Figure Legends

Figure S1. *Irf4*$^{-/-}$ B cells are not arrested at an immature stage and generate marginal zone B cells

A. Schematic depicting the generation of 1:1 mixed bone marrow chimeric animals. Chimeras were generated such that the CD45.1 expressing compartments in Groups A and B were reconstituted with *Irf4*$^{+/+}$ hematopoietic progenitors whereas the CD45.2 expressing compartments were reconstituted with *Irf4*$^{+/+}$ and *Irf4*$^{-/-}$ hematopoietic progenitors, respectively. Reconstituted mice were immunized with SRBC and GC B cells were analyzed on Day 7

B. Splenic B cells from control *Irf4*$^{+/-}$ or *Irf4*$^{-/-}$ mice do not harbor increased frequencies of B220^{+}CD93^{+} B cells; the phenotype of immature transitional B cells (left). Mature B cells, gated on B220^{+}CD93^{-} (left) were analyzed for the distribution of IgM and CD23; CD23^{lo}IgM^{hi} cells represent Marginal Zone B cells (right).

C. Analysis of CD23 expression within the B220^{+}CD93^{-} mature B cell subsets from (A).

D. IgM and IgD expression levels among B220^{+} splenic *Irf4*$^{+/-}$ and *Irf4*$^{-/-}$ B cells are shown (left). B220^{+}IgM^{hi}IgD^{lo} cells (left) are gated to analyze CD21^{hi}CD1d^{hi} B cells (right). Shown in panels A, B, and C are 2 mice of each genotype of more than five that were analyzed.

Figure S1, related to Figure 1.

Figure S2. Differing levels of IRF4 expression distinguish between GC B and plasma cells

A. Experimental design of adoptive transfer experiments involving *Irf4*$^{+/+}$ or *Irf4*$^{-/-}$ SW$_{HEL}$ donor B cells transplanted into CD45.1 hosts and immunized with HEL^{2X}SRBC.

B. Adoptively transferred SW$_{HEL}$ donor cells were analyzed by flow cytometry 4.5 days after immunization. SW$_{HEL}$ responder B cells are identified as CD45.1^{-}CD45.2^{+}B220^{+}HEL-binding cells. Flow cytometric analysis of intracellular levels of IRF4 and Bcl6 among HEL-specific B cells of *Irf4*$^{+/+}$ and *Irf4*$^{-/-}$ genotypes.

C. IRF4^{hi} HEL-specific B cells express higher levels of cytoplasmic HEL-specific Ig (cHEL) than their IRF4^{lo} counterparts. The left contour plot displays the distribution of IRF4 and cHEL. The middle histogram shows gating criteria used to separate the IRF4^{lo} and IRF4^{hi} populations. The right histogram shows the distribution of cHEL within the gated (middle panel) IRF4^{lo} and IRF4^{hi} populations.

D. Adoptively transferred, CFSE labeled, SW$_{HEL}$ B cells were analyzed by flow cytometry 3.5 days after immunization. SW$_{HEL}$ responder cells in the spleen were identified as DAPI^{-}CD45.1^{-}CD45.2^{+}B220^{+} and HEL-binding. Shown on the left is the distribution of B220 and HEL staining. On the right are the histograms depicting the respective CFSE dilution. The arrow in the *Irf4*$^{-/-}$ histogram highlights residual undivided cells. Results shown are representative of 3 experiments; n=6 mice.

E. Naive *Irf4*$^{+/+}$ SW_{HEL} B cells were sorted based on B220 expression and their ability to bind HEL. Antigen primed *Irf4*$^{+/+}$ and *Irf4*$^{-/-}$ SW_{HEL} B cells were sorted based on CD45.1^{-}CD45.2^{+}B220^{+} phenotype, and their ability to bind HEL. Sorting of antigen primed SW_{HEL} responder B cells was performed 4.5 days after adoptive transfer and immunization. Equal numbers of cells were sorted and then processed for RT-PCR analysis for Oct1 and Pax5 transcripts. Oct1 was used for normalization. Data shown are averages ±SEM from three independent experiments of two mice per group.

F. Quantitation of absolute numbers of cells analyzed from Fig. 2C. Numbers of HEL-specific, HEL-specific IRF4^{hi}Bcl6^{-}, and HEL-specific IRF4^{lo}Bcl6^{+} expressing cells from individual mice pooled from 3 experiments are plotted.

G. ELISpot analysis of plasma cells generated after adoptive transfer of SW_{HEL} B cells 4.5 days after immunization. Splenocytes were analyzed to enumerate HEL-specific plasma cells in response to HEL^{3X}, HEL^{2X}, HEL^{WT} antigen variants; shown are the average ±SEM of each group (n=3) from one representative experiment of four performed.

Figure S2, related to Figure 2.

Figure S3. Distinct modes of IRF4 action instruct GC B cell and plasma cell states

A. Schematic depicting the experimental design of *Irf4*-inducible (*Irf4*$^{-/-}$) HEL-specific SW_{HEL} donor B cells used in adoptive transfer experiments into (CD45.1, *Rosa*$^{M2rtTA/+}$) hosts, which were then immunized with HEL^{2X} conjugated to SRBC. Mice were administered water lacking DOX (- DOX), containing DOX throughout the 5 day experiment (+ DOX "sustained") or containing DOX for the first two days only (+ DOX "pulse").

B. The ratio of absolute numbers of HEL-specific IRF4^{hi}Bcl6^{-} to IRF4^{lo}Bcl6^{+} expressing cells on a per mouse basis from the experiment shown in Fig. 3A. Each point represents the analysis of a single mouse pooled from two experiments of three mice per group.

C. Quantitation of absolute numbers of cells analyzed from Fig. 3A. Each point represents the analysis of a single mouse pooled from two experiments of three mice per group.

D. Splenocytes analyzed by ELISpot to enumerate HEL-specific plasma cells in response to HEL^{2X} antigen after adoptive transfer of SW_{HEL} B cells 5 days after immunization; shown are the average ±SEM of each group (n=3) from one of two experiments that gave similar results.

E. Schematic depicting the experimental design of *Irf4*-inducible (*Irf4*$^{+/+}$) HEL-specific SW_{HEL} donor B cells used in adoptive transfer experiments into (CD45.1, *Rosa*$^{M2rtTA/+}$) hosts, which were then immunized with

HEL^{3X} conjugated to SRBC. Mice were administered water lacking DOX (- DOX) or containing DOX throughout the 5 day experiment (+ DOX "sustained").

F. The ratio of absolute numbers of HEL-specific $IRF4^{hi}Bcl6^{-}$ to $IRF4^{lo}Bcl6^{+}$ expressing cells on a per mouse basis from the experiment shown in Fig. 3D. Each point represents the analysis of a single mouse pooled from two experiments of three mice per group.

G. Quantitation of absolute numbers of cells analyzed from Fig. 3D. Each point represents the analysis of a single mouse pooled from two experiments of three mice per group.

H. Splenocytes analyzed by ELISpot to enumerate HEL-specific plasma cells in response to HEL^{3X} antigen after adoptive transfer of SW_{HEL} B cells 5 days after immunization; shown are the average ±SEM of each group (n=3) from one of two experiments that gave similar results.

Figure S3, related to Figure 3.

Figure S4. The genome binding landscape of IRF4 and PU.1 as a function of B cell activation and differentiation

A. Schematic of culture system utilizing B cells from the B1-8i anti-NP knock-in mouse. The B cells were stimulated with NP-Ficoll, CD40L, and IL-2/4/5. On Days 1 and 3, chromatin was crosslinked and processed for IRF4 and PU.1 specific ChIP coupled to massively parallel sequencing.

B. The differentiation status of antigen specific B cells at Day 3 of stimulation. Flow cytometry of anti-Syndecan-1 (plasma cells) and anti-IgG1 (class switched B cells) staining cells.

C. Distribution of IRF4 expression in Day 1 or Day 3 stimulated B cells as measured by intracellular staining and flow cytometry. (B) and (C) are representative of four experiments.

D. Union analysis of the IRF4 and PU.1 genomic enrichments (peaks); left, as a function of time and right, as a function of IRF4 and PU.1 co-occupancy.

E. The distribution of peaks as a function of genome annotation (NCBI37/mm9) within extragenic and intragenic regions (number represents % of total peaks falling within annotation category).

