hybridized to GeneChip Mouse Gene 1.0 ST microarrays (Affymetrix). Data were normalized by robust multiarray average summarization and quartile normalization with ArrayStar software (DNASTAR). The log<sub>2</sub>-transformed data were imported into the software of the R project (version 3.2.3). Differential expression analyses were performed using limma package of R and *P*-value were corrected by the Benjamini-Hochberg procedure<sup>45</sup>. Spearman's rank correlation coefficient plot was generated by R. Hierarchical clustering was performed with Euclidean distance and Ward clustering using of R.

**ChIP-seq.** ChIP was performed as described<sup>32</sup> with minor modifications. In brief, 1 × 10<sup>7</sup> activated CD4<sup>+</sup> T cells were collected, crosslinked with 1% formaldehyde at room temperature for 8 min, quenched with 1.25 M glycine and washed twice with PBS. Pellets were 'flash frozen' for storage at –80 °C. Chromatin was sonicated for 24 cycles of 20 s on and 50 s off per cycle with a Vivra-Cell VCX130PB and CV188 sonicators (Sonics & Material) in lysis buffer (10 mM Tris pH7, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.5% N-lauroylsarcosine). Chromatin was immunoprecipitated overnight at 4 °C with Dynabeads Protein A or G (Invitrogen) that had been pre-incubated with 5 µg of antibody: anti-IRF4 (sc-6059X; Santa Cruz Biotechnology), rabbit anti-BATF (prepared in-house)<sup>25</sup>, or rabbit anti-BATF3 (prepared in-house)<sup>13</sup>. Beads containing protein–DNA complexes were washed two times with RIPA buffer (10 mM Tris, pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.1% SDS, 0.1% sodium deoxycholate and 1% Triton-X), two times with RIPA buffer plus 0.3 M NaCl, two times with LiCl buffer (0.2 M LiCl, 0.5% NP-40 and 0.5% sodium deoxycholate), two times with Tris-EDTA-Triton-X buffer (10 mM Tris, pH 8.0, 1 mM EDTA, 0.2% Triton-X), and twice with Tris-EDTA. DNA fragments were eluted and reverse-crosslinked by incubation for 5 h at 65 °C in Tris-EDTA, pH 8.0, with 0.3% SDS, 1 mg/ml Proteinase K (New England Biolabs). DNA was purified by phenol-chloroform extraction followed by ethanol precipitation. Libraries for ChIP-seq were prepared from 1 ng of ChIPed DNA with a ThruPLEX-FD kit (Rubicon Genomics) and were sequenced with an Illumina HiSeq 2500 as single 'reads' extending 50 bases.

**Computational analysis for ChIP-seq.** ChIP-seq data sets were aligned to the mouse genome (NCBI37/mm9 assembly) by Bowtie software (version 1.1.1)<sup>46</sup> with the following parameters: –sam-best -p4 -m1 –chunkmbs 8000. Uniquely mapped reads were masked with Samtools<sup>47</sup> with blacklist of the ENCODE project<sup>48</sup> and the RepeatMasker program (which screens for interspersed repeats and low complexity) in the UCSC Genome Browser. Duplicated reads

are discarded using 'make tag directory' of Homer software package<sup>49</sup> with the parameter -tbp 1. Data were visualized with the 'makeUCSCfile' of Homer. Peaks from individual conditions were identified with 'findPeaks' of Homer with a 300 bp window and Poisson *P* value of  $<1 \times 10^{-10}$  and with normalized tag counts fourfold more than control (input sample). IRF4 peaks from individual of ChIP-seq experiments at primary stimulation conditions (10 ng/ml and 2 ng/ml on day 4 and 10 ng/ml on day 2) were merged using 'mergePeaks' of Homer and were centered and trimmed to 300 bp. Tag counts per peak were calculated with 'annotatePeaks.pl' of Homer. Spearman's rank correlation coefficient of tag counts on merged IRF4 peaks between each ChIP-seq experiment were performed by R. Peak related indicated genes were extracted from merged IRF4 peaks within transcription start site  $\pm 50$  kb using 'intersectBed' of BedTools package<sup>50</sup>. High-affinity peaks were defined as merged IRF4 peaks with tag counts with more than a threshold of Poisson *P* value of  $1 \times 10^{-10}$  in all three experiments (10 ng/ml and 2 ng/ml on day 4 and 10 ng/ml on day 2). Intermediate-affinity peaks were defined as merged IRF4 peaks with tag counts more than threshold in two of three experiments, and low-affinity peaks have tag counts more than threshold in one of three experiments. The heat map of binding intensity was generated using 'annotatePeaks.pl' of Homer and R. *De novo* motif analysis was performed from the top 3,000 peaks, which were ranked by sum of tag counts per peak of the three experiments (10 ng/ml and 2 ng/ml on day 4 and 10 ng/ml on day 2), using 'findMotifsGenome.pl' of Homer with 150 bp window. Putative motif loci of motifs from each category (high-affinity, intermediate-affinity and low-affinity) were extracted from merged IRF4 peaks with *de novo* motifs using 'annotatePeaks.pl', length of motifs were adjusted and motifs were merged to one bed file using 'intersectBed'. Motif logos were generated by the 'seqLogo' package of R.

**ChIP-exo analysis.** ChIP-exo analysis was performed as described<sup>51</sup> with minor modifications. In brief, 'ChIPed' DNA-antibody-bead complexes from  $30 \times 10^6$  cells per experiment were washed with RIPA buffer six times, and Tris-EDTA, pH 8.0, twice. The ChIP-exo libraries were made by the following enzymatic reactions with four washes between each reaction; End polishing by T4 DNA polymerase, Klenow DNA polymerase, T4 polynucleotide kinase, ligation of P7 exo-adaptor by T4 DNA ligase, nick repair by Phi29 DNA polymerase, exonuclease reaction by lambda exonuclease and cleaning-up single strand DNA by RecJf exonuclease. DNA was eluted and reverse-crosslinked as above. The libraries were generated by P7 primer extension with Phi29 DNA polymerase, ligation of P5 exo-adaptor with T4 DNA ligase, PCR amplification with Phusion polymerase for 12 cycles and were cleaned up by AMPure XP.

