A new pipeline SPICE identifies novel JUN-IKZF1 composite elements

Peng Li[*,#], Sree H. Pulugulla,[*], Sonali Das, Jangsuk Oh, Rosanne Spolski, Jian-Xin Lin, Warren J. Leonard[#]

Laboratory of Molecular Immunology

Immunology Center

National Heart, Lung and Blood Institute

National Institutes of Health

Bethesda, MD 20892

[*]These authors contributed equally.

[#] Correspondence to: P.L. (lip3@nhbli.nih.gov) or W.J.L. (leonardw@nhlbi.nih.gov)

Abstract

Transcription factor partners can cooperatively bind to DNA composite elements to augment gene transcription. Here, we report a novel protein-DNA binding screening pipeline, termed Spacing Preference Identification of Composite Elements (SPICE), that can systematically predict protein binding partners and DNA motif spacing preferences. SPICE de novo predicted known composite elements, including AP1-IRF composite elements (AICE) and STAT5 tetramers, and also predicted a range of novel binding partners, including JUN-IKZF1 composite elements, including at an upstream conserved noncoding region, CNS9, in the human *IL10* gene, which contains a non-canonical IKZF1 site. We confirmed cooperative binding of JUN and IKZF1 and showed that the activity of an *IL10*-luciferase reporter construct in primary B and T cells depended on both this site and the AP1 binding site within this composite element. Overall, our findings reveal an unappreciated global association of IKZF1 and AP1 and establish SPICE as a valuable new pipeline for predicting novel transcription binding complexes.

Introduction

Genome-wide studies investigating cooperative interactions of transcription factors (TFs), histone modifications, and higher-order chromatin conformation are critical for the systematic evaluation of mechanisms of gene expression. Chromatin immunoprecipitation coupled with DNA sequencing (ChIP-Seq) has been used to map genome-wide DNA binding proteins of interest, thereby enhancing our understanding of genomics and epigenomics (1, 2), with important discoveries related to disease-associated transcriptional regulation (3, 4), tissue-specificity of epigenetic regulation (5, 6), and chromatin organization (7). To extract biologically relevant information from ChIP-Seq data, many computational tools, such as Model-based Analysis for ChIP-Seq (MACS)(8), MUltiScale enrIchment Calling for ChIP-Seq (MUSIC)(9), and Zero-Inflated Negative Binomial Algorithm (ZINBA)(10), have been developed to understand transcription factor cooperation and interactions. Genomic regions that are significantly enriched represent candidates for protein/DNA-binding sites that can be used to predict binding motifs (11), and ChIP-Seq analysis can be integrated with other types of genomic assays, including RNA-Seq gene expression profiling and long-distance chromatin interaction analyses to elucidate mechanisms of genomic function (12, 13) and gene ontology (14). Moreover, given a set of identified ChIP-Seq peaks, motif enrichment analysis tools such as MEME (11, 15) can de novo discover the motif(s) corresponding to the TF's DNA binding domain. When binding to genomic DNA, some TFs may cooperatively bind and physically interact with specific partner TFs, and there often is preferred spacing between their binding sites (12, 16-19). These cooperative interactions at composite elements are critical for transcriptional regulation of genes and are used by cells to integrate diverse signals and potently drive transcription even at low concentrations of transcription factors. However, few computational

tools to date can predict composite elements and their optimal spacing using transcription factor ChIP-Seq data. SpaMo (space motif analysis) in MEME suite (15) can infer interactions between a pair of specific TFs at near sites on the DNA sequence, however, it cannot systematically predict TF-TF composite elements genome wide. Here, we report a new computational pipeline denoted "Spacing Preference Identification of Composite Elements" (SPICE), which can predict in a genome-wide fashion novel composite elements for pairs of transcription factors and the spacing between the elements. SPICE is versatile and efficient and can use any ChIP-Seq dataset to make predictions that can be tested to confirm cooperative binding and corresponding effects on gene expression. We found SPICE predictions to be highly accurate in predicting previously established binding partners/compositve motifs, and in this study we demonstrate genome-wide JUN-IKZF1 interactions. SPICE therefore has potentially broad utility and provides a powerful method to identify a range of new transcription-factor interactions and mechanisms underlying gene regulation.

Results and Discussion

The SPICE pipeline predicts known and novel protein complexes

The SPICE schematic pipeline is summarized in **Figure 1A** and detailed in the Methods. First, we aligned the sequenced reads from ChIP-Seq datasets to the reference genome and identified significant TF binding sites (peaks) using MACS. Second, we performed de novo peak analysis, designated the top-enriched motif among the identified peaks as the primary motif, kept peaks containing the primary motif, and eliminated sequences without primary matches. Third, we loaded secondary motifs using motif databases (e.g., HOCOMOCO, etc.), centered the primary motifs and scanned for secondary motifs within 500 bp of the primary motif, specifically

checking for enrichment of known TF DNA-binding motifs. Finally, we sorted the significant secondary motifs and spacing, and grouped those that were redundant. We then visualized their interaction matrices with the primary and secondary motifs as a heat map (see illustrative schematic in **Figure 1B**, where each red square indicates a predicted composite element containing the motifs at the corresponding primary and secondary motifs on the x- and y-axis, respectively). Based on the distance from the primary to secondary motif, we displayed the spacing distributions as bar graphs highlighting the most significant spacing preferences (**Figure 1C**). To validate the predictive accuracy of SPICE, we used our previously published IRF4 ChIP-Seq data in mouse pre-activated T cells to see if we could de novo re-discover AP-1/IRF4 composite elements (AICE, 5′ -TGAnTCA/GAAA-3′)(20). Indeed, SPICE successfully predicted these elements, including optimal spacing between AP1 and IRF4 as 0 or 4 bp (**Figure 1D; Figures S1A and S1B**), as previously reported (20), thus validating the pipeline. Furthermore, SPICE successfully predicted STAT5 tetramerization with an optimal spacing of 11-12 bps (**Figure S1C, D**), as previously reported (17). STAT1, STAT3, and STAT4 also have been reported to form tetramers (21-23), and indeed SPICE de novo predicted that these STATs also can form tetramers, whereas STAT2 was not predicted to form tetramers **(Table 1**), consistent with the lack of reports of such formation.

SPICE can predict known and novel composite elements from the human Encode database

To assess SPICE's ability to predict novel TF binding complexes, we utilized ChIP-Seq data from the ENCODE project, as standardized by the ENCODE Analysis Working Group; this allowed us to perform integrated analysis of all ENCODE data types based on uniform processing. We evaluated the performance of SPICE using ENCODE ChIP-Seq libraries from

Transcription Factor Binding Sites (TFBS) generated by ENCODE/Stanford/Yale/USC/Harvard (SYUH), which comprise a total of 343 libraries in 20 different cell lines, with K562 human erythroleukemia, GM12878 EBV-transformed B cell, and HeLa S3 human cervical carcinoma cell lines being heavily used to generate these libraries. For each selected TF ChIP-Seq library that contains a primary motif (**Figure S2**, x-axis), we extracted 500 bp of DNA sequence centered on the primary motif and searched for 401 known human transcription factor binding motifs using the Homo sapiens Comprehensive Model Collection database (HOCOMOCO, v11)(**Figure S2,** y-axis), and generated a 343 x 401 spatial interaction matrix, in which each cell was assigned an E-value indicating the significance of primary-secondary motif pairs (**Table 2**). The E-values were then log transformed (-log10(E-value)) and shown as an interaction heat map (**Figure S2**). As is evident, most of the motif pairs were not significant, suggesting that transcription factor composite elements are relatively rare events; in addition, some interactions that exist in some cell types may not be observed if they are not present in the cell lines that were used to generate the ENCODE datasets. After applying a filtering criterion that only included motif pairs with an E-value < $1e^{-10}$, the resulting interaction matrix had dimensions of 118 x 205 and exhibited reduced spatial complexity. Very interestingly, we could immediately identify multiple potential TF-TF composite elements (**Figure 1E**), which include putative associations of JUN with STAT3 and IKZF1, of CEBPβ with E2F, KLF, and BACH family members, and a range of others including ZNF241/NF-κB, GATA1/NDF1, and CTCF/ETS. Importantly, association of JUN with STAT3 has been known (24) and a functional CTCF-ETS1 interaction was recently reported (25), further validating the predictive power of SPICE. Regarding the CTCF-ETS association, SPICE successfully defined an optimal spacing of 8 bp for CTCF/ETS

complexes (**Figures S1E and S1F**), something that was not defined in the earlier study (25) and thereby further underscoring the robustness of our pipeline.

Most IKZF1 binding sites co-localize with JUN

We decided to focus on the putative JUN-IKZF1 given that both JUN and IKZF1 play critical roles in immune cells but have not been reported to physically associate or functionally cooperate. IKZF1 (Ikaros) has crucial functions in the hematopoietic system (26, 27) and is a regulator of immune cell development in early B cells and T cells (28, 29). JUN family factors are part of AP-1 complexes, forming homodimers or hetero-dimers with FOS family basic leucine zipper proteins and playing prominent roles in T cell biology (30, 31). By ChIP-Seq, IKZF1 and JUN extensively co-localized in the human cell line GM12878 (**Figure S3A)**, for example at multiple sites in the *TNFRSF8* and *IL10* loci. The *IL10* locus included a highly conserved CNS9 region (highlighted in a red box and labeled in **Figure S3B**), located approximately 9.1 kb upstream of the *IL10* promoter. Pairwise sequence alignment confirms the CNS9 region is highly conserved between human and mouse (**Figure S3C**). BATF, a FOS family member, also bound at these loci (**Figure S3B)**, consistent with BATF's ability to partner with JUN in AICEs (20). To further characterize IKZF1-JUN complexes and validate their existence in vitro and in vivo, we performed ChIP-Seq experiments with antibodies to IKZF1 and JUN in primary mouse B cells that were either untreated or treated with LPS or LPS + IL-21 for 3 hours. We used B cells given the importance of IKZF1 for B cell development (32). When cells were stimulated with LPS + IL-21, we identified a total of 9197 and 14433 binding sites for IKZF1 and JUN, respectively (**Figure 2A**). The de novo motif analysis (HOMER) on the IKZF1 peaks (**Figure 2B**) indicated the most enriched motifs include ETS-like motif (ranks 1st) and