F. IRF4 (and PU.1) and IRF4 (not PU.1) peaks were chosen at random to be validated by ChIP-qPCR in antigen stimulated B1-8i B cells (Day 3). The positive control represents the Igkappa 3' enhancer. Three negative control regions were used to define the background, represented by the light green bar. Genomic coordinates of peaks are (1) chr10:44182566-44182566 (2) chr6:122520252-122520386 (3) chr10:21862750-21862977 (4) IgKappa 3' Enhancer (5) chr4:44683859-44684229 (6) chr4:32488557-32488800 (7) chr16:24151472-24151830 (8) chr18:69804123-69804696 (9) chr13:5863844-5864107 (10) chr1:108464037-

108464433 (11) chr2:127834302-127835161 (12) chr16:11385880-11386238 (13) chr10:44428699-44429007 (14) chr13:30824584-30824875 (15) chr9:32459894-32460227 (16) chr6:47571104-47571526 (17) chr7:71075755-71076040. Data are representative of two experiments.

G. Alignment of IRF4 and/not PU.1 binding peaks with genomic DNase I hypersensitive sites (HS). Coincidence frequency (%) is displayed based on alignment IRF4 and/not PU.1 binding peaks with DNaseI sites in naive B cells or fibroblasts from the ENCODE database.

H. Identification of flanking GAAA sequences to the ISRE and AP-1 motifs in our peak sets. For ISRE and AP-1, positions at 0 and 11 or 0 and 8, respectively, are directly abutting the motif. For ISRE analysis we only plot GAAA on the same strand as the motif. Background frequencies (gray lines) were obtained from random genomic sequences containing the ISRE or AP-1 consensus sequence (GAAAnnGAAA or TGAnTCA). The frequency plotted in the y-axis indicates the distribution of GAAA over all motifs with a nearby GAAA. At Day 1, 12% of ISRE motifs and 64% of AP-1 motifs had a nearby GAAA; at Day 3, 50% of ISRE motifs and 60% of AP-1 motifs had a nearby GAAA. Among the random sequence consensus hits, 27% of ISRE motifs and 40% of AP-1 motifs had a nearby GAAA.

I. Locations of motifs in ChIP-regions relative to peak maxima (centered at 200bp).

J. IRF4 binds the ISRE as a homodimer. Indicated COOH-truncation mutants of IRF4 were synthesized by coupled *in vitro* transcription and translation. Equal amounts of protein were used based on ^{35}S labeling and SDS-PAGE followed by autoradiography (not shown). Each mutant was used by itself or in combination as indicated above the gel. Mixing of IRF4 proteins of different sizes in EMSAs would be predicted to generate a heterodimeric complex with intermediate mobility compared to homodimeric complexes. Accordingly, a mixture of IRF4 proteins, 1-340aa and 1-419aa yielded a protein-DNA complex of intermediate mobility. The bottom panel displays the densitometric quantiation of each lane in the EMSA; densitometry is represented as a histogram. Results are representative of two experiments performed.

K. IRF4 and BATF co-binding to regions containing AICE motifs. B1-8i anti-NP B cells were stimulated with NP-Ficoll, CD40L, and IL2/4/5 prior to processing for ChIP-qPCR and analysis of 6 independent regions as indicated. Data are averaged from two independent experiments and average enrichment ±SEM. (% input chromatin) is shown for two experiments performed. The presumptive AICE motif present in regions analyzed are depicted below following the genomic coordinates of the peaks analyzed; red type indicates the AP-1 element and green type indicates the IRF element. Peaks are (1) chr11:76548055-76548473 TTTCTCAGAATGAGTCACA (2) chr11:51574783-51575047 TATGAGTCAGAGAGA (3) chr5:122738571-122738864 GCTTTGAGTCTGACTCAAT (4) chr13:63676131-63676286

TATGACTCAGAATCC (5) chr10:44100849-44101057 TTGAAATGACTCTGT (6) chr10:44164097-44164330 AGTCTTGTTATGAGTCATC (7) negative control.

Figure S4, related to Figure 4.

Figure S5. IRF4 binding regulates chromatin alterations at the *Prdm1* locus

A. Histogram of the IRF4 and PU.1 ChIPseq tag enrichments at the Blimp1 locus. The six peaks (P1 through 6, CNS9) identified in the ChIPseq analyses are indicated in black text. Additional regions analyzed by qPCR (panels B, C) as reference include the promoter (P1a), alternate promoter (P1b), IL21RE (Kwon et al., 2009), and two Conserved Non-coding Sequences upstream (CNS1) or in the intron (CNS5) of Blimp-1 are indicated in red text. The CNS1 amplicon served as a negative control.

B. Timecourse of IRF4 and PU.1 binding, and histone modifications in unstimulated and stimulated B1-8i anti-NP B cells. Cells were stimulated with NP-Ficoll, CD40L, and IL2/4/5 prior to processing for ChIP-qPCR. Data are averaged from two independent experiments and enrichment (% input chromatin) is represented as a heat map. The scale for % Input of each antibody precipitate is shown at the left of the heat map block. Analyzed regions are shown in (A).

C. IRF4 is required for the binding of PU.1 and the H3K4me1 and H3K27Ac chromatin marks at the *Blimp1* locus. ChIP-qPCR was performed on chromatin isolated from activated $Irf4^{+/+}$ and $Irf4^{-/-}$ B cells; enrichment (% input chromatin) is represented as a heat map. Increasing black intensity denotes increased enrichment; scale bar is shown at the left of each block. Data are averaged from three independent experiments.

D. IRF4 and PU.1 bind the EICE motif present within the targeted region upstream of the Bcl6 gene. Binding reactions using the EICE DNA probe derived from the Bcl6 region were carried out in the presence of the indicated PU.1 and IRF4 proteins and antibodies.

Figure S5, related to Figure 5.

Figure S6. Analysis of the IRF4 regulomes in naive, GC B, and plasma cells

A. We considered three groups of genes: GC only genes (white), GC+PC genes (grey) and PC only genes (black).

B. We plot the EMBER log-odds matrix for the three new behavior dimensions using the same style as for the expression patterns in Fig. 6 (Fig. 6B gives the third dimension of these matrices, showing behaviors in GC B cells versus plasma cells) in regards to Day 3 peaks of IRF4 (and PU.1) and IRF4 (not PU.1). We see that the associated expression patterns inferred *in vitro* (Fig. 6) is consistent with what is inferred *in vivo* for IRF4 (and PU.1) and IRF4 (not PU.1).

C. We plot the log-ratio and *p*-value for over-representation of each gene group among Day 3 peak group's EMBER targets. Bars are colored by the *p*-value of the observation (significant results are orange or red). We see no significant over-representation of GC only genes for either peak group. For GC+PC genes, we see strong over-representation for IRF4 (and PU.1) and moderate over-representation for IRF4 (not PU.1) peak groups. For PC only, we see weak over-representation for IRF4 (and PU.1) and strong over-representation for IRF4 (not PU.1) peak groups.

D. Gene ontology of inferred IRF4 target genes. The inferred target genes of IRF4 (and PU.1) and IRF4 (not PU.1) peak groups are associated with distinct functional classes of genes. The inferred target genes derived from the EMBER analysis in (Fig. 6A) were used to discover enriched classes of gene functions using the David ontology tool (***david***.*abcc.ncifcrf.gov/*). The top 3 ontology terms corresponding to Molecular Functions (MF), Cellular Component (CC) and Biological Process (BP) are reported along with their corresponding p-value.