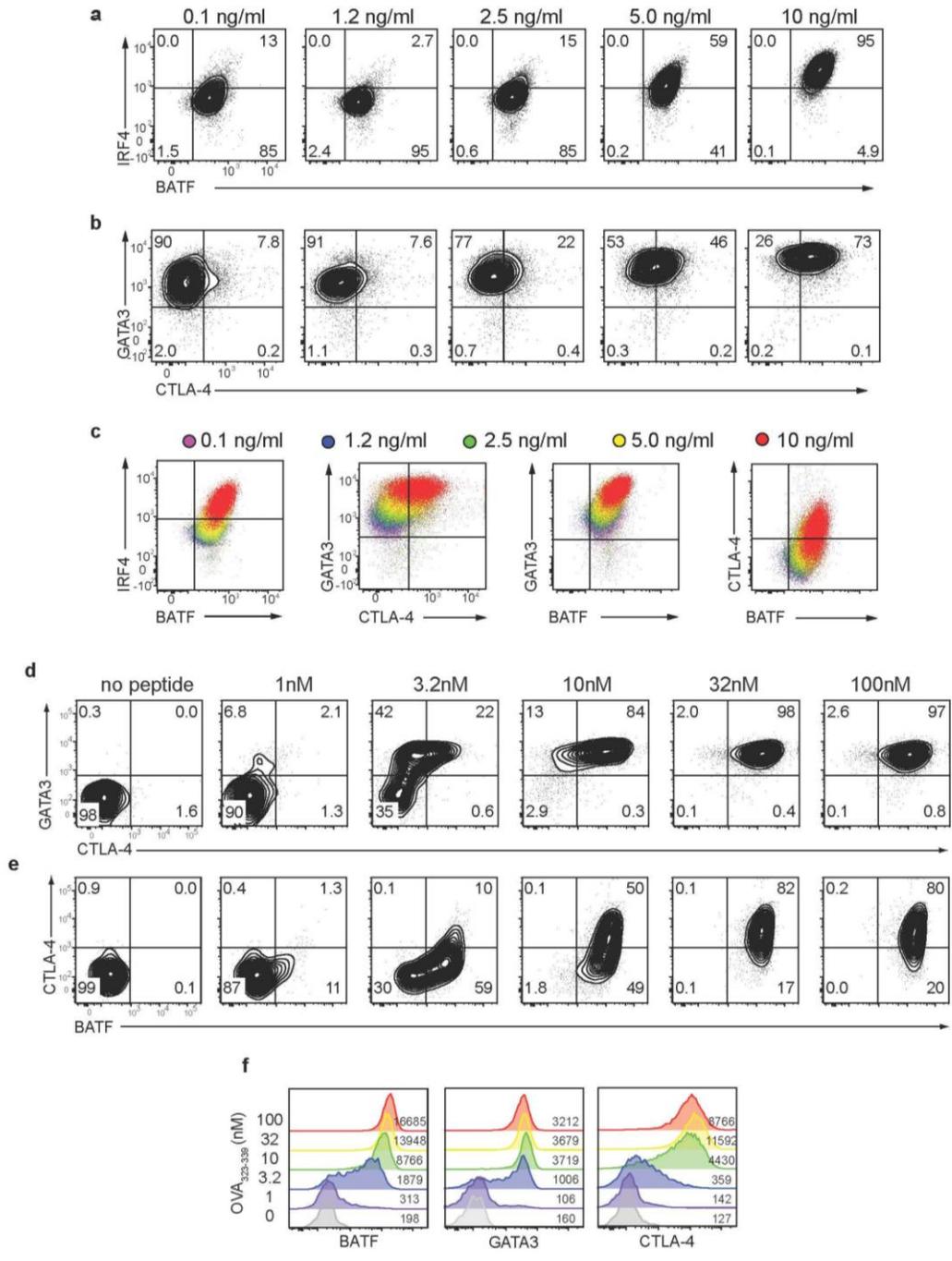
**Computational analysis of ChIP-exo analysis.** ChIP-exo data sets were aligned and masked as above. We kept duplicated tags and shrunk 50 bp of reads to first 1 bp of 5' position for further analysis. For visualizing, four replicates of BATF and two replicates of IRF4 were combined and normalized to 10 million tags. Bedgraphs were generated by 'genomeCoverageBed' of BedTools. ChIP-exo tag counts per base of motif  $\pm 50$  bp were measured by 'coverageBed' of BedTools and were visualized by R after normalization to

10 million tags per experiments. We chose putative narrow exonuclease stopped position (exo binding region) around motifs and non-binding region (control region) based on the mean ChIP-exo tag counts around consensus motifs and motifs predicted by ChIP-seq (Fig. 5d and Supplementary Fig. 6a). We applied 20 bp of exo-binding region and control regions for BATF and 10 bp for IRF4. We defined the motif sites with BATF and IRF4 ChIP-exo binding that satisfied the following criteria: threshold, total tag counts on target region of four experiments for BATF or two experiments for IRF4 were more than the threshold (Poisson *P* value of  $1 \times 10^{-6}$ ); significance of exo binding region versus control region, differential analysis between log<sub>2</sub>-transformed ChIP-exo tag counts on exo-binding region (eight regions for BATF; four regions for IRF4) and ChIP-exo tag counts on control regions (16 regions for BATF; 8 regions for IRF4) in each experiment were performed by Welch's *t*-test with Storey's correction<sup>52</sup> ( $P < 0.05$ ); and change in binding, the mean of tag counts on exo-binding region was more than twofold higher than the mean of tag counts on control region. Motif logos were generated by the 'seqLogo' package of R.

**Statistical analysis.** All statistical analyses were performed using Prism (GraphPad Software) or R.

**Data availability.** Data have been deposited in the GEO database: microarray data, GSE85173; ChIP-seq and ChIP-exo data, GSE85172.

41. Dignam, J.D., Lebovitz, R.M. & Roeder, R.G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475–1489 (1983).
42. Szabo, S.J. *et al.* Identification of cis-acting regulatory elements controlling interleukin-4 gene expression in T cells: roles for NF- $\kappa$ B and NF-ATc. *Mol. Cell. Biol.* **13**, 4793–4805 (1993).
43. Darrah, P.A. *et al.* Multifunctional TH1 cells define a correlate of vaccine-mediated protection against *Leishmania major*. *Nat. Med.* **13**, 843–850 (2007).
44. Shooshtari, P. *et al.* Correlation analysis of intracellular and secreted cytokines via the generalized integrated mean fluorescence intensity. *Cytometry A* **77**, 873–880 (2010).
45. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
46. Langmead, B. *et al.* Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
47. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
48. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
49. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
50. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
51. Serandour, A.A. *et al.* Development of an Illumina-based ChIP-exonuclease method provides insight into FoxA1-DNA binding properties. *Genome Biol.* **14**, R147 (2013).
52. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440–9445 (2003).