EBF1-like motif (ranks 2nd), which are comparable to the previously identified IKZF1 motifs (33), as well as an IRF-like motif (ranks 3rd), etc., all of which ostensibly could be utilized by IKZF1 to bind to DNA. Strikingly, 8125 out of 9197 (88.3%) of IKZF1 peaks co-localized with JUN (**Figure 2C**), and the Reactome Pathway Analysis from the 8125 co-localized peaks suggested that the IKZF1-JUN complex might play broad biological roles as there were multiple enriched pathways, including "Cytokine Signaling in Immune System", "Signaling by Interleukins", TLR4-related cascades, and more. (**Figure 2D** and **Figure S4**). IKZF1 binding was induced when cells were treated with LPS or LPS + IL-21, while JUN binding was prominent only with LPS + IL-21 treatment. IKZF1 sites globally co-localized with JUN (**Figure 2E**), including at known regulatory elements, for example at the *Prdm1* and *Il10* genes (**Figure 2F**)(20), which are critical for plasma cell differentiation in IL-10-producing B cells (34). Importantly, the CNS9 region was identified not only in the human *IL10* gene (**Figure S3B**) but also in the mouse *Il10* gene (**Figure 2F**), suggesting that this conserved region might be biologically important and dependent on the binding of both IKZF1 and JUN.

Cooperative binding of IKZF1 and JUN at the *IL10* upstream region

To identify potential physical interactions between IKZF1 and JUN, we performed co-immunoprecipitation experiments in the MINO cell line, a human mantle cell lymphoma B cell line. We confirmed that both JUN and IKZF1 were expressed in these cells, as assessed by western blotting (**Figure 3A**). Although we could not co-immunoprecipiate IKZF1 with anti-JUN, JUN was weakly co-precipitated by anti-IKZF1, suggesting these factors might form a complex (**Figure 3A**), consistent with their binding in proximity in a genome-wide fashion as shown in **Figure 2C and 2E**. The gene encoding IL-10 in both humans and mice is expressed in

both T cells and B cells (35) and contains a canonical AP-1 motif and a nearby sequence (GTTGCAGTTTC) that is an 11 bp match for the variant 3rd IKZF1 motif peak in **Figure 2B.**. These motifs were located in the CNS9 region (~9.1 kb upstream of the *IL10* promoter) (**Figures 2F and S3B**), which is a highly conserved non-coding region in the mouse and human genes (**Figure S3C**) and is critical for regulation of the *IL10* gene (20, 36). We wished to assess whether this might be a variant IKZF1 motif and performed EMSAs using wild-type or mutant probes corresponding to CNS9 (**Figure 3B**; the GTTGCAGTTTC motif 3 region and AP1 motifs are underlined) and MINO nuclear extracts. A strong complex formed with the WT CNS9 probe (**Figure 3C**, lane 2), but the intensity was weaker when the variant IKZF1 sequence was mutated (lane 4) and abolished when the AP-1 motif (lane 6) or both motifs (lane 8) were mutated (**Figures 3C and 3D**). We next evaluated the presence of IKZF1 and JUN by super-shifting with antibodies. As expected, a strong complex was seen only in the presence of nuclear extract (**Figure 3E**, lane 2 vs. 1). Neither mouse nor rabbit IgG affected the complex (lanes 3 and 4), but anti-IKZF1 consistently altered the appearance of the shift and anti-JUN gave a discrete slower mobility super-shifted band, consistent with both IKZF1 and JUN being part of the complex (lanes 5 and 6).

Both IKZF1 and JUN are critical for cooperative binding and transcriptional activation

We next performed EMSAs using the WT IR700 probe and 25, 50, 100, and 200 molar excess of cold competitor oligonucleotides corresponding to the WT or the variant IKZF1 mutant, AP-1 mutant, or double mutant CNS9 probes. The WT competitor dramatically reduced the binding activity and the variant IKZF1 mutant competitor moderately reduced the intensity of the shift, whereas AP-1 mutant or IKZF1/AP-1 double mutant oligonucleotides had little if any effect

(**Figures 4A and 4B**). To determine whether IKZF1 and AP-1 proteins cooperatively bound to the CNS9 region, we used recombinant IKZF1 and cJUN/cFOS AP1 protein to perform EMSAs with wild-type and mutant IR700 probes corresponding to the *IL10* CNS9 site (data shown in **Figure 4C** and quantified in **4D**). With the wild-type probe, no binding activity was observed with IKZF1 alone (**Figures 4C and 4D, lane 2**), but binding occurred with AP-1 (**lane 3**), and stronger binding occurred when both IKZF1 and AP1 proteins were added (**lane 4**), suggesting cooperative binding of these factors. Binding was not detected to the AP-1 mutant probe (**lane 8**) and was lower with the variant IKZF1 mutant probe (**lane 6**), suggesting that the AP-1 site is absolutely required for the cooperative binding and that the putative variant IKZF1 site had low affinity in its own right but cooperated with the AP1 site to allow cooperative binding of the factors. Correspondingly, activity of the WT *IL10* luciferase reporter was potently induced by LPS in primary B cells compared to vector but not when IKZF1, AP-1, or both motifs were mutated (**Figure 4E**), suggesting both IKZF1 and AP-1 are important for transcriptional activation of this gene. Besides B cells, IKZF1 has been shown to be a regulator of *Il10* gene expression in mouse T cells, in which *Il10* expression was significantly lower compared with wild-type T cells in response to TCR stimulation or during Th2 differentiation in *Ikzf1*$^{-/-}$ cells (37). We therefore also analyzed activity in T cells. *IL10* reporter activity was potently induced, but expression was diminished when the variant IKZF1 motif, or particularly the AP1 motif or both motifs were mutated (**Figure 4F**), demonstrating cooperative binding and transcriptional activation by IKZF1 and AP1 proteins in this setting as well. These results provide mechanistic insights into the actions of IKZF1 and underscore the power of the SPICE pipeline for identifying new transcription factor cooperative effects.

We have developed a computational pipeline, SPICE, that can predict novel transcription factor binding partners and optimal DNA spacing of motifs. In this report, we analyzed the ENCODE database using the SPICE pipeline, successfully predicting a range of known interactions as well as a multitude of novel interactions. Here, we focused on JUN-IKZF composite elements, an association not previously demonstrated. ChIP-Seq analysis demonstrated that IKZF1 co-localizes with JUN in a genome-wide manner, and that the major two motifs identified were canonical IKZF1 motifs, but that non-canonical variant motifs were also identified, including at the *IL10* gene locus. We could coprecipitate JUN with anti-IKZF1, and moreover, we found functional cooperation between IKZF1 and AP1 motifs. Specifically, we found that IKZF1 and JUN not only cooperatively bound to the CNS9 region in the *IL10* locus but that both IKZF1 and AP1 motifs were critical for transcriptional activity of a reporter construct in both B and T cells. Although we focused on the *IL10* locus, we demonstrated genome-wide co-localization of these factors, and reactome pathway analysis linked the genes in which they were found to a range of pathways including but not limited to cytokine signaling pathways (**Figure 2**); thus, the physiological relevance of cooperation of these transcription factors may be broad, potentially including in the context of cellular function and diseases. The identification of JUN-IKZF1complexes demonstrates that our computational approach and novel pipeline, SPICE, can predict novel transcription factors binding complexes and their optimal spacing, providing a valuable new way to discover composite elements that contribute to a highly specific pattern of gene transcriptional regulation and to better understand gene regulatory networks in the immune response.

Acknowledgements:

This work was supported by the Division of Intramural Research, NHLBI. Next generation sequencing was performed at the NHLBI Sequencing Core.

Figure Legends

Figure 1. Spacing Preference Identification of Composite Elements (SPICE) predicts known and novel composite elements and spacing preferences. (**A**) Schematic step-by-step flowchart of the SPICE pipeline to predict transcription factor composite elements and their spacing preferences (see Methods for details). (**B**) Schematic of a heat map depicting hypothetically enriched significant motif pairs (shown as solid red boxes). This interaction matrix of primary-secondary motifs depicts the types of data that would be generated as potential composite elements after a ChIP-Seq derived primary motif region is compared with the motif database for secondary motifs based on distribution of spacing between primary and secondary motifs. The red boxes indicate composite elements based on the the primary and secondary motifs on the x- and y-axis, respectively. (**C**) Schematic bar graphs showing spacing distribution of secondary motifs centered on a given primary motif, with bars highlighted in red indicating the most preferred spacing. (**D**) SPICE successfully predicts and validates known composite elements. Shown is validation of SPICE's ability to rediscover the known AP-1/IRF4 composite element (AICE). **(E)** Potential transcription factor composite elements identified from SPICE. Heat map of the significant motif pairs, with at least one pair of motifs exhibiting an E-value < $1e^{-10}$. This 118 x 205 spatial interaction matrix using Transcription Factor Binding Sites (TFBS) from the ENCODE project represents a subset of the primary and secondary motifs shown in the interaction matrix shown in **Figure S2**. The x-axis represents the primary motifs derived from ChIP-Seq libraries, and the y-axis indicates the motifs from HOCOMOCO database. Potential

novel combinatory complexes are highlighted in black boxes and yellow font (e.g., CTCF/ETS composite elements and JUN/IKZF composite elements).