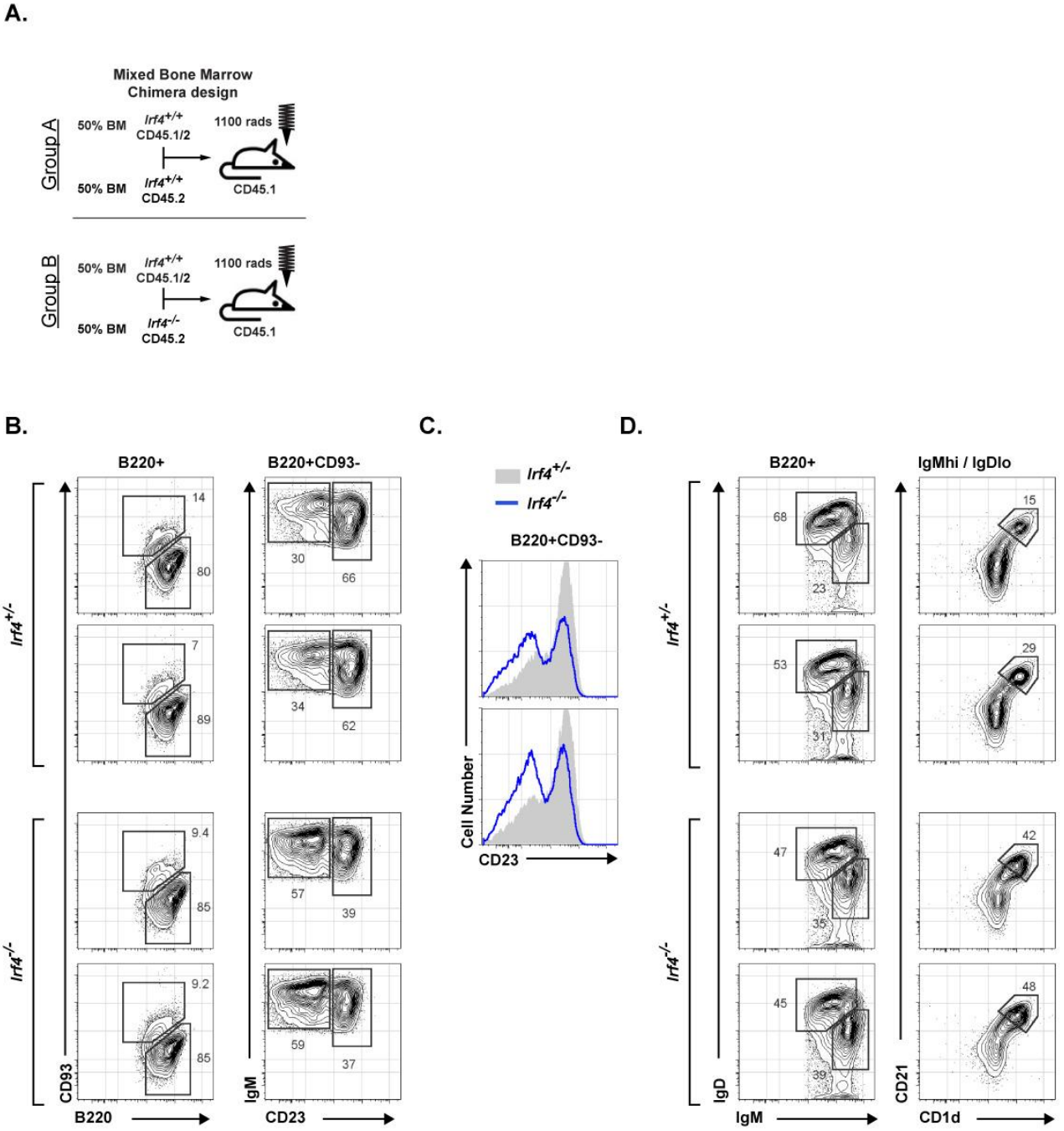
E. IRF4 (and PU.1) and IRF4 (not PU.1) genome targeting are associated with diverse patterns of transcriptional regulation. Results of the behavior dimensions shown here are identical to Fig. 6A; however, they are shown adjacent to additional differential gene expression dimensions for comparison. Genome-wide expression data was derived from purified B cells isolated from wild type (*Irf4*$^{+/+}$), mutant (*Irf4*$^{-/-}$), or *Irf4*-inducible mice that were stimulated with CD40L and cytokines. Total RNA was isolated on Days 0, 1 and 3 and processed for Affymetrix arrays. The Irf4-inducible mice were also cultured in the presence of 16ng/mL or 100ng/mL doxycycline (DOX) to induce the expression of the *Irf4* transgene, indicated by the triangle on the x-axis. To integrate the transcriptome analysis with the ChIPseq binding data, we first defined a set of pair-wise gene expression comparisons (specifically wild type cells versus *Irf4*$^{-/-}$ cells or Irf4-inducible cells (on the *Irf4*$^{-/-}$ background) with and without DOX) as indicated on the x-axis. Differential expression was sorted into 5 discrete bins, denoted in the scale bar. Using μ as mean expression and σ as standard deviation, we define a large upregulation (++) as $3(\sigma_1 + \sigma_2) < \mu_2 - \mu_1$ (which implies $\mu_2 > \mu_1$ as well, indicating an increase in expression). A small up-regulation (+) is $(\sigma_1 + \sigma_2) < \mu_2 - \mu_1 \leq 3(\sigma_1 + \sigma_2)$. No change (0) is $-(\sigma_1 + \sigma_2) \leq \mu_2 - \mu_1 \leq (\sigma_1 + \sigma_2)$. Down-regulation classifications (-- and -) are defined analogously to up-regulation classifications by reversing the order of μ_2 and μ_1. Using EMBER, the pair-wise differential expression data was integrated with ChIPseq peaks using genes within a ±100 kbp genomic interval corresponding to the Day 1 or Day 3 IRF4 (and PU.1) or IRF4 (not PU.1) peak sets (see text). Regulated target genes were inferred based on over represented patterns of differential gene expression within peak associated genes versus genome wide differential gene expression. The numbers in the upper left hand boxes of each graph indicate the number of peaks that scored above a significance threshold and were assigned at least one target

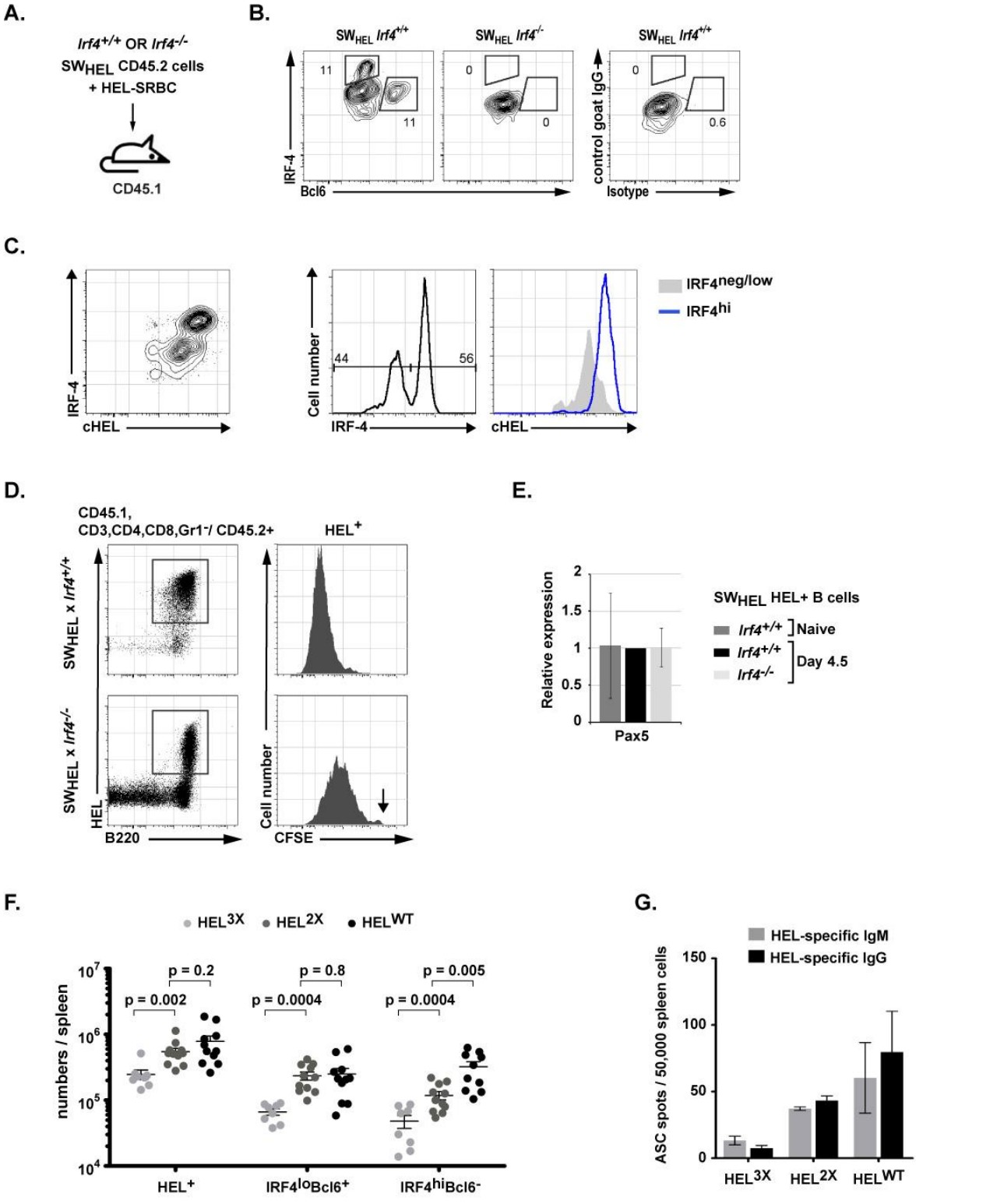
gene. The size of each bar indicates the correlation of differential expression with a given transcription factor binding event such that larger bars correspond to greater significance.