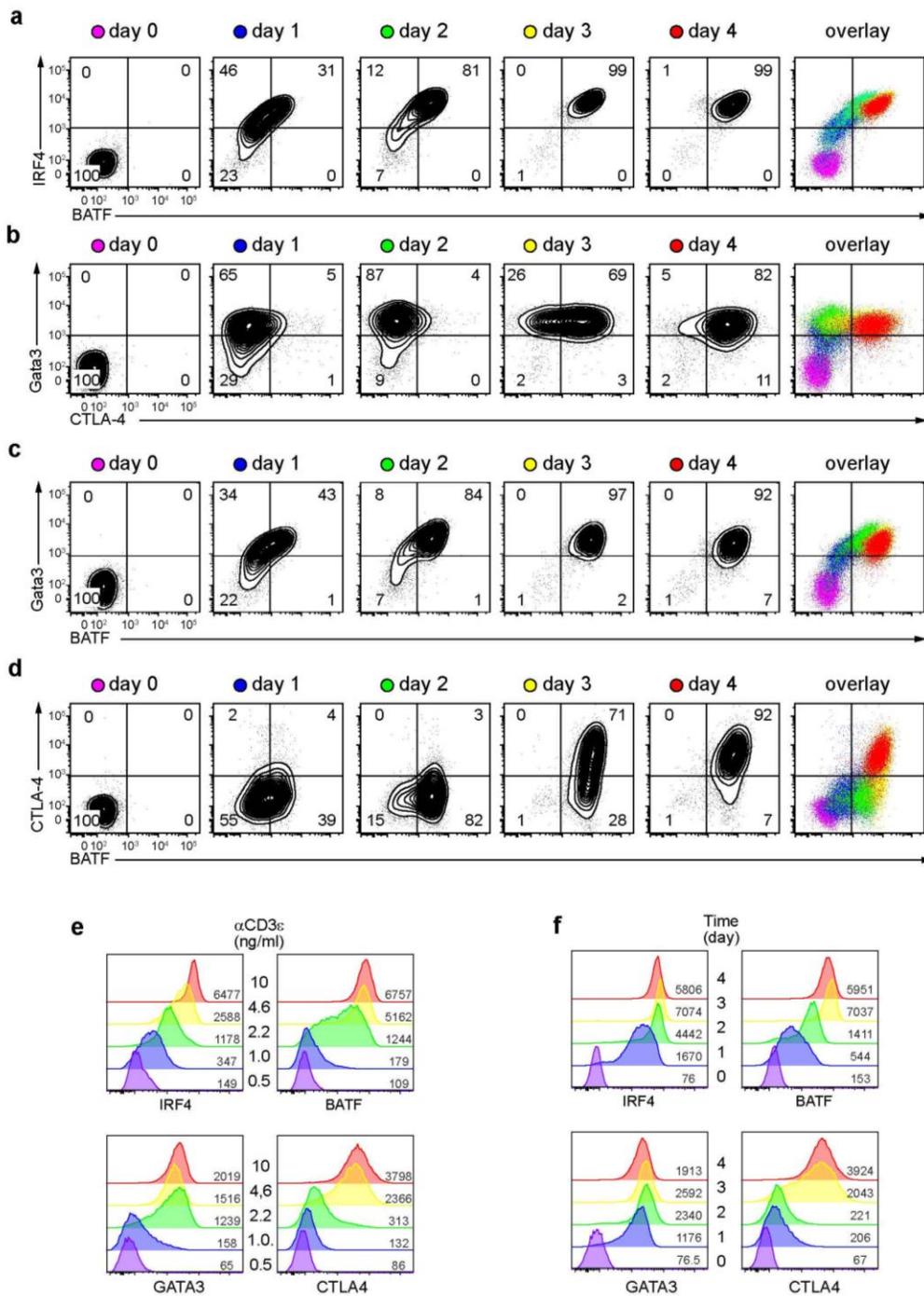


**Supplementary Figure 1**

GATA3 and CTLA-4 are differentially sensitive to graded expression of BATF and IRF4 in TH2 cells.

(a) Flow cytometry analyzing IRF4 and BATF expression on day 4 of secondary activation in WT CD4<sup>+</sup> T cells cultured under TH2 conditions (anti-IFN- $\gamma$ , anti-IL-12 and IL-4) with anti-CD28 and the indicated concentration of anti-CD3 $\epsilon$  crosslinked by plate-bound anti-hamster IgG. Numbers indicate the percentage of live CD4<sup>+</sup> cells

in each quadrant. **(b)** Flow cytometry analyzing GATA3 and CTLA-4 expression on day 4 in WT CD4<sup>+</sup> T cells cultured as in **(a)**. **(c)** Overlays of flow cytometry data from **(a)** and **(b)** showing expression of the indicated proteins at various doses of plate bound anti-CD3 $\epsilon$ . Data are representative of two biological replicates **(a-c)**. **(d,e)** Flow cytometry analyzing GATA3, CTLA4 and BATF on day 4 of primary activation of naïve KJ126<sup>+</sup> CD4 T cells from DO11.10 mice activated under T<sub>H</sub>2 conditions with the indicated concentration of ovalbumin (323-339) peptide and dendritic cells from Balb/c mice. Data are representative of two biological replicates. **(f)** Histograms and MFI for BATF, GATA3 and CTLA4 expression from **(d,e)**.