Figure 2. JUN co-localizes with IKZF1. (**A**) Number of ChIP-Seq peaks identified for IKZF1 and JUN in experiments from primary mouse B cells that were treated with LPS + IL-21. (**B**) De novo motif discovery (HOMER) from IKZF1 peaks in primary mouse B cells treated with LPS + IL-21. The most significantly enriched motifs and their corresponding p-values are shown. (**C**) Venn diagram showing that most (88.3%) IKZF1 peaks co-localize with JUN. (**D**) Reactome pathways analysis from the IKZF1-JUN co-localized peaks reveal that cytokine related pathways were most enriched, including "Cytokine Signaling in Immune System", "Signaling by Interleukins", and several TLR4-related pathways/cascades. (**D**) Heat map showing that IKZF1 binding was potently activated by LPS or LPS + IL-21 treatment and that IKZF1 binds in proximity to JUN. Shown are normalized ChIP-Seq signals ± 3 kb, centered on IKZF1 peak summits. (**E**) IGV browser file shows IKZF1 and JUN bound to the *Prdm1* and *Il10* genes. The red boxes indicate known regulatory elements. The position of a known upstream cis-regulatory element (CNS9) that controls *Il10* expression is indicated.

Figure 3. Cooperative binding of IKZF1 and JUN at an *IL10* upstream site, CNS9. (A) Co-immunoprecipitation reveals IKZF1 physically associates with JUN. Nuclear protein lysates from 25 x 10^6 cells were used for each IP reaction. Protein extracts (input) were immunoprecipitated with antibodies to IKZF1, JUN, or normal IgG (mouse and rabbit) and resolved by SDS-PAGE. Western blotting was then performed with antibodies to IKZF1 or JUN. (**B**) Wild-type and mutant probes at the *IL10* CNS9 upstream site; variant IKZF1 and AP-1

motifs are underlined, with mutant nucleotides colored in red. (**C**) Representive EMSA with *IL10* upstream probes (wild-type, or IKZF1, AP-1, or IKZF1/AP-1 mutants; see panel B) and nuclear extracts from MINO cells. 100 nM of IR700-labeled probes was used per EMSA reaction. 5 μg of MINO nuclear extract was used in lanes 2, 4, 6, an 8; no nuclear extract was added in control lanes 1, 3, 5, and 7. (**D**) Average intensities from three independent EMSA experiments (one of which is shown in panel C). Band intensities were calculated using Image Studio software (LI-COR). (**E**) EMSA super-shifting was performed with 5 μg of nuclear extracts from MINO cells and mouse IgG, rabbit IgG, anti-IKZF1, and anti-JUN. Nuclear lysates were pre-incubated for 20 minutes on ice with 1 μg of the indicated antibodies prior to addition of probe. EMSAs were performed at least 3 times.

Figure 4. Both IKZF1 and JUN are critical for cooperative binding and transcriptional activation. (A, B) Cooperative binding of IKZF1 and AP1. EMSA were performed with a WT *IL10* probe corresponding to CNS9 and nuclear extracts from MINO cells stimulated with LPS + IL-21. Un-labeled "cold" wild-type or mutant double-stranded oligonucleotides (25, 50, 100, or 200-fold molar excess relative to the IR700-labeled probes; see Fig. 4B) were added to 5 μg nuclear extracts (prepared from MINO cells treated with LPS + IL-21) prior to the addition of 100nM of WT IR700-labeled probe. Shown are a representative EMSA (A) and summary of relative band intensities from three independent experiments (B). Relative intensities were calculated using Image Studio software (LI-COR). (**C, D**) EMSA using human recombinant AP-1 or IKZF1 with *IL10* CNS9 upstream probes (wild-type, or IKZF1/AP-1 mutants). 300 ng of recombinant AP1 protein (150 ng of each cJUN and cFOS) and 1 μg of recombinant IKZF1 were used as indicated. Band intensities from three independent experiments are shown in panel D. (**E**)

WT or mutant reporter constructs were transfected via electroporation into mouse primary B cells pre-stimulated with LPS for 24 hours. After electroporation, cells were treated with LPS for 24 hours before dual luciferase activity was measured (n=3; mean ± S.D.). (**F**) WT or mutant reporter constructs were transfected into mouse primary pre-activated $CD8^+$ T cells and the cells were then treated as indicated for 24 h (n = 3; mean ± S.D.). RLU, relative light units.

Figure S1. SPICE successfully predicts AICE, STAT5 tetramers, and CTCF/ETS complexes. SPICE successfully predicted previously identified composite elements, including AICE (AP-1/IRF4 composite elements), based on BATF, JUN and IRF4 ChIP-Seq data in TCR pre-activated mouse T cells (**A, B**), STAT5 tetramer complex formation based on STAT5 ChIP-Seq in TCR pre-activated and IL-2-treated T cells (**C, D**), and CTCF/ETS composite elements based on available ETS2 ChIP-Seq data (**E, F**).

Figure S2. SPICE predicts transcription factor composite elements using the Encode ChIP-Seq data. Heat map of the motif interaction matrix of ChIP-Seq libraries from Transcription Factor Binding Sites (TFBSs) from the ENCODE project. The x-axis represents the primary motifs derived from 343 Encode ChIP-Seq libraries, and the y-axis indicates the 401 motifs from HOCOMOCO database. The color scale represents the -log10 transformed E-value, which is the lowest p-value of any spacing of the secondary motif times the number of secondary motifs.

Figure S3. JUN-IKZF composite elements in GM12878 cells. (A) Heat map of genome-wide ChIP-Seq binding intensities in a ± 3 kb genomic region centered on summits of combined IKZF1 and JUN peaks in GM12878 cells from the ENCODE project; two biological replicate

experiments are shown in the heat map. (**B**) Co-localization of IKZF1, JUN, and BATF peaks at the *IL10* and *TNFRSF8* genomic loci. A conserved cis-regulatory element in the human *IL10* and mouse *Il10* loci, CNS9, is highlighted in the red box. (**C**) Sequence alignment reveals CNS9 region is highly conserved in human and mouse.

Figure S4. Gene-concept network depicts the linkages of genes and biological concepts of the top 5 Reactome Pathways that were enriched in IKZF1-JUN shared peaks. The highlighted red dots represent the top 5 enriched pathways and the dot size indicates the number of genes involved in each pathway.

Table 1. SPICE can predict possible tetramers for STAT family member proteins. SPICE is capable of identifying canonial STAT binding motifs, also known as interferon gamma-activated sequence (GAS) motifs, and then predicts STAT tetramer formation and the optimal spacing for tetramer formation for STAT1, STAT3, STAT4, STAT5A, and STAT5B, whereas STAT2 was predicted to not form tetramers. The various STAT family members were activated in the indicated cell types by the indicated cytokine, ChIP-Seq derived canonical GAS motifs, percentage of GAS motifs, likelihood of tetramer formation, and predicted optimal spacing were determined.

Table 2. The spatial interaction matrix of primary motifs from ChIP-Seq libraries and secondary motifs from HOCOMOCO database. Each row represents one of the 401 motifs from HOCOMOCO database, and each column is derived motifs from 343 TFBS ChIP-Seq libraries. Each cell has an E-value indicating the significance of motif pairs, which is the lowest

p-value of any spacing of the secondary motif times the number of secondary motifs. The value of 1 indicates there is no interaction between motif pairs.

Materials and Methods:

SPICE pipeline

All mapped files of Transcription Factor Binding Sites (TFBS) were downloaded from ENCODE project (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeSydhTfbs/), from a collection of libraries generated by Stanford/Yale/USC/Harvard (SYUH). The following are the step-by-step analysis corresponding to the flowcart in Figure 1A.

"Idenfity TF peaks": MACS 1.4.2(8) was used to identifiy significant peaks for each transcription factor using input IgG as control, *p-value* threshold was set as $1e^{-10}$. Alternatively, one can directly download SYUH pre-defined peaks-- i.e., SydhTfbsGm12878JundIggrabPk.narrowPeak. *"De novo motif analysis"*: MEME (multiple EM for motif elicitation) was used to to "de novo" discover consensus primary motifs from the 1000 most significant TF binding sites (sorted by *-log10(p-value)* in peak file). *"Using top-enriched motif as primary motif"*: The top-enriched motif (sorted by *e-value*, which is based on its log likelihood ratio, width, sites, the background letter frequencies) was selected as the primary motif . *"Loading all sequences", "Determining best primary motif matches", and "Eliminating sequences without primary matches"*: If there was indeed a primary motif from a particular TF ChIP-Seq, we extracted 500 bp DNA sequences centered on TF peak summits and only kept sequences containing primary motif matches and eliminated sequences without primary matches. *"Loading secondary motifs from database" and "Scanning for secondary motifs"*: We performed a spacing distribution analysis using SpaMo from MEME suite(15) from DNA sequences containing primary motif to search for the strongest secondary motif from a motif

database (HOCOMOCO, v11) to identify significantly enriched spacings. The relative spacings of the primary and secondary motifs in all identified ChIP-Seq peaks were counted and the probability of each spacing was calculated. A *p-value* was assigned based on the Binomial Test, using the number of observed secondary spacings falling into the given 1-bp bin, adjusted for the number of bins tested. *"Sorting significant secondary motifs and spacing" and "Grouping redundant secondary motifs*": After all the calculations were done, SPICE ranked the non-redundant secondary motifs with the best spacing, or gap (e.g., 4 bp) that passed the significance threshold (*p-value* < 0.01) in order of significance. Similar secondary motifs belonging to the same cluster (e.g., ETS1 and ELF1 belong to the same ETS cluster) were grouped together and ranked in order of *p-value*. We performed the above step-by-step analysis iteratively for each individual TF ChIP-Seq and generated the final result of interaction sparse matrix to plot in heat map.

Cell culture and reagents

MINO cells were cultured in RPMI-1640 medium (Gibco, 11875-003) containing 10% fetal bovine serum (FBS, Gemini Bio-Products, 100-106) and 1% penicillin-streptomycin (Gemini Bio-Products, 400-109). Mouse splenic B cells were isolated using the EasySep™ Mouse B cell isolation Kit (STEMCELL technologies, 19854) according to the manufacturer's protocol. Primary mouse B cells were cultured in RPMI-1640 medium containing 10% FBS, 1% penicillin-streptomycin and 50 μM 2-mercaptoethanol (Gibco, 21985023). LPS (Escherichia coli O111:B4) was purchased from Sigma-Aldrich (L3012). Recombinant human and mouse IL-21 protein was purchased from R & D systems.