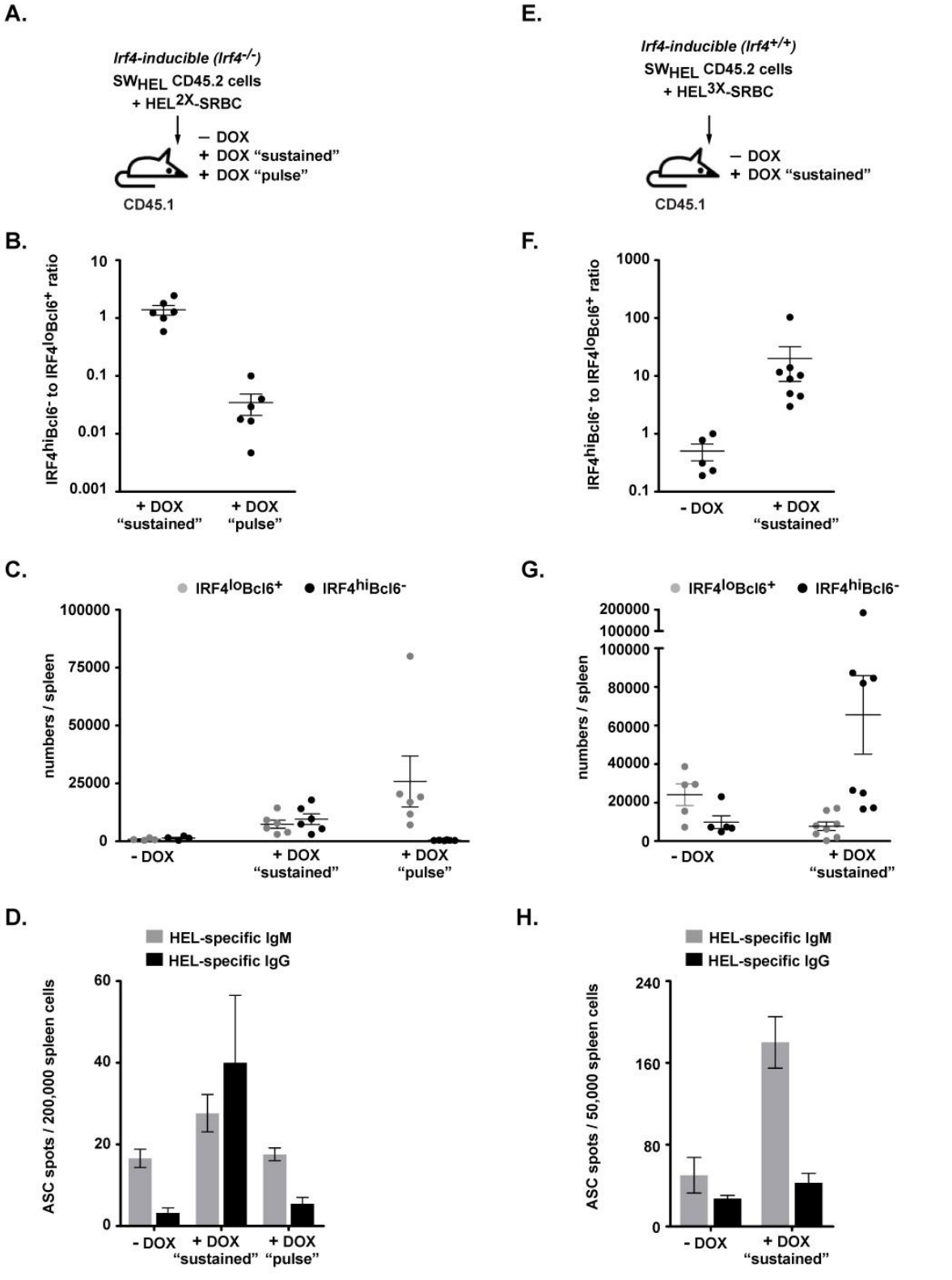
Figure S6, related to Figure 6.

Figure S7. An integrated model for the three distinct modes of IRF4 genome targeting in relation to its intracellular concentrations and B cell fate dynamics. Upon B cell activation by antigen and T-B cognate interactions, IRF4 expression is transiently activated and its concentration is regulated by strength of BCR and co-receptor signaling. Kinetic control dictates that cells experiencing a low production rate of IRF4 can undergo Ig gene diversification; either CSR or CSR and SHM, the latter within the context of the GC B cell program. On the other hand, B cells expressing high IRF4 concentrations differentiate into plasma cells at the expense of undergoing Ig gene diversification. At lower concentrations, IRF4 is proposed to assemble with either PU.1, Spi-B or BATF containing AP-1 heterodimers on EICE or AICE motifs, respectively, due to high affinity cooperative binding. The latter mode of genome targeting is transient given the signaling dependent activities of AP-1 transcription factors. These two modes of IRF4 genome targeting serve to engage B cell activation and GC B cell gene programs. As IRF4 accumulates to higher concentrations, it assembles as a homodimer on ISRE motifs and engages the plasma cell program of gene expression. During positive selection of high affinity GC B cells, BCR and CD40 signaling reactivate IRF4 expression and kinetic control is reinstated to enable plasma cell differentiation of those cells that attain the highest concentrations of IRF4.