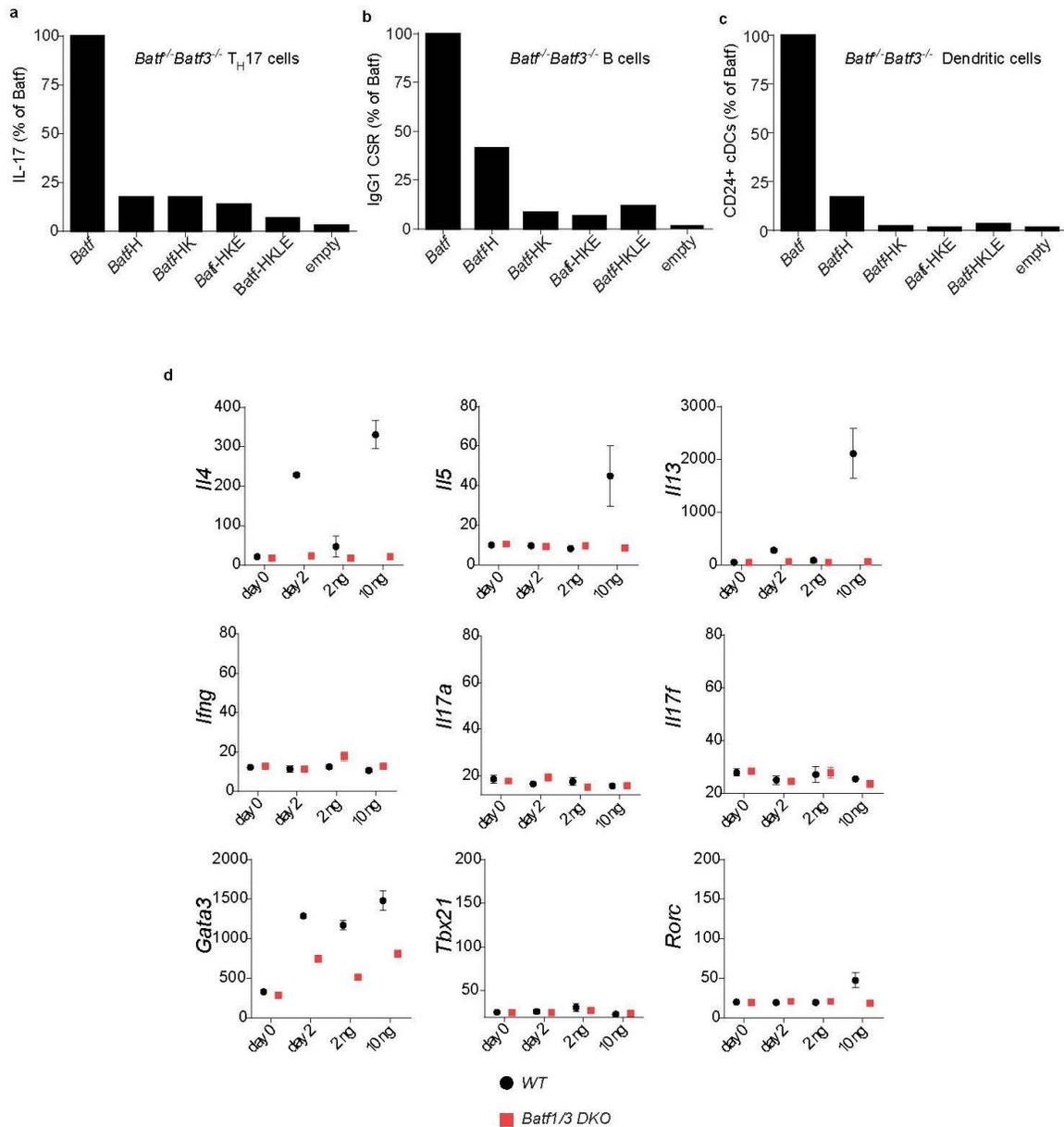


**Supplementary Figure 2**

Kinetics of the expression of BATF, IRF4, GATA3 and CTLA4 in  $T_H2$  cells.

(a) Flow cytometry analyzing IRF4 and BATF expression at the indicated time after primary activation of WT cells cultured under  $T_H2$  condition (anti-IFN- $\gamma$ , anti-IL-12 and IL-4) using 10 ng/ml of anti-CD3 $\epsilon$  and 4  $\mu$ g/ml

of anti-CD28 crosslinked with plate-bound anti-hamster IgG. **(b)** Flow cytometry analyzing CTLA-4 and GATA3 expression in T<sub>H</sub>2 cells cultured as in **(a)**. **(c)** Flow cytometry analyzing BATF and GATA3 expression in T<sub>H</sub>2 cells as in **(a)**. **(d)** Flow cytometry analyzing BATF and CTLA-4 expression in T<sub>H</sub>2 cells as in **(a)**. Data are representative of three biological replicates. **(e, f)** Histograms and MFI for factor expression from **(a-d)**.

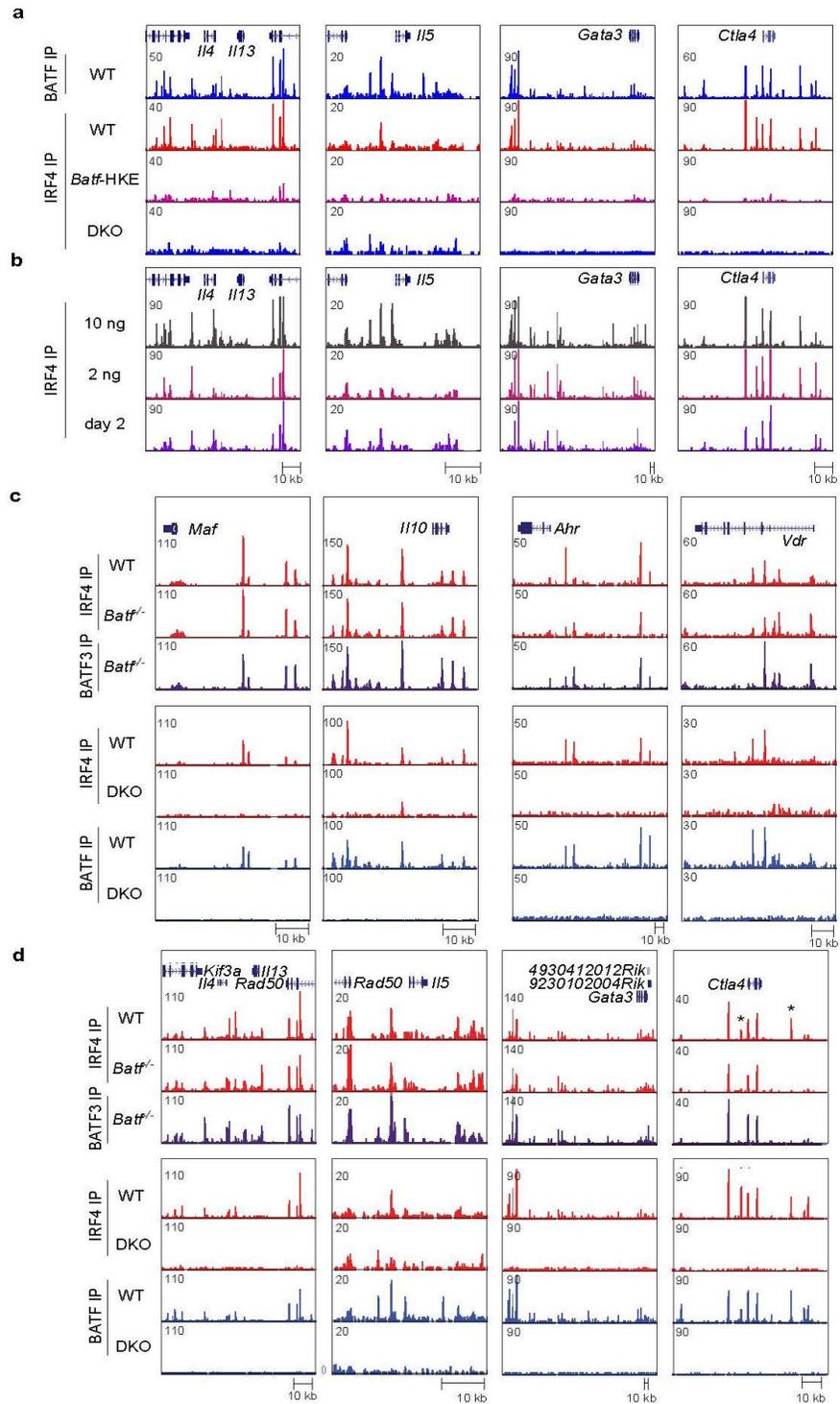


### Supplementary Figure 3

Analysis of *Batf* mutations and phenotype of *Batf1-Batf3* DKO cells.