Mice

Wild-type 6 -10 weeks old female C57BL/6 background mice were purchased from Charles River Laboratories (strain #556). All experiments with mice were conducted under protocols approved by the NHLBI Animal Care and Use committee. NIH guidelines for use of animals in intramural research were followed.

Electrophoretic mobility shift assays

MINO cells were treated with LPS (1 μg ml^{-1}) and IL-21 (100 ng ml^{-1}) for 3 hours. 10^7 cells were washed with 1 ml of ice-cold PBS and resuspended in 1 ml of ice-cold lysis buffer (10 mM HEPES pH 7.9, 10 mM KCL, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT and protease inhibitors). After 10 minutes of incubation on ice, 32 μl of 10% NP-40 was added and vortexed for 10 seconds. The samples were centrifuged at 13,000 rpm for 30 seconds, and the supernatants was discarded. Pellets were washed with 1 ml of ice-cold lysis buffer and centrifuged at 13000 rpm for 30 seconds, supernatants were discarded, and 100 μl of nuclear extraction buffer (20 mM HEPES pH 7.9, 0.4 M NaCl, 1 mM EDTA, 1 mM EGTA, 1mM DTT and protease inhibitors) was added to each tube and vortexed. The samples were incubated on ice for 30 minutes and spun out at 13,000 rpm for 10 minutes. The supernatant (nuclear extract) was transferred to a fresh tube and the pellet was discarded. The protein concentration was measured using the Pierce™ BCA protein assay kit – reducing agent compatible (Cat. No. 23250). IRDye 700-labeled and non-labeled probes were purchased from IDT DNA technologies. The EMSA reactions were set up according to the manufacturer's protocol (LI-COR, P/N: 829-07910). 5 μg of MINO nuclear extract was used per reaction. For super-shift assays, the nuclear extract was pre-incubated for 20 minutes on ice with anti-cJUN rabbit mAb (clone 60A8) from Cell

Signaling technologies (#9165S) or with anti-IKZF1 (clone IK14), provided by Dr. Katia Georgopoulos, Massachusetts General Hospital, Boston, MA. Recombinant human cJUN (ab84134), cFOS (ab56280) and IKZF1 (ab169877) proteins were purchased from Abcam. Equimolar concentrations of recombinant cJUN and cFOS proteins were mixed together and incubated on ice for 20 minutes to generate recombinant AP1 protein for EMSA. 300 ng of recombinant AP1 and 1 μg of recombinant IKZF1 was used per EMSA reaction. 10X orange loading dye was purchased from LI-COR (P/N 927-10100). The samples were run on a 4 - 12% TBE gel (Invitrogen, EC62352BOX) in 0.5 X TBE buffer (KD medical, RGE-3330).

Co-immunoprecipitation assays

MINO cells were stimulated for 3 hours with LPS (1 μg ml^{-1}) and IL-21 (100 ng ml^{-1}), and nuclear co-IP was performed according to the manufacturer's protocol (Active Motif, 54001). 25 x 10^6 cells were used per IP reaction. Protein lysates were immunoprecipitated with antibodies to cJUN (Cell Signaling Technologies, 9165S), IKZF1 (Sigma-Aldrich, MABE913), normal mouse IgG (Santa Cruz, sc-2025) and normal rabbit IgG (Cell Signaling Technologies, 2729S). 5 μg of antibody was used per reaction. The protein-antibody complex was captured using Pierce™ protein A/G magnetic beads (88802). NuPage™ LDS sample buffer (4X) (Invitrogen, NP0007) and NuPage™ sample reducing agent (10X) (Invitrogen, NP0004) was used to prepare the IP samples for denaturing western blot. The samples were loaded onto NuPage™ 4 - 12% bis-tris gels (Invitrogen, NP0335BOX). After gel electrophoresis, the proteins were transferred onto a PVDF Low fluorescence membrane (Bio-Rad, 1620261). The membranes were blocked for 1 hour at room temperature with 5% skim milk in Tris-buffered Saline (TBS). After washing three times with TBS, the membrane was incubated overnight at

$4^{o}C$ with primary antibodies to cJUN (Cell Signaling Technologies, 9165S), or IKZF1 (Sigma-Aldrich, MABE913). Blots were then washed and probed with secondary antibodies: IRDye 680RD Goat anti-rabbit IgG (LI-COR, P/N 926-68071) or IRDye 800CW Goat anti-mouse IgG (LI-COR, P/N 926-32210) at 1:10000 dilution. The membranes were imaged using ODYSSEY CLx (LI-COR).

Luciferase reporter assays

Mouse splenic B cells were pre-stimulated with LPS (1 μg ml^{-1}) for 24 hours. 5 x 10^6 cells were electroporated with 1 μg of NanoLuc pNL3.1 (Promega, N1031) vector and 0.2 μg of control vector, pGL4.S4 luc TK (Promega) using the P3 Primary cell 4D-nucleofector X kit S from Lonza (V4XP-3032) according to the manufacturer's protocol. The wild-type and mutant oligos were cloned into pNL3.1 vector using XhoI and HindIII restriction sites. After electroporation, the cells were stimulated with 1 μg ml^{-1} LPS for 24 hours. Dual luciferase reporter activity was measured using the Nano-Glo Dual-Luciferase Reporter assay (Promega, N1630). The luciferase activity data shown is relative to the control pGL4.S4 luc TK activity. Primary mouse $CD8^+$ T cells were purified from spleens from WT C57BL/6 mice. Cells were activated for 24 hours with anti-CD3 (2 μg ml^{-1}) and anti-CD28 (1 μg ml^{-1}). Cells (0.5 x 10^6) were then transfected with 1 μg nanoluciferase reporter constructs by nucleofection and stimulated immediately for 24 hours with IL-21 or anti-CD3 + anti-CD28 or the combination of IL-21 with anti-CD3 + anti-CD28. Luciferase activity was measured 24 hours later and expressed as RLU relative to the control plasmid.

ChIP-Seq library preparation

For ChIP-Seq experiments, 8-10 million B cells were treated with either LPS (1 μg ml^{-1}) or LPS + mouse IL-21 (100 ng ml^{-1}) for 3 hours. Immunoprecipitation was performed overnight at $4^{o}C$ using antibodies to cJUN (Cell Signaling Technologies, 9165S), IKZF1 (Active Motif, 39291), normal mouse IgG (Santa Cruz, sc-2025), or normal rabbit IgG (Cell Signaling Technologies, 2729S). ChIP-Seq DNA libraries were prepared with the KAPA LTP Library Preparation Kit and barcoded with NEXTflex DNA barcodes, quantified, and sequenced on an Illumina HiSeq 3000 system.

Bioinformatics analyses

Sequenced reads (50 bp, single end) were obtained with the Illumina CASAVA pipeline and mapped to the mouse genome mm10 (GRCm38, December 2011) using Bowtie 2.2.6. Only uniquely mapped reads were retained. The mapped outputs were converted to browser-extensible data files, which were then converted to binary tiled data files (TDFs) using IGVTools 2.4.13 for viewing on the IGV browser (https://software.broadinstitute.org/software/igv/). TDFs represent the average alignment or feature density for a specified 20 bps window size across the genome. Peak calling was performed by MACS 1.4.2(8) relative to an IgG library as input control and the p-value threshold was set as $1e^{-10}$. Only non-redundant reads were analyzed for peak calling. ChIP-Seq binding intensity heat maps were analyzed and visualized using deepTools. ChIP-Seq co-localization analysis was performed using R package “ChIPseeker”, and Reactome pathway analysis was performed using R package “ClusterProfiler”.

Data availability

All data generated or analyzed during this study are included in the manuscript and supporting file. The raw and processed datasets of ChIP-Seq are available in Gene Expression Omnibus

with accession GSE230035 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE230035, temporary reviewer's token: uvkzskimxzsjhup)

References

1. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823-37.
2. Roberson ED, Scearce-Levie K, Palop JJ, Yan F, Cheng IH, Wu T, et al. Reducing endogenous tau ameliorates amyloid beta-induced deficits in an Alzheimer's disease mouse model. Science. 2007;316(5825):750-4.
3. Deardorff MA, Bando M, Nakato R, Watrin E, Itoh T, Minamino M, et al. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. Nature. 2012;489(7415):313-7.
4. Izumi K, Nakato R, Zhang Z, Edmondson AC, Noon S, Dulik MC, et al. Germline gain-of-function mutations in AFF4 cause a developmental syndrome functionally linking the super elongation complex and cohesin. Nat Genet. 2015;47(4):338-44.
5. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473(7345):43-9.
6. Mikkelsen TS, Xu Z, Zhang X, Wang L, Gimble JM, Lander ES, et al. Comparative epigenomic analysis of murine and human adipogenesis. Cell. 2010;143(1):156-69.
7. Sutani T, Sakata T, Nakato R, Masuda K, Ishibashi M, Yamashita D, et al. Condensin targets and reduces unwound DNA structures associated with transcription in mitotic chromosome condensation. Nat Commun. 2015;6:7815.
8. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.
9. Harmanci A, Rozowsky J, Gerstein M. MUSIC: identification of enriched regions in ChIP-Seq experiments using a mappability-corrected multiscale signal processing framework. Genome Biol. 2014;15(10):474.
10. Rashid NU, Giresi PG, Ibrahim JG, Sun W, Lieb JD. ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. Genome Biol. 2011;12(7):R67.
11. Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics. 2011;27(12):1696-7.
12. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012;148(1-2):84-98.
13. Li P, Mitra S, Spolski R, Oh J, Liao W, Tang Z, et al. STAT5-mediated chromatin interactions in superenhancers activate IL-2 highly inducible genes: Functional dissection of the Il2ra gene locus. Proc Natl Acad Sci U S A. 2017;114(46):12111-9.
14. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010;28(5):495-501.
15. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37(Web Server issue):W202-8.
16. Karczewski KJ, Tatonetti NP, Landt SG, Yang X, Slifer T, Altman RB, et al. Cooperative transcription factor associations discovered using regulatory variation. Proc Natl Acad Sci U S A. 2011;108(32):13353-8.