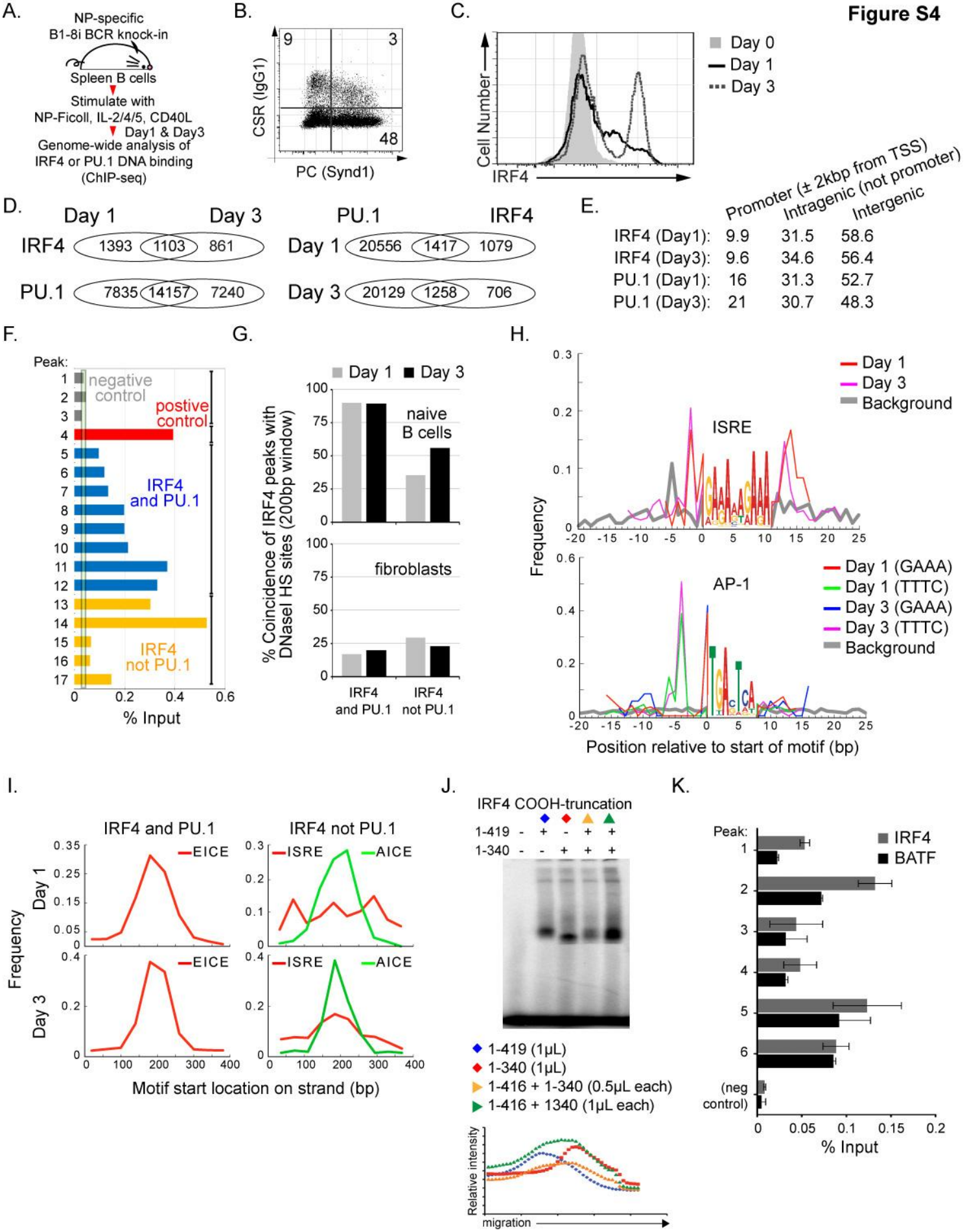
Figure S7, related to Figures 1-6.


A.
Mixed Bone Marrow
Chimera design
Group A
50% BM
Irf4+/+
CD45.1/2
1100 rads
50% BM
Irf4+/+
CD45.2
CD45.1
Group B
50% BM
Irf4+/+
CD45.1/2
1100 rads
50% BM
Irf4-/-
CD45.2
CD45.1
B.
B220+
B220+CD93-
Irf4+/-
Irf4-/-
14
80
7
89
9.4
85
9.2
85
30
66
34
62
57
39
59
37
CD93
B220
IgM
CD23
C.
Irf4+/-
Irf4-/-
B220+CD93-
Cell Number
CD23
D.
B220+
IgMhi / IgDlo
Irf4+/-
Irf4-/-
68
23
53
31
15
29
47
35
45
39
42
48
IgD
IgM
CD21
CD1d


A.
Irf4+/+ OR Irf4-/-
SWHEL CD45.2 cells
+ HEL-SRBC
CD45.1
B.
SWHEL Irf4+/+
SWHEL Irf4-/-
SWHEL Irf4+/+
IRF-4
Bcl6
control goat IgG
Isotype
11
11
0
0
0
0.6
C.
IRF-4
cHEL
Cell number
44
56
IRF-4
cHEL
IRF4neg/low
IRF4hi
D.
CD45.1,
CD3,CD4,CD8,Gr1-/ CD45.2+
HEL+
SWHEL x Irf4+/+
SWHEL x Irf4-/-
HEL
B220
Cell number
CFSE
E.
Relative expression
Pax5
SWHEL HEL+ B cells
Irf4+/+ Naive
Irf4+/+ Day 4.5
Irf4-/-
F.
HEL3X
HEL2X
HELWT
numbers / spleen
p = 0.002
p = 0.2
p = 0.0004
p = 0.8
p = 0.0004
p = 0.005
HEL+
IRF4loBcl6+
IRF4hiBcl6-
G.
HEL-specific IgM
HEL-specific IgG
ASC spots / 50,000 spleen cells
HEL3X
HEL2X
HELWT

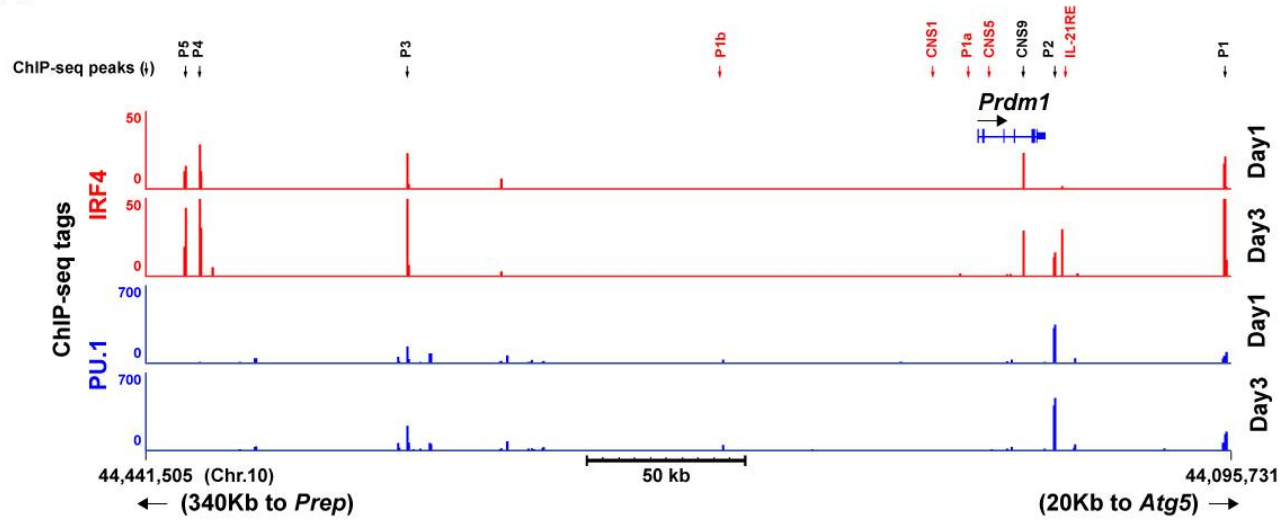

A.
Irf4-inducible (Irf4-/-)
SWHEL CD45.2 cells
+ HEL2X-SRBC
– DOX
+ DOX "sustained"
+ DOX "pulse"
CD45.1
B.
IRF4hiBcl6- to IRF4loBcl6+ ratio
10
1
0.1
0.01
0.001
+ DOX "sustained"
+ DOX "pulse"
C.
IRF4loBcl6+
IRF4hiBcl6-
numbers / spleen
100000
75000
50000
25000
0
- DOX
+ DOX "sustained"
+ DOX "pulse"
D.
HEL-specific IgM
HEL-specific IgG
ASC spots / 200,000 spleen cells
60
40
20
0
- DOX
+ DOX "sustained"
+ DOX "pulse"
E.
Irf4-inducible (Irf4+/+)
SWHEL CD45.2 cells
+ HEL3X-SRBC
– DOX
+ DOX "sustained"
CD45.1
F.
IRF4hiBcl6- to IRF4loBcl6+ ratio
1000
100
10
1
0.1
- DOX
+ DOX "sustained"
G.
IRF4loBcl6+
IRF4hiBcl6-
numbers / spleen
200000
100000
100000
80000
60000
40000
20000
0
- DOX
+ DOX "sustained"
H.
HEL-specific IgM
HEL-specific IgG
ASC spots / 50,000 spleen cells
240
160
80
0
- DOX
+ DOX "sustained"