(a) *Batf1/3* DKO CD4 T cells were cultured under T<sub>H</sub>17 conditions (anti-IFN- $\gamma$ , anti-IL-12, anti-IL-4, IL-6, TGF- $\beta$  and IL-1 $\beta$ ) on plate-bound anti-CD3 $\epsilon$  and anti-CD28 and were infected on day 1 of primary stimulation with empty retrovirus or retrovirus containing WT *Batf* or the indicated *Batf* mutant (*Batf*-H: *Batf* H55Q, *Batf*-HK: *Batf* H55Q K63D, *Batf*-HKE: *Batf* H55Q K63D E77K, *Batf*-HKLE: *Batf* H55Q K63D L56A E77K.) On day 5 after secondary, cells were stimulated with PMA/ionomycin and analyzed for IL-17 expression. Shown is

the activity of each mutant for reconstitution of IL-17 positive cells relative to WT *Batf* in the same experiment. Bars showed the mean of two experiments. **(b)** *Batf1/3* DKO B cells were cultured with LPS and IL-4, and infected on day 1 with empty retrovirus or retrovirus containing the indicated *Batf* mutant described in **(a)**. Class switch recombination to IgG1 was analyzed on day 4. Shown is the activity of each mutant for reconstitution of IgG1 positive cells relative to WT *Batf* in the same experiment. Bars showed the mean of two experiments. **(c)** *Batf1/3* DKO bone marrow cells were cultured with Flt3 ligand and infected on day 1 with empty retrovirus or retrovirus containing WT *Batf* or the indicated *Batf* mutant described in **(a)**. CD24<sup>+</sup> CD172a<sup>-</sup> CD11c<sup>+</sup> MHCII<sup>+</sup> B220<sup>-</sup> SiglecH<sup>-</sup> cells were analyzed on day 9. Shown is the activity of each mutant for reconstitution of CD24<sup>+</sup> DCs relative to WT *Batf* in the same experiment. Bars show the mean of three experiments. **(d)** Microarray expression of cytokines and transcription factors in WT and T<sub>H</sub>2 cells on day 0 or 2 days after primary activation with 10ng/ml anti-CD3ε (day2) or 4 days after primary activation with 2ng/ml or 10ng/ml anti-CD3ε as indicated. Shown is the mean and SEM for three biological replicates.

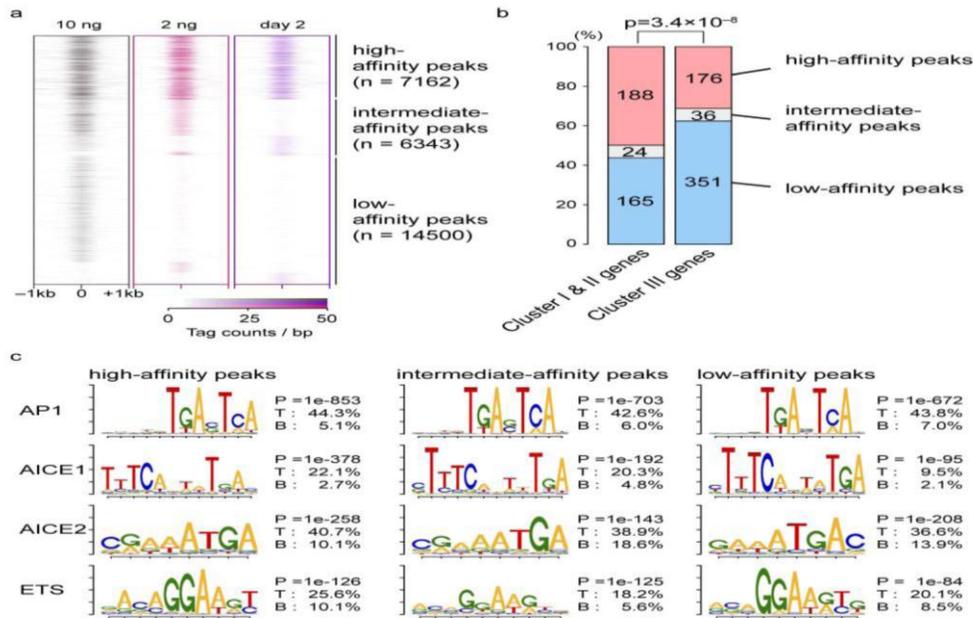


#### Supplementary Figure 4

T<sub>H</sub>2 cell-related genes have low sensitivity to the binding of BATF and IRF4.

(a) ChIP-seq for BATF in WT T<sub>H</sub>2 cells and IRF4 in WT and *Batf1/3* DKO T<sub>H</sub>2 cells (DKO) and in *Batf1/3*