17. Lin JX, Li P, Liu D, Jin HT, He J, Ata Ur Rasheed M, et al. Critical Role of STAT5 transcription factor tetramerization for cytokine responses and normal immune function. Immunity. 2012;36(4):586-99.
18. Liu Z, Little JW. The spacing between binding sites controls the mode of cooperative DNA-protein interactions: implications for evolution of regulatory circuitry. J Mol Biol. 1998;278(2):331-8.
19. Jolma A, Yin Y, Nitta KR, Dave K, Popov A, Taipale M, et al. DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature. 2015;527(7578):384-8.
20. Li P, Spolski R, Liao W, Wang L, Murphy TL, Murphy KM, et al. BATF-JUN is critical for IRF4-mediated transcription in T cells. Nature. 2012;490(7421):543-6.
21. Vinkemeier U, Cohen SL, Moarefi I, Chait BT, Kuriyan J, Darnell JE, Jr. DNA binding of in vitro activated Stat1 alpha, Stat1 beta and truncated Stat1: interaction between NH2-terminal domains stabilizes binding of two dimers to tandem DNA sites. EMBO J. 1996;15(20):5616-26.
22. Zhang X, Darnell JE, Jr. Functional importance of Stat3 tetramerization in activation of the alpha 2-macroglobulin gene. J Biol Chem. 2001;276(36):33576-81.
23. Vinkemeier U, Moarefi I, Darnell JE, Jr., Kuriyan J. Structure of the amino-terminal protein interaction domain of STAT-4. Science. 1998;279(5353):1048-52.
24. Zhang X, Wrzeszczynska MH, Horvath CM, Darnell JE, Jr. Interacting regions in Stat3 and c-Jun that participate in cooperative transcriptional activation. Mol Cell Biol. 1999;19(10):7138-46.
25. Pham D, Moseley CE, Gao M, Savic D, Winstead CJ, Sun M, et al. Batf Pioneers the Reorganization of Chromatin in Developing Effector T Cells via Ets1-Dependent Recruitment of Ctcf. Cell Rep. 2019;29(5):1203-20 e7.
26. Mullighan CG, Su X, Zhang J, Radtke I, Phillips LA, Miller CB, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med. 2009;360(5):470-80.
27. Ramirez J, Lukin K, Hagman J. From hematopoietic progenitors to B cells: mechanisms of lineage restriction and commitment. Curr Opin Immunol. 2010;22(2):177-84.
28. Nutt SL, Kee BL. The transcriptional regulation of B cell lineage commitment. Immunity. 2007;26(6):715-25.
29. Kleinmann E, Geimer Le Lay AS, Sellars M, Kastner P, Chan S. Ikaros represses the transcriptional response to Notch signaling in T-cell development. Mol Cell Biol. 2008;28(24):7465-75.
30. Riera-Sans L, Behrens A. Regulation of alphabeta/gammadelta T cell development by the activator protein 1 transcription factor c-Jun. J Immunol. 2007;178(9):5690-700.
31. Carr TM, Wheaton JD, Houtz GM, Ciofani M. JunB promotes Th17 cell identity and restrains alternative CD4(+) T-cell programs during inflammation. Nat Commun. 2017;8(1):301.
32. Sellars M, Kastner P, Chan S. Ikaros in B cell development and function. World J Biol Chem. 2011;2(6):132-9.
33. Yoshida T, Georgopoulos K. Ikaros fingers on lymphocyte differentiation. Int J Hematol. 2014;100(3):220-9.
34. Wang YH, Tsai DY, Ko YA, Yang TT, Lin IY, Hung KH, et al. Blimp-1 Contributes to the Development and Function of Regulatory B Cells. Front Immunol. 2019;10:1909.
35. Saraiva M, O'Garra A. The regulation of IL-10 production by immune cells. Nat Rev Immunol. 2010;10(3):170-81.

36. Lee CG, Kang KH, So JS, Kwon HK, Son JS, Song MK, et al. A distal cis-regulatory element, CNS-9, controls NFAT1 and IRF4-mediated IL-10 gene activation in T helper cells. Mol Immunol. 2009;46(4):613-21.
37. Umetsu SE, Winandy S. Ikaros is a regulator of Il10 expression in CD4+ T cells. J Immunol. 2009;183(9):5518-25.

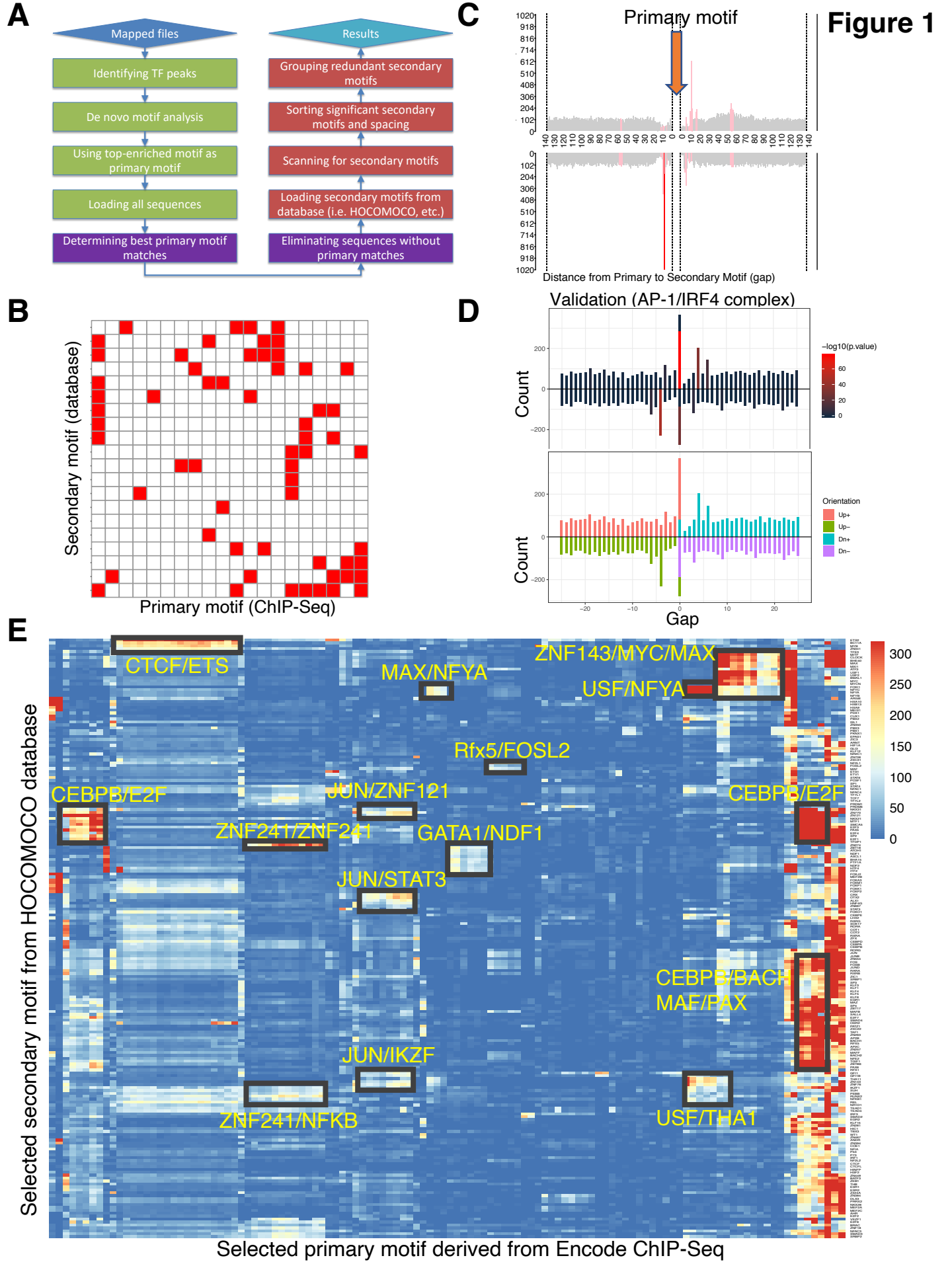


Figure 1
A
Mapped files
Identifying TF peaks
De novo motif analysis
Using top-enriched motif as primary motif
Loading all sequences
Determining best primary motif matches
Results
Grouping redundant secondary motifs
Sorting significant secondary motifs and spacing
Scanning for secondary motifs
Loading secondary motifs from database (i.e. HOCOMOCO, etc.)
Eliminating sequences without primary matches
B
Secondary motif (database)
Primary motif (ChIP-Seq)
C
Primary motif
Distance from Primary to Secondary Motif (gap)
D
Validation (AP-1/IRF4 complex)
Count
−log10(p.value)
Orientation
Up+
Up−
Dn+
Dn−
Gap
E
CTCF/ETS
MAX/NFYA
ZNF143/MYC/MAX
USF/NFYA
Rfx5/FOSL2
CEBPB/E2F
JUN/ZNF121
ZNF241/ZNF241
GATA1/NDF1
JUN/STAT3
CEBPB/BACH
MAF/PAX
JUN/IKZF
ZNF241/NFKB
USF/THA1
Selected secondary motif from HOCOMOCO database
Selected primary motif derived from Encode ChIP-Seq