Ochiai et al
Figure S4

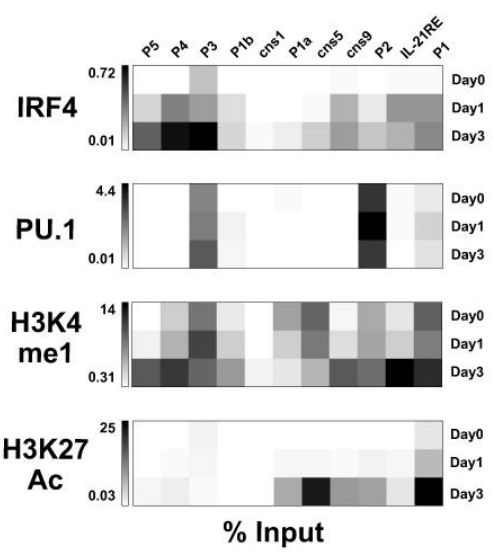


Ochiai et al
Figure S5

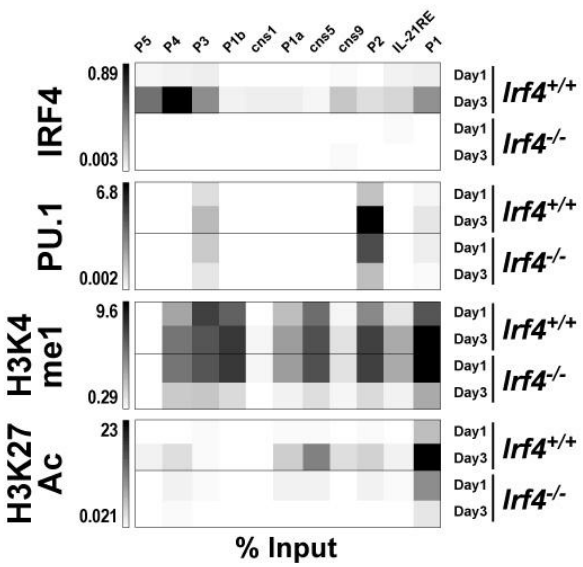
A.



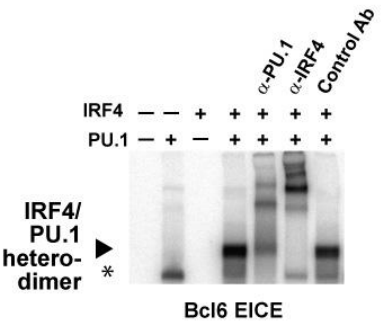
B.

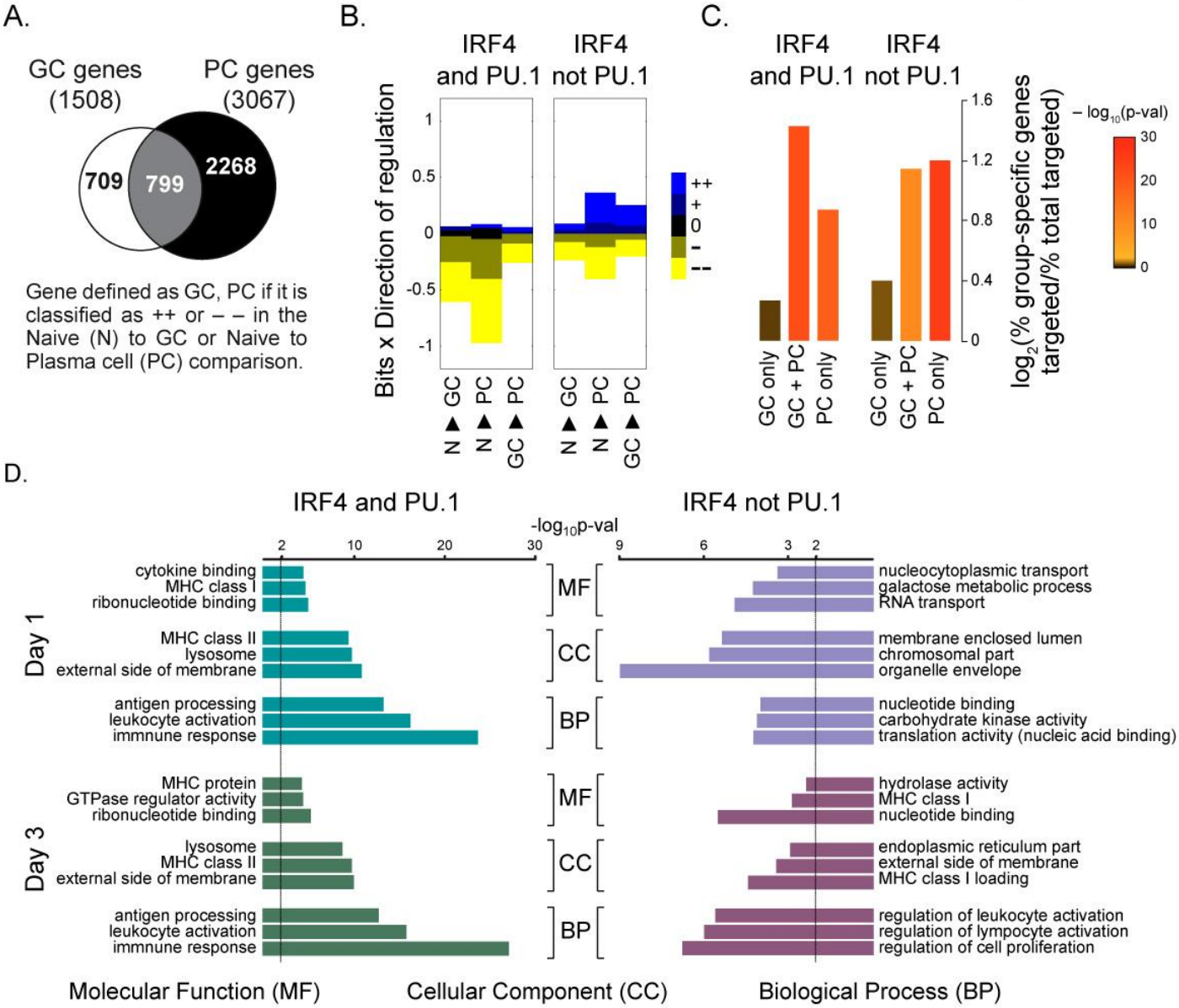


C.



D.




A.
GC genes (1508)
PC genes (3067)
709
799
2268
Gene defined as GC, PC if it is classified as ++ or – – in the Naive (N) to GC or Naive to Plasma cell (PC) comparison.
B.
IRF4 and PU.1
IRF4 not PU.1
Bits x Direction of regulation
++
+
0
–
– –
N ▲ GC
N ▲ PC
GC ▲ PC
C.
IRF4 and PU.1
IRF4 not PU.1
log2(% group-specific genes targeted/% total targeted)
– log10(p-val)
GC only
GC + PC
PC only
D.
IRF4 and PU.1
IRF4 not PU.1
-log10p-val
Day 1
Day 3
MF
CC
BP
cytokine binding
MHC class I
ribonucleotide binding
MHC class II
lysosome
external side of membrane
antigen processing
leukocyte activation
immnune response
MHC protein
GTPase regulator activity
ribonucleotide binding
lysosome
MHC class II
external side of membrane
antigen processing
leukocyte activation
immnune response
nucleocytoplasmic transport
galactose metabolic process
RNA transport
membrane enclosed lumen
chromosomal part
organelle envelope
nucleotide binding
carbohydrate kinase activity
translation activity (nucleic acid binding)
hydrolase activity
MHC class I
nucleotide binding
endoplasmic reticulum part
external side of membrane
MHC class I loading
regulation of leukocyte activation
regulation of lympocyte activation
regulation of cell proliferation
Molecular Function (MF)
Cellular Component (CC)
Biological Process (BP)
E.
IRF4 and PU.1
IRF4 not PU.1
Bits x Direction of Regulation
Day 1 peaks
Day 3 peaks
Day 0 to 1
Day 1 to 3
Day 0 to 3
Day 0
Day 1
Day 3
DOX
Day 1
DOX
Day 3
Irf4+/+
Irf4+/+
Irf4-/-
Irf4-inducible (Irf4-/-)
Irf4-/-