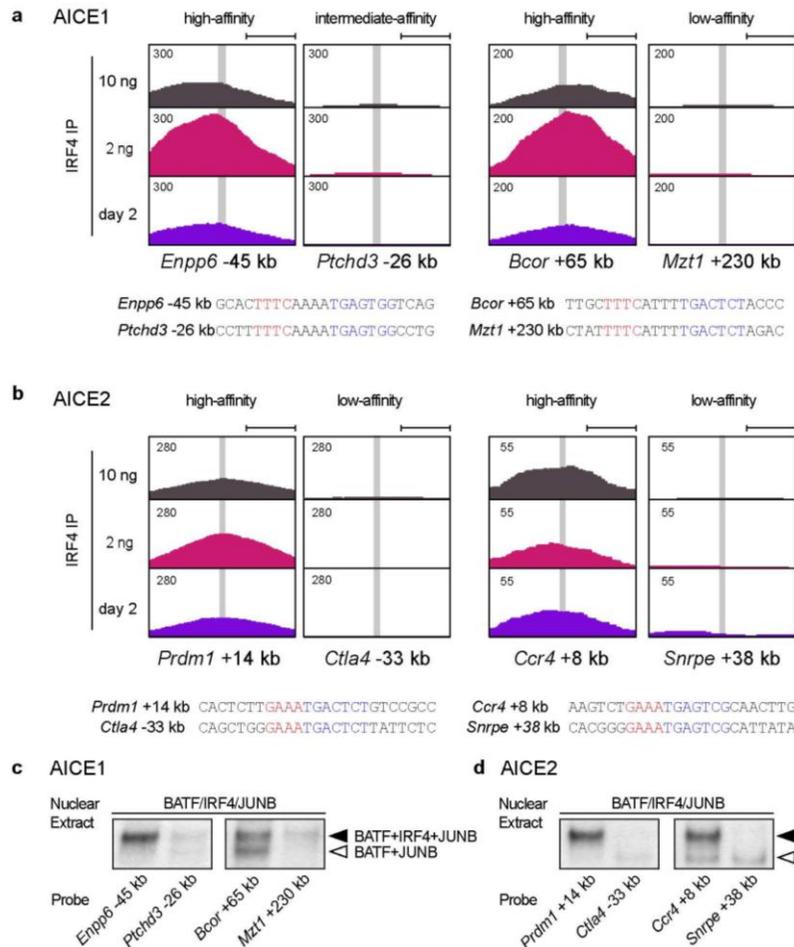
DKO T<sub>H</sub>2 cells reconstituted with retroviral *Batf*-HKE (*Batf*-HKE). Cells were prepared for ChIP after PMA/ionomycin stimulation on day 4 of secondary stimulation. **(b)** IRF4 ChIP-seq in WT T<sub>H</sub>2 cells cultured with 10 ng/ml of anti-CD3ε for 4 days, 2.2 ng/ml of anti-CD3ε for 4 days, or 10 ng/ml of anti-CD3ε for 2 days. **(c)** ChIP-seq for BATF3 in *Batf*<sup>-/-</sup> T<sub>H</sub>2 cells and IRF4 in WT and *Batf*<sup>-/-</sup> T<sub>H</sub>2 cells within genes from Cluster I/II and Cluster III. ChIP-seq was performed after PMA/ionomycin activation on day 4 of secondary stimulation and is compared with data from Figure 4a. **(d)** ChIP-seq for BATF3 in *Batf*<sup>-/-</sup> T<sub>H</sub>2 cells and IRF4 in WT and *Batf*<sup>-/-</sup> T<sub>H</sub>2 cells within T<sub>H</sub>2 related genes. ChIP-seq was performed after PMA/ionomycin activation on day 4 of secondary stimulation and is compared with data from Supplementary Figure 4a. \*Peaks that do not bind BATF3 in *Batf*<sup>-/-</sup> T<sub>H</sub>2 cells.



### Supplementary Figure 5

Affinity of IRF4 ChIP-seq peaks correlates with BATF-dependent gene expression, but peak affinity cannot be distinguished by *de novo* motif analysis.

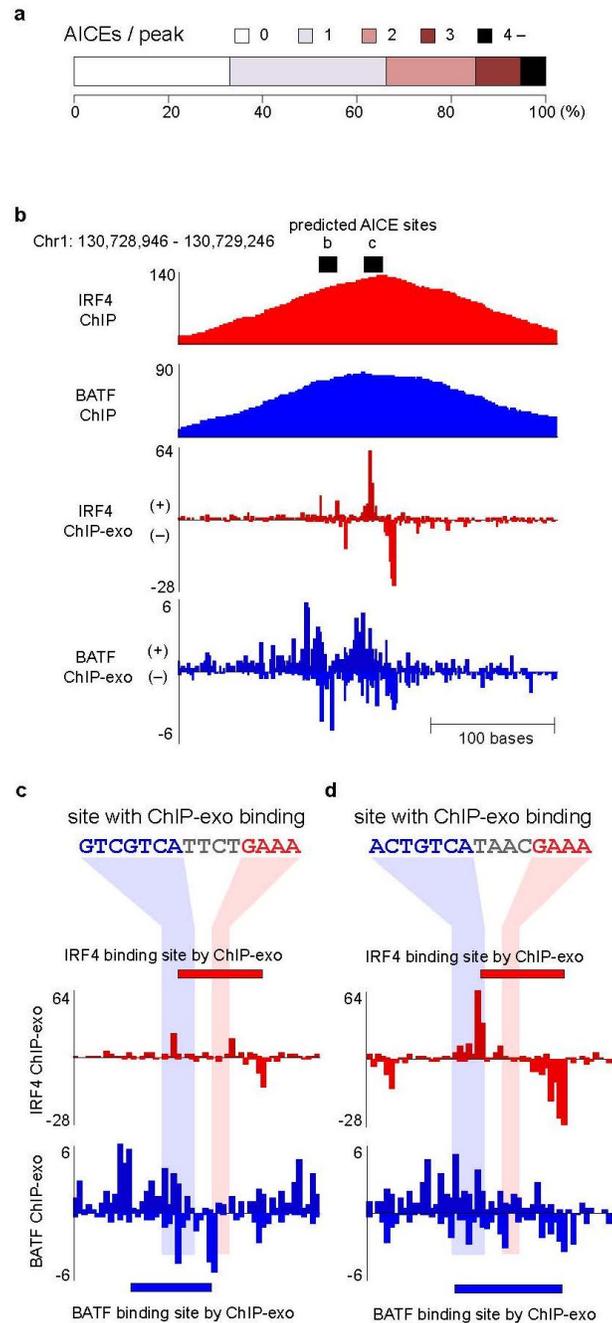
(a) Merged IRF4 ChIP-seq peaks from three conditions of primary stimulation were categorized as high-affinity (present in three conditions), intermediate-affinity (two conditions) and low-affinity (one condition). Color map shows the intensity of IRF4 binding centered on the peak  $\pm 1$  kb. (b) The number of high-affinity peaks, intermediate-affinity peaks and low-affinity peaks within  $\pm 50$  kb from TSS of Cluster I and II genes or Cluster III genes. Chi-square test,  $p = 3.417e-08$ , standardized residuals, high affinity peaks: 5.74, low sensitivity peaks: -5.61. (c) *De novo* motif analysis for the top 3000 IRF4 peaks in each category ranked by sum of tag counts of three conditions. P, final enrichment p-value; T, number of target sequences with motif as percent of total targets; B, number of background sequences with motif as percent of total background.



### Supplementary Figure 6

The flanking region regulates the affinity of a motif for the BATF–IRF4 complex.