Figure 2

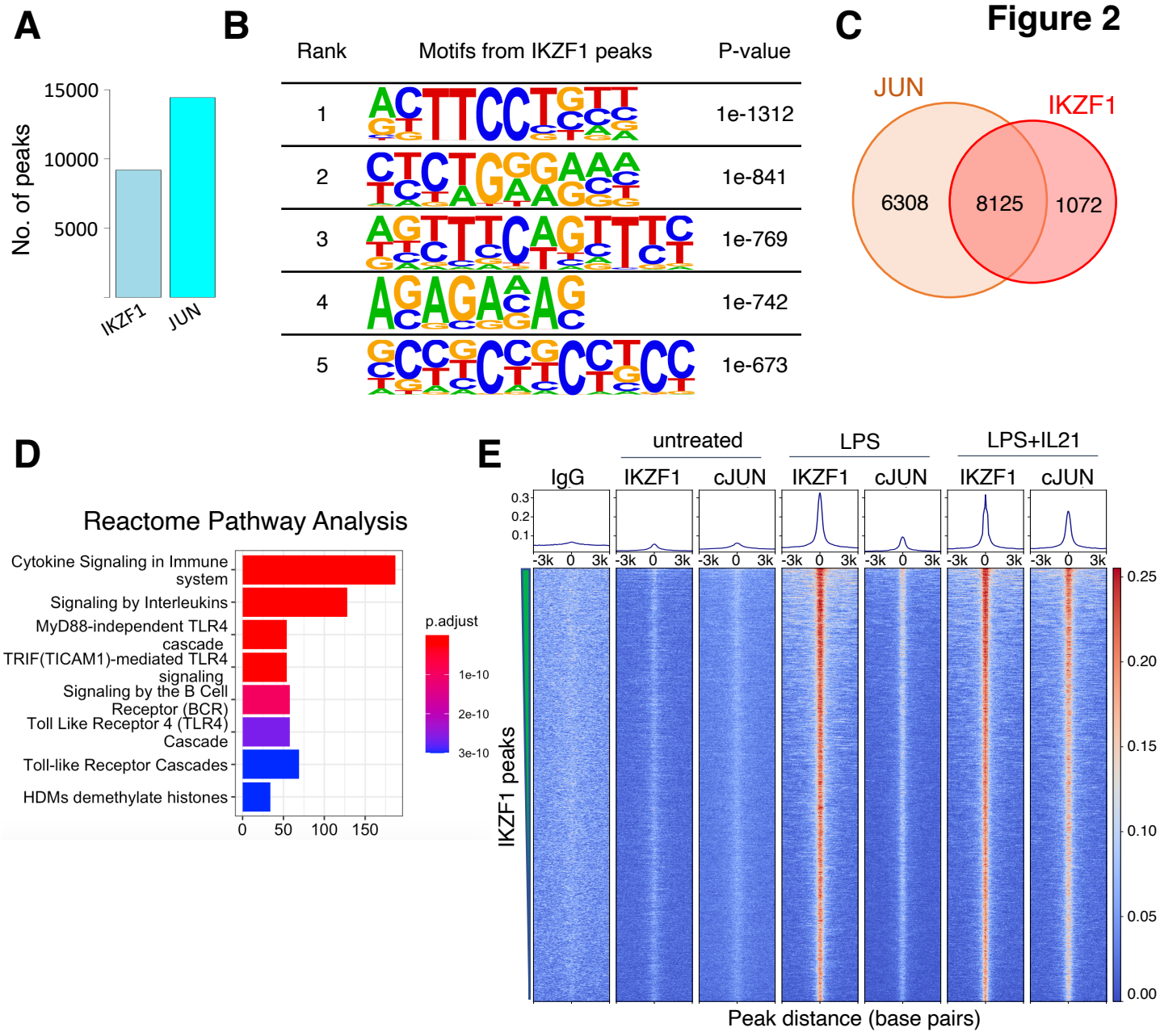

A
No. of peaks
15000
10000
5000
IKZF1
JUN
B
Rank
Motifs from IKZF1 peaks
P-value
1
1e-1312
2
1e-841
3
1e-769
4
1e-742
5
1e-673
C
JUN
IKZF1
6308
8125
1072
D
Reactome Pathway Analysis
Cytokine Signaling in Immune system
Signaling by Interleukins
MyD88-independent TLR4 cascade
TRIF(TICAM1)-mediated TLR4 signaling
Signaling by the B Cell Receptor (BCR)
Toll Like Receptor 4 (TLR4) Cascade
Toll-like Receptor Cascades
HDMs demethylate histones
0
50
100
150
p.adjust
1e-10
2e-10
3e-10
E
untreated
LPS
LPS+IL21
IgG
IKZF1
cJUN
IKZF1
cJUN
IKZF1
cJUN
0.3
0.2
0.1
-3k
0
3k
IKZF1 peaks
0.25
0.20
0.15
0.10
0.05
0.00
Peak distance (base pairs)
F
anti-IKZF1
anti-JUN
Prdm1
anti-IKZF1
anti-JUN
CNS9
Il10

Figure 3

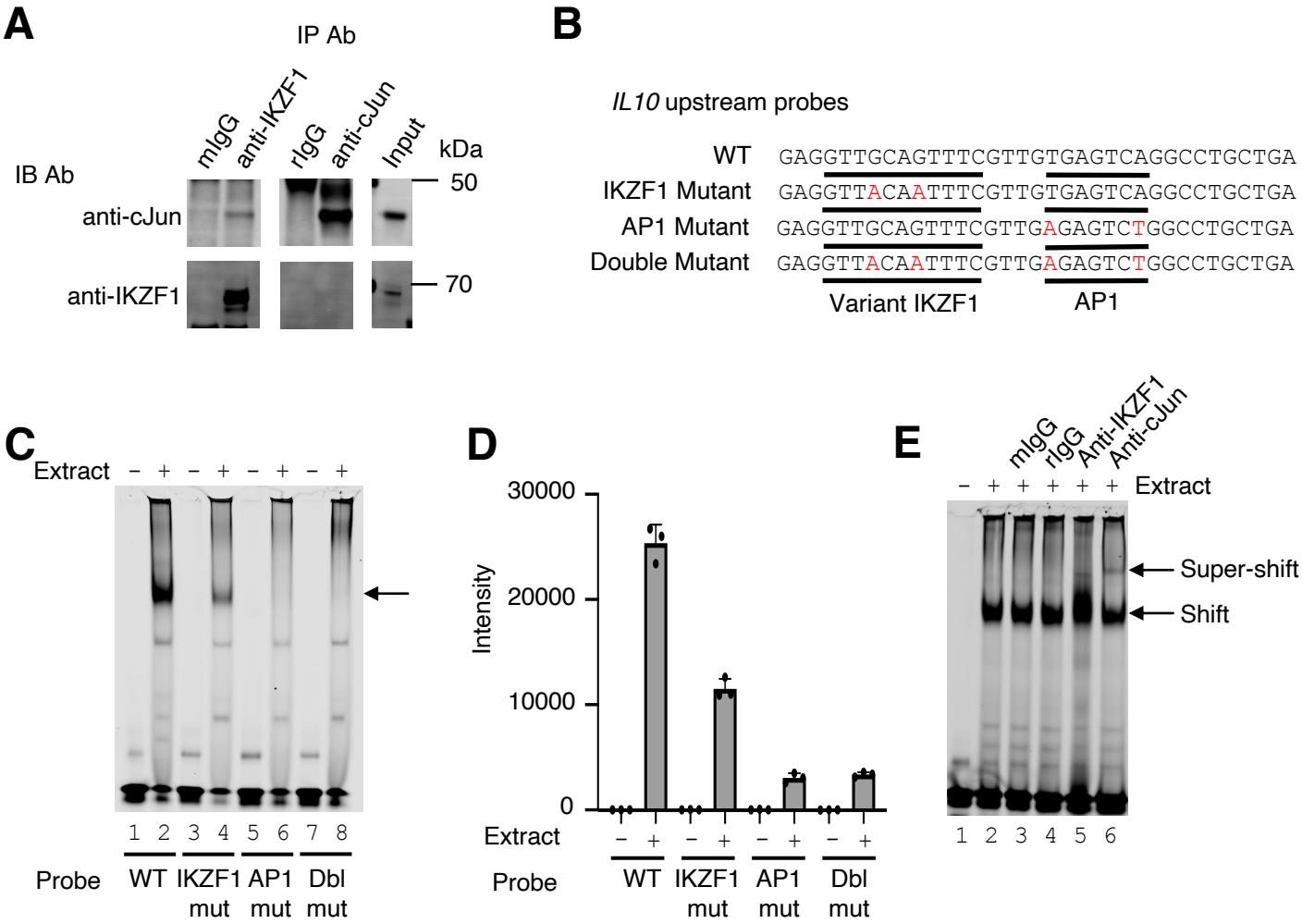

A
IP Ab
mIgG
anti-IKZF1
rIgG
anti-cJun
Input
kDa
IB Ab
anti-cJun
50
anti-IKZF1
70
B
IL10 upstream probes
WT GAGGTTGCAGTTTCGTTGTGAGTCAGGCCTGCTGA
IKZF1 Mutant GAGGTTACAATTTCGTTGTGAGTCAGGCCTGCTGA
AP1 Mutant GAGGTTGCAGTTTCGTTGAGAGTCTGGCCTGCTGA
Double Mutant GAGGTTACAATTTCGTTGAGAGTCTGGCCTGCTGA
Variant IKZF1
AP1
C
Extract − + − + − + − +
1 2 3 4 5 6 7 8
Probe WT IKZF1 mut AP1 mut Dbl mut
D
Intensity
30000
20000
10000
0
Extract − + − + − + − +
Probe WT IKZF1 mut AP1 mut Dbl mut
E
mIgG
rIgG
Anti-IKZF1
Anti-cJun
− + + + + + Extract
Super-shift
Shift
1 2 3 4 5 6