Ochiai et al
Figure S7

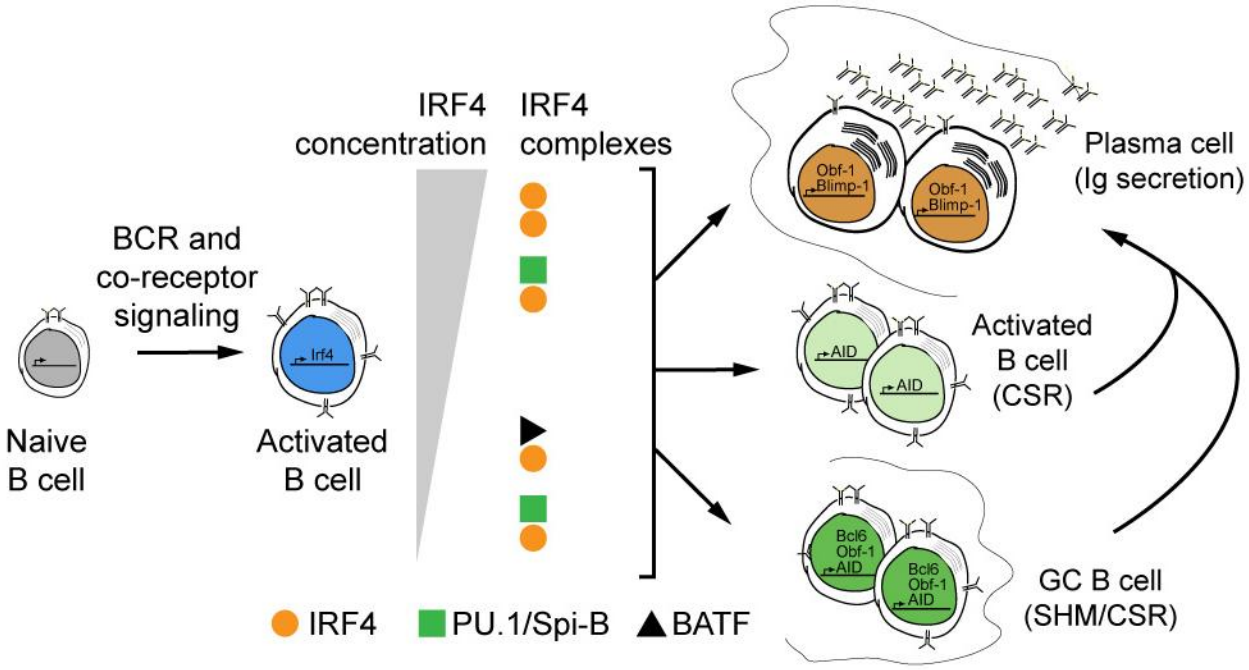


Table Legends

Table S1. Peak finding data processing statistics. Peaks were called using QuEST (Valouev et al., 2008) against an input containing 30.8 M reads (all read counts reported are the number of uniquely aligned reads to the mouse genome (NCBI37/mm9). Peaks were called at a QuEST threshold of 15, and the threshold was raised subsequently to achieve an FDR of at most 1% (in some cases the FDR was below 1% at 15). Individual replicates were combined to create a final set of peaks for each TF at each day by coincident peak analysis. For IRF4, all peaks that were coincident in at least two of three replicates were retained. For PU.1, all peaks coincident between the two replicates were retained. Two peaks were called coincident if their peak maxima were less than 200 bp apart.

Table S2. EMBER data. 13 Excel worksheets: "All genes on microarray" gives all microarray probe sets with the discrete classifications for the 10 *in vitro* behavior dimensions used by EMBER plus the 3 *in vivo* behavior dimensions used in SFigure 4. Probe sets with only a single behavior of "-1" are those that were always expressed at a low level over all conditions; these are ignored in EMBER. "ALLGENES…" worksheets give the potential targets for each of the four peak groups studied. Peak lines (5 columns starting with "chr") are immediately followed by gene lines (same data as in "All genes", starting with the probe set ID) listing the potential targets for that peak. Some peaks have zero potential targets. "TARGETS…" worksheets again list the peaks for the four peak groups, but with only those genes that are selected as EMBER targets. The first column of the gene lines gives the score for that gene. "MODEL…" worksheets give the log-odds matrix (plotted in the expression patterns) for each of the four peak groups, as well as the threshold used for determining target genes.

Table S3. Gene Ontology using DAVID. Excel file that details the statistics and genes of each category shown in Fig. S6D.

Table S4. Primers used in this study.

		Day 1			Day 3		
TF	Replicate	# reads	# peaks	FDR	# reads	# peaks	FDR
IRF4	1	15.2 M	825	0.97%	12.4 M	2784	0.90%
	2	11.4 M	9326	0.88%	14.1 M	1408	0.97%
	3	12.8 M	8929	0.86%	29.6 M	5595	0.39%
PU.1	1	12.2 M	27321	0.16%	9.5 M	34410	0.64%
	2	14.6 M	38052	0.10%	14.8 M	34061	0.07%

Validation of ChIP seq peaks by ChIP-qPCR (Fig. S4F)		
chr2:127834302-127835161	Forward	GTGTGTGTGTGCTGGGGCGT
	Reverse	CCGCTGTACCGGATCACCCCT
chr13:5863844-5864107	Forward	TGGTTCCTTTTGCACAGACCTGCC
	Reverse	TGGTCGTGGGTGAAAGTTCCTCT
chr1:108464037-108464433	Forward	AGGAGGGTTGGGCATAAGGCT
	Reverse	GGAAATGCCTGCAAGCGGCAAG
chr9:32459894-32460227	Forward	TCTCCCTCTCCTCCCCTCCCT
	Reverse	AGGTGGCTGGAGGTTTAGTGTTCA
chr4:44683859-44684229	Forward	CATACAGCCCTTAGCATCTC
	Reverse	GAGTTCTGTCTAGCCGAAGA
chr4:32488557-32488800	Forward	TGTAGAGGGAACTGATGACTCAAA
	Reverse	GCTCATCTGAGTGTAATCTGGGTA
chr18:69804123-69804696	Forward	TCTTAGCACCCAGTGAAGTT
	Reverse	TCACGTACAAGGATGAATGA
chr16:11385880-11386238	Forward	TGGGCAATGATTTAATTTCT
	Reverse	CCCTGGACACAGTTCAGTAT
chr10:44428699-44429007	Forward	TGTATCATCAAGGCAAATGA
	Reverse	CAAGAAAGGTCACCCACTAA
chr13:30824584-30824875	Forward	AGCGCACATGCAGGCATGGGAAA
	Reverse	CACTCACTATCACCACTCGCGGGCA
chr6:47571104-47571526	Forward	GACTCACAGCCAGAGATAGC
	Reverse	GCTCACAGGAAACTGAAGAC
chr7:71075755-71076040	Forward	AAGTAGGCAGAGTCTCACCA
	Reverse	CAGAGTATCCAGAAGGAACG
chr16:24151472-24151830	Forward	AAGTTGAAAGGGAAGAATCC
	Reverse	AAGGTGCAATCAGCATAAAT
chr6:122520252-122520386 (negative control)	Forward	TGGGAGCCCAGGCAGAAAGA
	Reverse	CGTCTTCAGGAAGAGCTGCCTTT
chr10:44182361-44182566 (negative control)	Forward	CCCTGACAATGTTTGTCCTA
	Reverse	CGGACAAGAAGAGCAAGTTA
chr10:21862750-21862977 (negative control	Forward	TGTGAAGGCCTACTTCAGAT
	Reverse	CCAAGCCTATTGCAGATTAC
chr6:70685455-70685889 (positive control; IgKappa 3' Enhancer)	Forward	TCATAGCTACCGTCACACTG
	Reverse	AACAGATGTGCCTAAGGTTC
Blimp-1 locus ChIP-qPCR (Fig. S5)		
P1	Forward	ACAGAGCTCAGCTGAAACAT
	Reverse	CTAGCTGCTCTGAGTGAACC
IL21RE	Forward	ACCGTTGAAAGACGGTGACT
	Reverse	AATGGCCCACATCATCTGTG
CNS9	Forward	AAGCTGCTAAGTGGGAGAGT
	Reverse	GCATAATTTAGCGTTTGGTG
CNS5	Forward	CTCGCTGCCTTTACATTACT
	Reverse	GCTGGTCTTTACAATTCGTC
P1a	Forward	AGATTCCCAGTCCTTGTTG
	Reverse	GACGGTCTGATTCACTCCT
CNS1	Forward	CCCTGACAATGTTTGTCCTA
	Reverse	CGGACAAGAAGAGCAAGTTA
P1b	Forward	CCCATTTCTTAAGGTAGCAA
	Reverse	CGAGAAAGCTTAGTTTGCAT
P3	Forward	GCCAATGCTGAATTACTTTC
	Reverse	GAATAATTGCCATACCCAGA
P4	Forward	GCCACATCATATGGAAAGTT
	Reverse	TTCAATAAGGACACCTCCAC
P5	Forward	TGTATCATCAAGGCAAATGA
	Reverse	CAAGAAAGGTCACCCACTAA
P2	Forward	TATCTGCCACTTCCTCTTTC
	Reverse	ATGCAAATCTTCTCTGCTGT
RT-qPCR of GC B cell transcripts (Fig. 2B and 3C)		
AID	Forward	AGGGAGTCAAGAAAGTCACG
	Reverse	CAGGAGGTGGCACTATCTCT
Blimp-1	Forward	CCCTCTGAAGAAACAGAATG
	Reverse	GCTTGTGCTGCTAAATCTCT
Bcl6	Forward	GCAGTTTAGAGCCCATAAGA
	Reverse	GTACATGAAGTCCAGGAGGA
Obf-1	Forward	CCAGTGAAGGAGCTACTGAG
	Reverse	GTCACTGTGGAAGCAGAAAC
Pax5	Forward	CGTGTTTGAGAGACAGCAC
	Reverse	AAGTTGGCTTTCATGTCATC
Oct1	Forward	GAGAAGTGGCTAAATGATGC
	Reverse	CATATTGAGCTGTTCAGCAA
EMSA		
EICE (IgKappa 3' Enhancer)		AAGACCCTTTGAGGAACTGAAAACAGAACC
ISRE (Blimp-1 CNS9)		CAACTGAAACCGAGAAAGC
ISRE MT1		CAACTCGTACCGAGAAAGC
ISRE MT2		CAACTGAAACCCGTAAAGC
ISRE MT1/2		CAACTCGTACCCGTAAAGC
EICE (Bcl6)		TGACAAGAGGGAAGTGAAAGTGCCCATTG