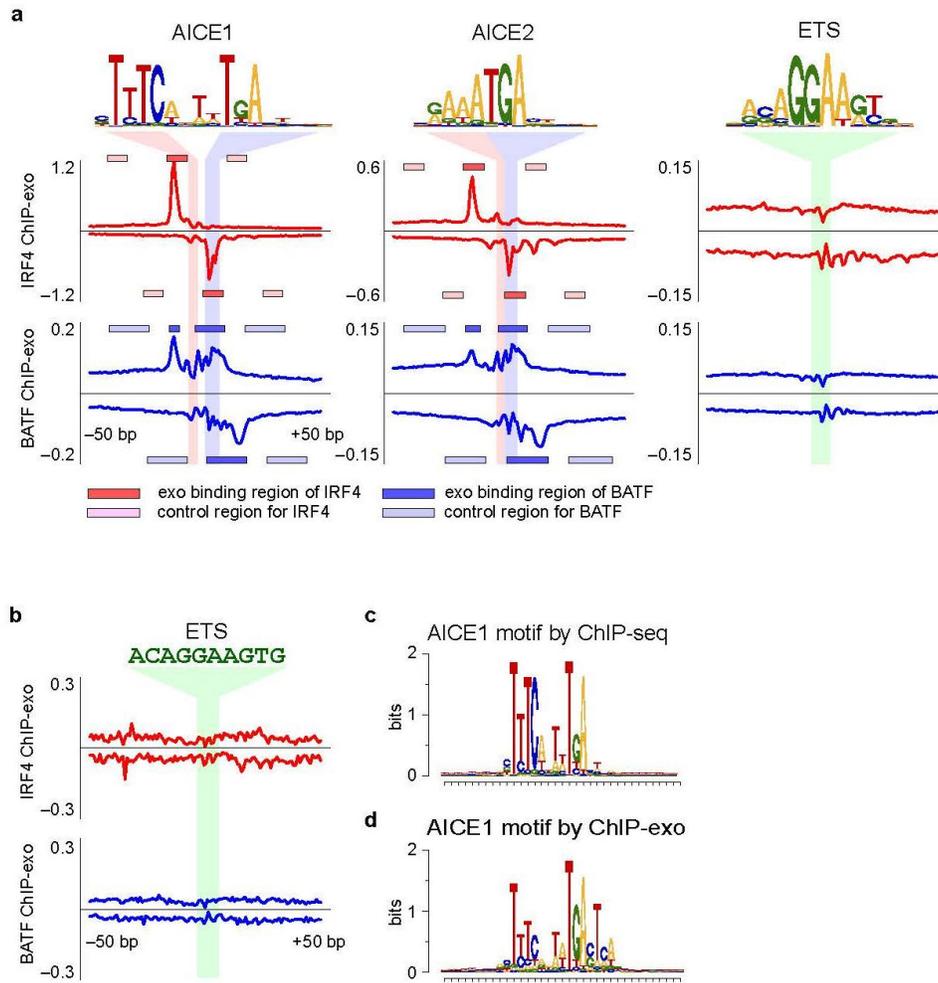
(a–d) ChIP-seq for IRF4 using the indicated primary activation condition of mouse  $T_H2$  cells. Shown are pairs of AICE1-containing peaks, *Enpp6* -45 kb and *Ptchd3* -26 kb, or *Bcor* +65 kb and *Mzt1* +230 kb with identical AICE1 sequences but different flanking regions (a) or pairs of AICE2-containing peaks, *Prdm1* +14 kb and *Ctla4* -33 kb, or *Ccr4* +8 kb and *Snrpe* +38 kb with identical AICE2 sequences but different flanking regions (b). (c) EMSA using nuclear extracts of HEK293FT cells expressing BATF, JUNB and IRF4 with probes based on sequences in (a). Open triangle: BATF/JUNB. Solid triangle: BATF/JUNB/IRF4. (d) EMSA using nuclear extract of HEK293FT cells expressing BATF, JUNB and IRF4 with probes based on sequences in (b). Data were representative of two experiments (c,d). Full sequence of probes are shown in Supplementary Table 2 (c,d).



**Supplementary Figure 7**

Multiple AICE motifs within single IRF4 ChIP-seq peaks.

. (a) Proportion of IRF4 peaks with multiple AICE motifs identified by *de novo* motif analysis. (b) ChIP-seq analysis and ChIP-exo analysis of BATF and IRF4 in WT T<sub>H</sub>2 cells. Black bars: predicted AICE sites. (c,d) Example of site from (a) with ChIP-exo binding.



### Supplementary Figure 8

AICE motif sites identified by *de novo* motif analysis show ChIP-exo binding, but ETS motif sites do not.

**(a)** Mean ChIP-exo tag counts of BATF and IRF4 binding on AICE1, AICE2 and ETS motifs within IRF4 ChIP-seq peaks that were identified by *de novo* motif analysis. **(b)** Mean ChIP-exo tag counts of BATF and IRF4 on consensus ETS motifs within IRF4 ChIP-seq peaks showing  $\pm 50$  bp flanking regions; red box: exo binding region of IRF4, light red box: control region for IRF4, blue box: exo binding region of BATF, light blue box: control region for BATF. **(c)** AICE1 motifs predicted by *de novo* motif analysis of IRF4 ChIP-seq. **(d)** AICE1 motifs identified from sites with higher ChIP-exo tag count on exo binding region of both BATF and IRF4 than control region.

## Supplementary Table 1

List of clustered genes.