Figure 4

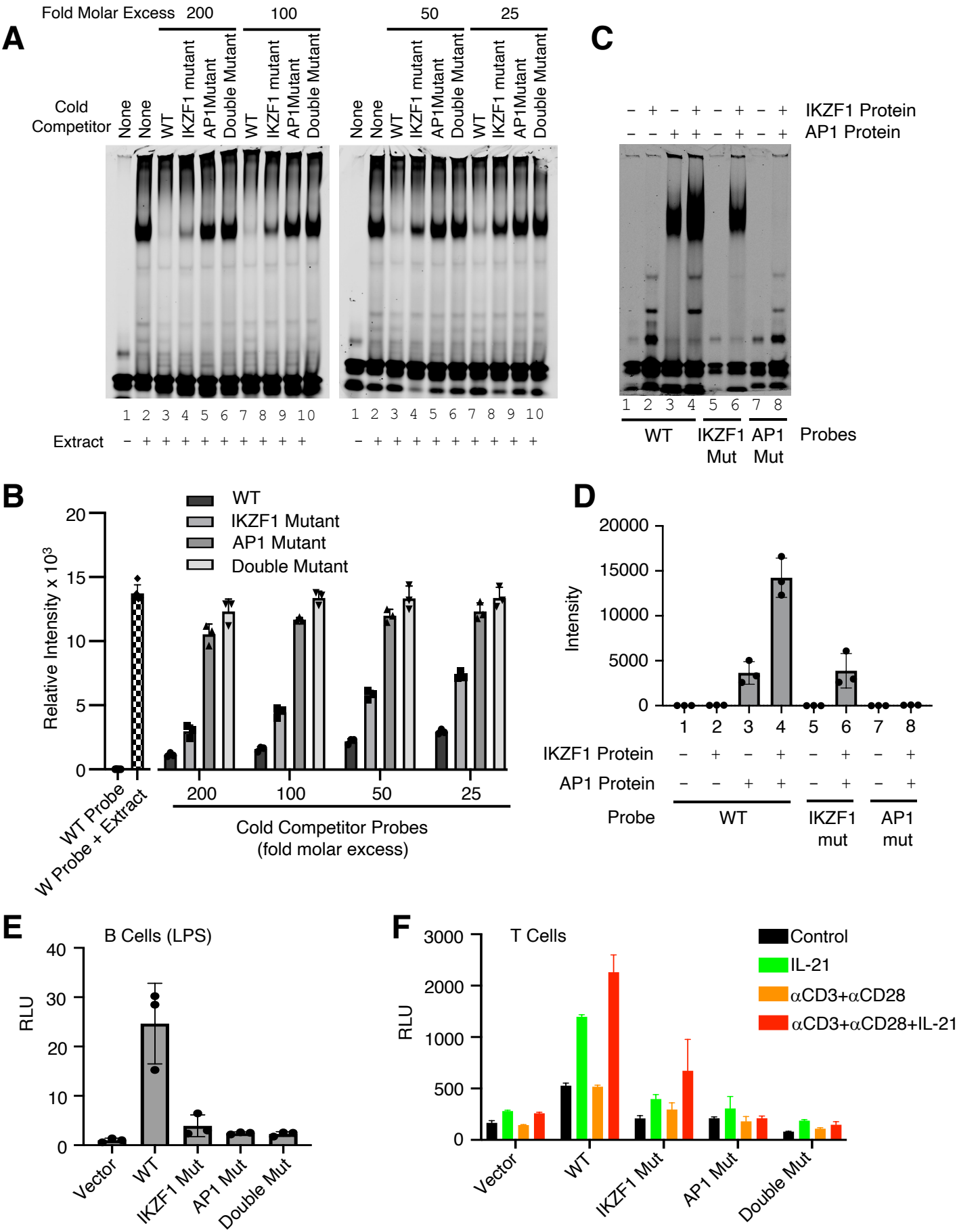

A
Fold Molar Excess 200 100 50 25
Cold Competitor None None WT IKZF1 mutant AP1Mutant Double Mutant WT IKZF1 mutant AP1Mutant Double Mutant None None WT IKZF1 mutant AP1Mutant Double Mutant WT IKZF1 mutant AP1Mutant Double Mutant
1 2 3 4 5 6 7 8 9 10
Extract − + + + + + + + + +
C
IKZF1 Protein
AP1 Protein
1 2 3 4 5 6 7 8
WT IKZF1 Mut AP1 Mut Probes
B
WT
IKZF1 Mutant
AP1 Mutant
Double Mutant
Relative Intensity x 10³
WT Probe
W Probe + Extract
200 100 50 25
Cold Competitor Probes
(fold molar excess)
D
Intensity
IKZF1 Protein
AP1 Protein
Probe WT IKZF1 mut AP1 mut
E
B Cells (LPS)
RLU
Vector WT IKZF1 Mut AP1 Mut Double Mut
F
T Cells
RLU
Control
IL-21
αCD3+αCD28
αCD3+αCD28+IL-21
Vector WT IKZF1 Mut AP1 Mut Double Mut

Figure S1

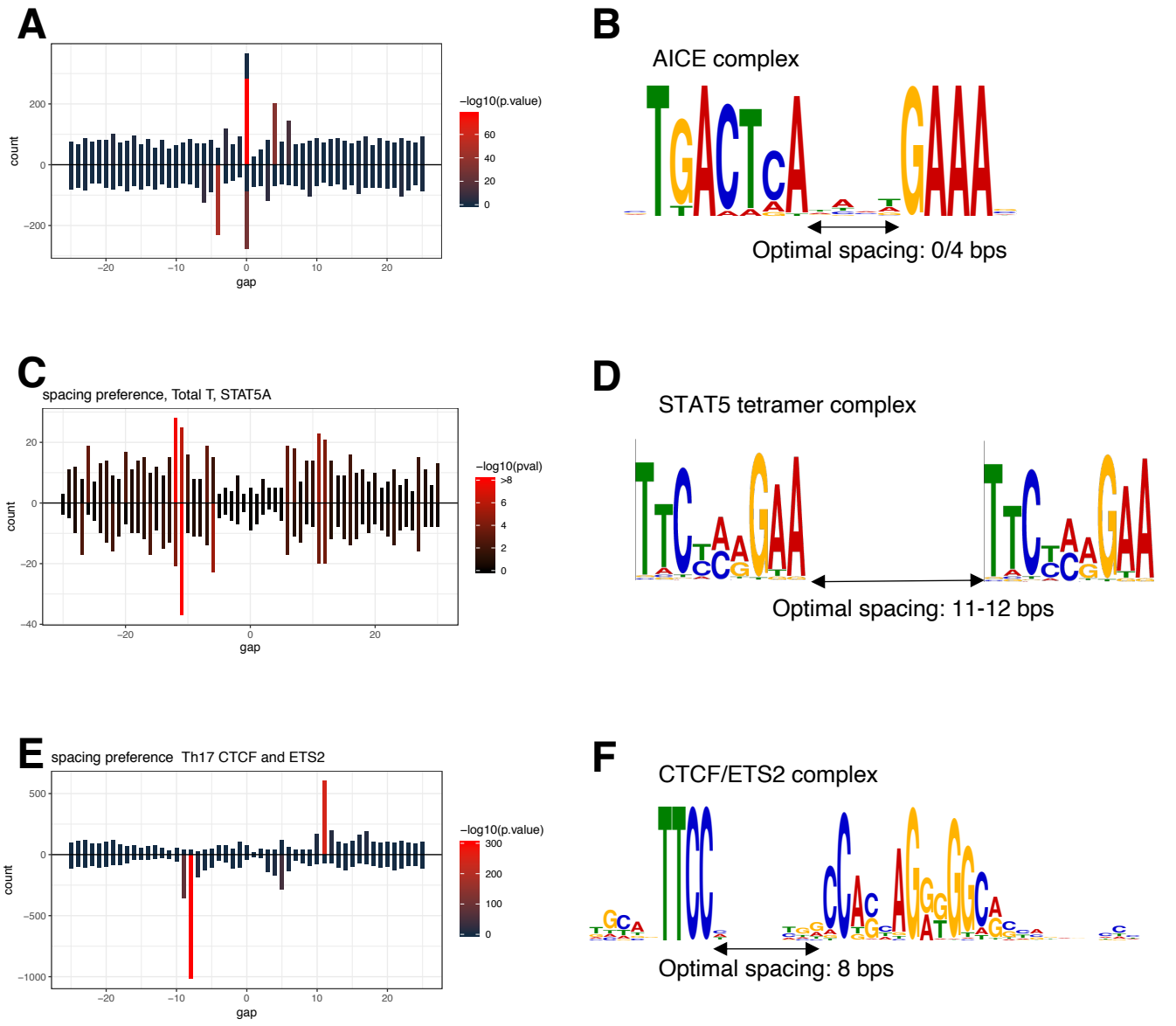

A
count
gap
−log10(p.value)
B
AICE complex
Optimal spacing: 0/4 bps
C
spacing preference, Total T, STAT5A
count
gap
−log10(pval)
D
STAT5 tetramer complex
Optimal spacing: 11-12 bps
E
spacing preference Th17 CTCF and ETS2
count
gap
−log10(p.value)
F
CTCF/ETS2 complex
Optimal spacing: 8 bps