Cluster I	Cluster II		Cluster III				Cluster IV	
Batf	Adamtsl3	Olf60	1600014C10Rik	Fam183b	Nedd4	Tnfrsf26	4930523C07Rik	Itih5
Cd24a	Ak4	Penk	4932438H23Rik	Fchsd2	n-R5s64	Tnfrsf8	4930562F07Rik	Kcnmb1
Ctla2a	Aldoc	Prdm1	9030617O03Rik	Fosl2	P4ha2	Tspan6	Add3	Klf3
Gp49a	Anxa2	Rbpj	AA467197	Frm4b	Padi2	Ubash3b	Al504432	Klk1b22
Ifitm3	Anxa5	Rnf19b	Abi3	Galc	Parp16	Ube2d2b	Ankrd55	Lax1
Il10	Asb2	Rps6ka5	Adamtsl4	Gatm	Paxbp1	Vdr	Arhgap20	Lin7a
Il21	Atxn711	S100a6	Adarb1	Gatsl3	Pde1a	Vldlr	B630019A10Rik	Marcksl1
Il24	B3galt2	Scd1	Afg3l2	Gbp2	Phxr4	Whsc111	Bcl2a1b	Mir363
Inhba	Bhlhe40	Scd2	Ahr	Gbp2b	Pik3ap1	Zak	Bcl2a1c	Ms4a4b
Lilrb4	Ccr2	Selp	Ajuba	Gem	Pkp2	Zc3h12c	Bzw2	Ms4a4c
Ly6a	Ccr4	Slc2a3	Akr1c18	Gimap7	Plcd1	Zcchc24	Ccl22	Ms4a6b
Mt2	Ccr8	Snord118	Anxa4	Gipr	Plin2		Ccr9	Ms4a6c
Nts	Cdkn1a	Snx9	Appl2	Gja1	Pparg		Cd226	Nipal1
	Crip1	St3gal1	Armxc3	Gm24622	Ppp1r3b		Cd44	Nod1
	Cyp11a1	Stom	Arrb1	Gpd2	Ppp2r3a		Cd69	Nr4a3
	Dusp16	Tanc2	Atf3	Gpnmb	Prnp		Cd74	Nrp1
	Dusp6	Tigit	Atp2b4	Gpr65	Pth		Cd79b	Nsg2
	Egln3	Trpm6	Atp6v0d2	Grina	Ptpn4		Cd83	Ostn
	Epas1	Ttc39c	B430306N03Rik	Gsn	Ptpnj		Cd9	Parp8
	Etv6	Vim	Basp1	H2-Q10	Rab33a		Cd96	Pdlim1
	Fads2		Batf3	Hao	Rabgap11		Cep85l	Phtf2
	Fam46a		Bcl3	Hamp	Rapgef5		Chn2	Ptger4
	Focad		Casp6	Hif1a	Rbp1		Clec2i	Pyhin1
	Gcnt1		Cd200r1	Hipk2	Rhoq		Cmah	Qser1
	Glrx		Cd63	Ifitm6	Rnf128		Csf1	Rel
	Gm14005		Cd93	Il12rb2	Rnf208		Csn1s2a	Serpinb6b
	Gpm6b		Cdh17	Il13	Sccpdh		Csn2	Slc14a1
	Gpr174		Cpd	Il18r1	Sdc3		Cth	Slc9a7
	Gpr68		Creb3l2	Il1r2	Sdcbp2		Ctss	Slfm1
	Gstk1		Crem	Il4	Selenbp1		Dapl1	Slfm3
	Gzmb		Ctla2b	Il5	Selenbp2		Dhrs3	Smpdl3a
	H2-Ab1		Ctla4	Inha	Selm		Dock10	Spr2a2
	Hgf		Cxcr2	Itga3	Sema4f		Emp3	St8sia6
	Hsd17b7		Cyp2s1	Itga7	Serpinf1		Fam129a	Tbc1d4
	Htr1b		Cysltr1	Jup	Sft2d2		Fcrls	Tcf7
	Ier3		D10Bwg1379e	Kif3a	Slc17a6		Foxp1	Tdgf1
	Igfbp4		Dennd3	Klhl6	Smim3		Foxp3	Tdgf1-ps1
	Ighe		Dgat1	Lama5	Smox		Gen1	Tespa1
	Il12rb1		Dnm1	Lamc1	Socs3		Gm10021	Tex35
	Il1r1l		Dntt	Lpxn	Stfa2		Gm10838	Tlr7
	Itpr1		Dock5	Ly75	Stfa2l1		Gm8995	Tnfrsf13c
	Klf6		Ecm1	Maged1	Stra6		Gm9900	Tnfsf10
	Lif		Egfr	Map3k5	Sval1		Gpr183	Tnfsf11
	Maf		Emp1	Mboat2	Tcp11l2		H60b	Tnfsf9
	Mfsd2a		Enah	Mir22	Tec		Hdc	Tox
	Moxd1		Ernm	Mir22hg	Tgfb3		Hist1h1a	Trat1
	Nckap1		F730043M19Rik	Mt1	Tgm1		Ifi203	Trim12c
	Ndr1		Fam114a1	Myo10	Tgm2		Ilgp1	Txk
	Nfil3		Fam115c	Nabp1	Timp1		Ipcef1	Utf1
	n-R5s54		Fam132a	Ncs1	Tmem140		Isg20	

## Supplementary Table 2

### Sequences of EMSA probes

Probe	Sequence
<i>Enpp6</i> -45 kb	ACCGCAC <u>TTT</u> CAAAA <u>TGAGTGG</u> TCAGAAT
<i>Ptchd3</i> -26 kb	TTTCCTT <u>TTT</u> CAAAA <u>TGAGTGG</u> CCTGCAA
<i>Bcor</i> +65 kb	GTGTTTGC <u>TTT</u> CATTT <u>TGACTCT</u> ACCCTCAA
<i>Mzt1</i> +23 kb	CTCCCTAT <u>TTT</u> CATTT <u>TGACTCT</u> AGACATCT
<i>Ccr4</i> +8 kb	CTGAAAGTCT <u>GAAATGAGTCG</u> CAACTTGCGA
<i>Snrpe</i> +38 kb	AAATCACGGG <u>GAAATGAGTCG</u> CATTATATGT
<i>Prdm1</i> +14 kb	TCACTCTT <u>GAAATGACTCT</u> GTCCGCCT
<i>Ctla4</i> -33 kb	ACAGCTGG <u>GAAATGACTCT</u> TATTCTCA
C to T <i>Ctla4</i> -33 kb	ACAG <u>T</u> GG <u>GAAATGACTCT</u> TATTCTCA
<i>Ctla4</i> +30 kb	TGACTTGT <u>GAAATGAGTC</u> CAGAGTCATTC
T to C <i>Ctla4</i> +30 kb	TGAC <u>C</u> TGT <u>GAAATGAGTC</u> CAGAGTCATTC
<i>Bcl11b</i> +30 kb	TGATAGTGCA <u>GAAATGAGTC</u> CAGAGTCAAAG
T to C <i>Bcl11b</i> +30 kb	TGATAG <u>C</u> GCA <u>GAAATGAGTC</u> CAGAGTCAAAG
rs231735-G	TCATCCTGGGTCT <u>GATATGAGTGA</u> AGTCCACCCT
rs231735-T	TCATCCTGG <u>T</u> TCT <u>GATATGAGTGA</u> AGTCCACCCT

Red indicates IRF4 binding motif, blue indicates BATF binding motif, and mutated nucleotide is indicated with underline.

## Supplementary Figure 3

### Sequences of enhancer oligos

Probe	Sequence
rs2317354 no AICE forward	AGCTTAACTCATCCTGG <u>C</u> TCTG <u>T</u> TGCAATACAAGTCCACCCTG
rs2317354 no AICE reverse	GATCCAGGGTGGACTT <u>G</u> TATTGCAACAGAGCCAGGATGAGTTA
rs2317354-G forward	AGCTTAACTCATCCTGGGTCT <u>GATATGAGTGA</u> AGTCCACCCTG
rs2317354-G reverse	GATCCAGGGTGGACT <u>TCACTCA</u> <u>TATC</u> AGACCAGGATGAGTTA
rs2317354-T forward	AGCTTAACTCATCCTGG <u>T</u> TCT <u>GATATGAGTGA</u> AGTCCACCCTG
rs2317354-T reverse	GATCCAGGGTGGACT <u>TCACTCA</u> <u>TATC</u> AGAA <u>C</u> AGGATGAGTTA

Red indicates IRF4 binding motif, blue indicates BATF binding motif, and mutated nucleotides and SNP are indicated with underline.