Figure S2

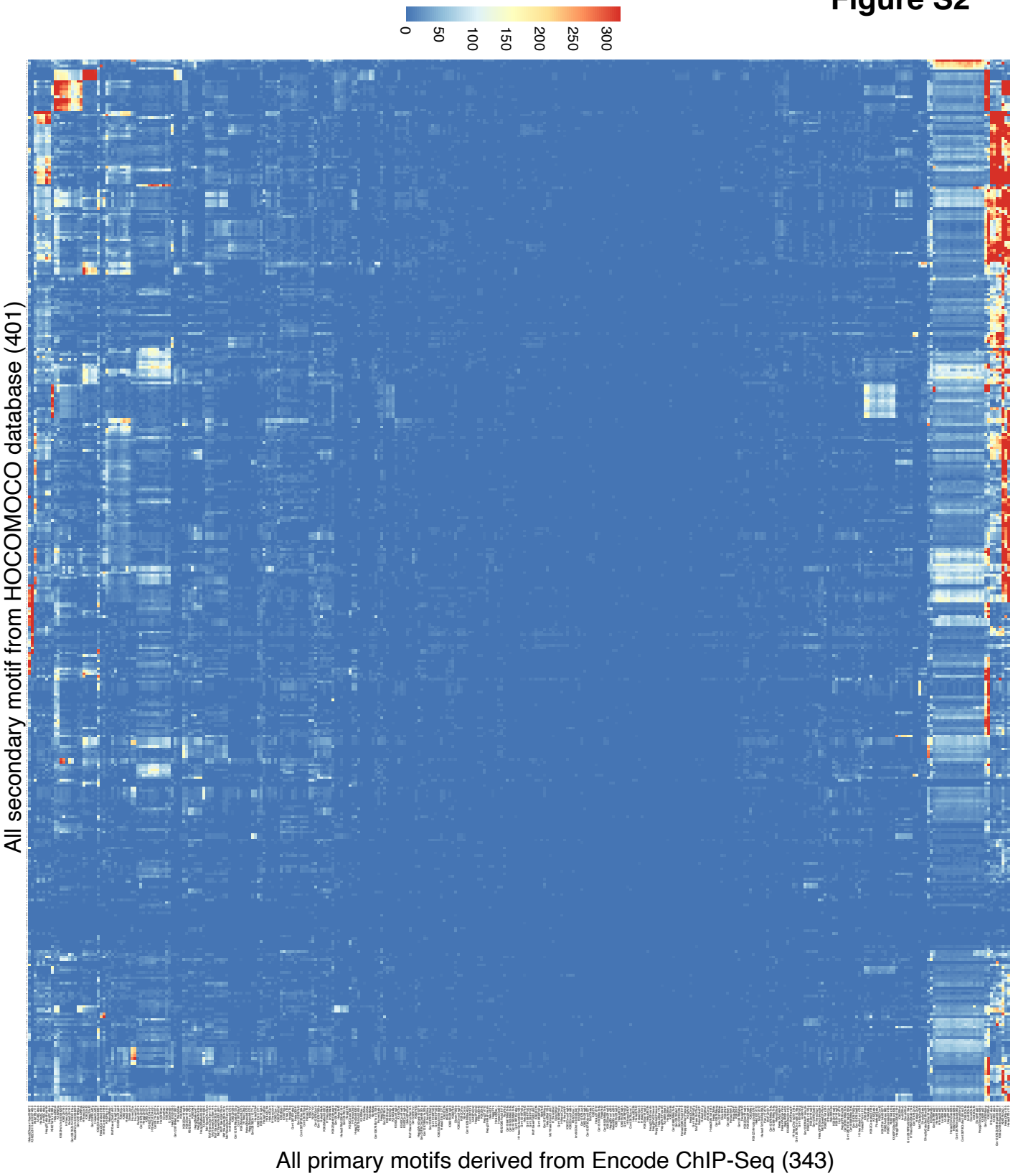
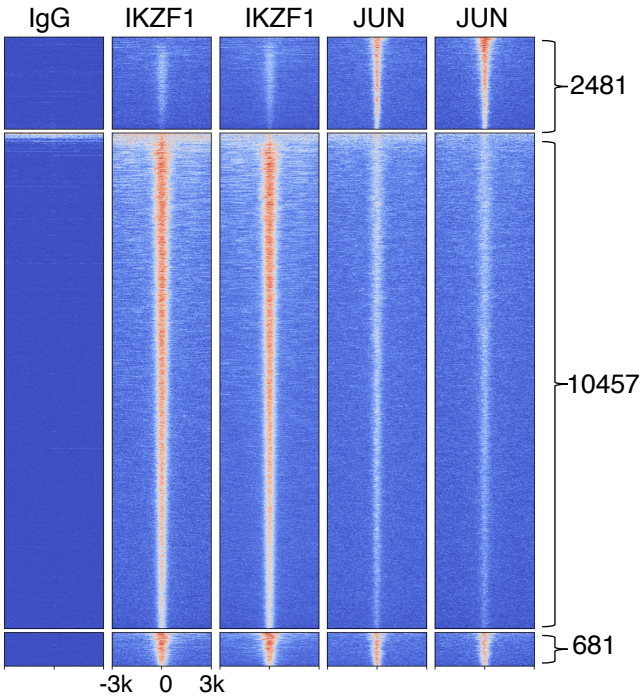
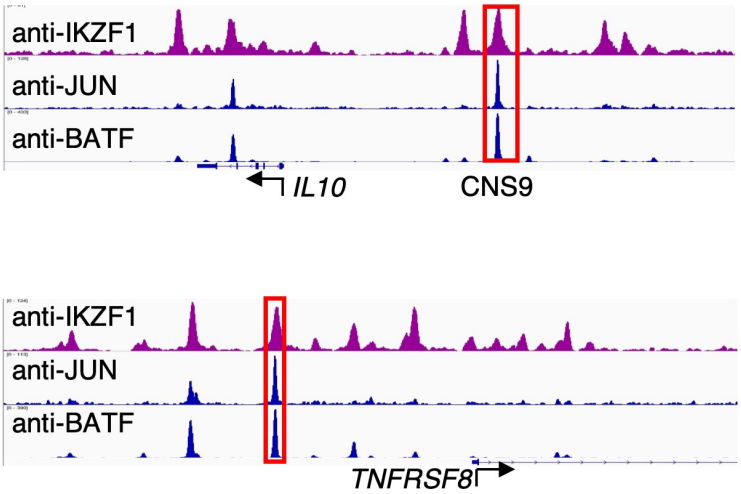

0
50
100
150
200
250
300
All secondary motif from HOCOMOCO database (401)
All primary motifs derived from Encode ChIP-Seq (343)

Figure S3

A



B



C

```
Human_CNS9:  CACTTTGAGGTTGCAGTTTCGTTGTGAGTCAGGCCTGCTGAGAGCAGT-
             ..|||||||||||||||||||||||||||||||||||||||||||| |
Mouse_CNS9:  TGCTTTGAGGTTGCAGTTTCGTTGTGAGTCAGGCCTGCTGAGAGCA-TC
```

Figure S4

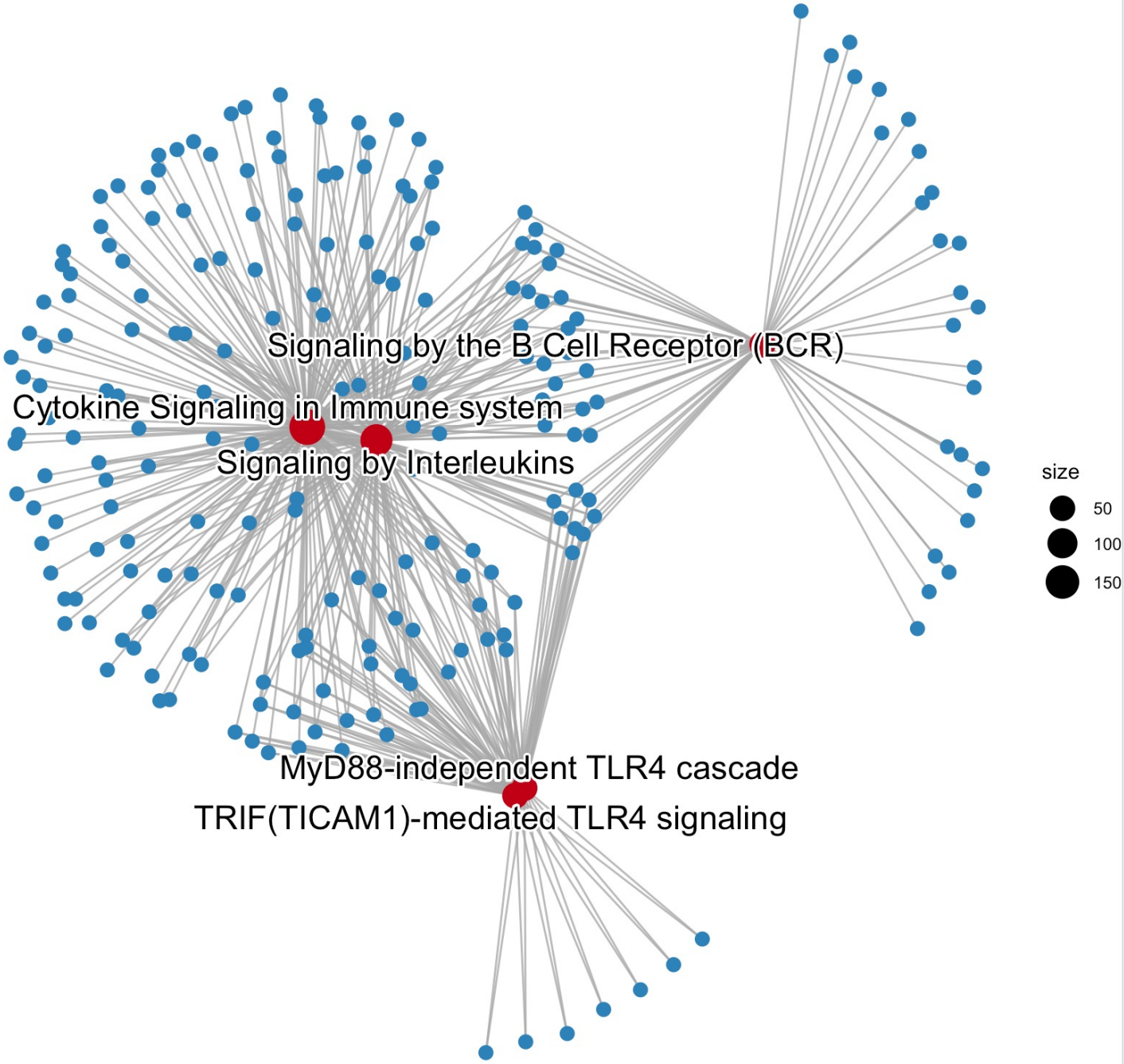
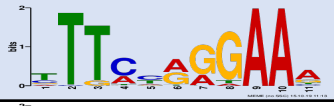

Signaling by the B Cell Receptor (BCR)
Cytokine Signaling in Immune system
Signaling by Interleukins
MyD88-independent TLR4 cascade
TRIF(TICAM1)-mediated TLR4 signaling
size
50
100
150

Table 1

STATs (Conditions)	Identified canonical GAS motif	GAS%	Tetramer likelihood	Optimal Spacing (bp)
STAT1 (MΦ, +IFN-γ)		87.4%	****	10
STAT2 (Th1, +IFN-γ)		38.2%	NA	NA
STAT3 (CD8T, +IL-21)		87.0%	***	11?
STAT4 (Th1)		95.4%	****	11-12
STAT5A (Total T, +IL-2)		99.8%	****	6-7, 11-12
STAT5B (Total T, +IL-2)		100%	****	6-7,11-12