Experimental Procedures

Mice

The generation and genotyping of the *Irf4*$^{-/-}$ and *Irf4*-inducible mice has been previously described (Mittrucker et al., 1997; Sciammas et al., 2011). The B1-8i gene targeted mice were a generous gift of K. Rajewsky (Sonoda et al., 1997). SW_{HEL} mice have been previously described (Phan et al., 2003) and were used to cross to *Irf4*$^{-/-}$ mice. Conditional *Irf4*$^{fl/fl}$ mice were crossed to *CD19*$^{+/CRE}$ mice. Mice were housed in specific pathogen-free conditions and were used and maintained in accordance of the Institutional Animal Care and Use Committee guidelines.

Cells and culture conditions

Splenic B cells were isolated by sequential hypotonic lysis and magnetic bead separation. Briefly, a single cell suspension of spleen cells was treated with Tris buffered ammonium chloride to lyse red blood cells and then a negative selection, bead-based immunopurification strategy was employed that depletes T, NK, and macrophage cells using a cocktail of biotinylated antibodies specific for CD4, CD8, CD3, CD49b, Gr1, CD11b, Ter119 (BD Biosciences) and adsorption to Streptavidin microbeads (Miltenyi Biotec Inc.). The unbound cells (typically greater than 90% B220 positive) were washed and placed into culture at cell densities ranging between 0.25 to $0.5x10^6$ cells/ml, depending on the experiment. Cells were stimulated with recombinant human IL-2 (Roche) 100U/ml, recombinant murine IL-4 (R&D Systems) 5ng/ml, recombinant mouse IL-5 (R&D Systems) 1.5ng/ml, as indicated. CD40L membranes were prepared from insect cells as described (Kehry and Castle, 1994) and used at 0.1µg/ml. NP(12)-Ficoll was obtained from BioSearch Technologies Inc. used at 0.01 ng/ml. In some cases cells were labeled with CFSE (Invitrogen) at 1.5uM prior to culturing.

Flow cytometry

Data were collected on the LSRII (BD) and was analyzed using FlowJo software (Tree Star, Inc.).

For quantitation of plasma cells and cells that had undergone CSR in stimulated B1-8i cells: Cells were washed in staining buffer, PBS containing BSA (0.5%, w/v) and sodium azide (0.05%, w/v) and blocked with anti-FcR (clone 2.4G2) hybridoma supernatant. Anti-mouse antibodies specific for Syndecan-1, IgG1, and isotype controls were purchased from BD. Dead cells were excluded from the analysis using DAPI (Invitrogen).

For measurement of intracellular IRF4 and Bcl6: When analyzing stimulated B1-8i cells, Intracellular detection of IRF4 was performed by fixing cells in 1% (w/v) paraformaldehyde for 10 minutes and then permeabilizing with PBS-based buffer containing 0.03% (w/v) saponin (Sigma). Staining buffer containing 0.3% Saponin was used to incubate the cells with a polyclonal goat anti-IRF4 or normal goat IgG (Santa Cruz Biotechnology) in the presence of 5% (v/v) donkey and rabbit serum. IRF4 labeling was detected using a Cy5-coupled donkey anti-goat antibody (Jackson Immunoresearch). When analyzing SW_{HEL} responder cells, after surface staining,

cells were fixed for 30 min and then permeabilized with the Fix/Perm staining kit (eBioscience). Cells were then stained with anti-Bcl6 (BD) and IRF4, as above.

For analyzing B cells after *in vivo* immunizations: Splenocytes from SRBC immunized mice were prepared and washed in PBS-based staining buffer containing 2% FCS. Cells were stained with anti-CD45.1, CD45.2, B220, CD95, and GL7 (or PNA), all from BD, or PNA-fitc (Sigma). In adoptive transfer experiments using SW_{HEL} B cells, two sequential steps were used to identify the HEL binding B cells; i) incubation with unlableled HEL (Sigma) and then after washing ii) detection with anti-HEL using fluorochrome-conjuated HyHEL9 mAb which detects an epitope of HEL that does not compete with that of the knock-in B cells (based on the HyHEL10 monoclonal antibody). During this second step, fluorochrome-conjugated anti-CD45.1, CD45.2, and B220 antibodies (from BD) were used. After fixation and permeabilization, cells were stained for IRF4 and Bcl6 (or controls) as described above.

For sorting GC B cells: 10 mice were immunized with SRBC and 7 days later RBC-depleted splenocytes were prepared. Next, a magnetic bead enrichment was performed (negative selection) by collecting cells that were not bound by anti-biotin magnetic beads bound to anti-IgD, anti-CD11c, or anti-CD43 using the EasySep system (StemCell Technologies). Unbound cells were then labeled with fluorochrome-conjugated antibodies specific for B220, CD95, and GL7 as well as fluorochrome-conjugated Streptavidin to gate out residual biotin-labeled cells. DAPI was used to gate out dead cells. After sorting on a FACSAria, cells were washed and then processed for ChIP-qPCR, see below.

Generation of 50%:50% mixed bone marrow chimeric mice and SRBC immunization

Bone marrow was collected from the long leg bones of wildtype CD45.1/2 heterozygous mice and from $Irf4^{+/+}$ or $Irf4^{-/-}$ (both CD45.2) mice and mixed together in equal numbers. Cells were injected into irradiated (2 x 550 rads; the night and morning before adoptive transfer, respectively) wildtype CD45.1 mice within 3hrs of the second irradiation. Eight weeks following adoptive transfer, mice were immunized with sheep RBC (Lampire Biologicals) i.p. and the GC response was quantified by flow cytometry seven days later.

Adoptive transfer of SW_{HEL} B cells

Detection of donor SW_{HEL} responder cells was performed by gating on $B220^{+}CD45.2^{+}CD45.1^{-}$ cells that bound HEL antigen.

Wild type SW_{HEL} experiments: Splenocytes from SW_{HEL} mice (CD45.2) were prepared and the frequency of HEL binding cells was determined via a two step flow cytometric method (see above). 50,000 HEL+ cells were injected into wildtype CD45.1 hosts in combination with 200 million SRBC (Lampire Biologicals) conjugated via EDCI (Sigma) to either HEL^{3X}, HEL^{2X}, or HEL^{WT} (prepared fresh). 4.5 days post immunization, splenocytes were isolated and processed for flow cytometry and HEL-specific ELISpot.

Irf4$^{-/-}$ SW_{HEL} experiments: SW_{HEL} mice were crossed to *Irf4*$^{-/-}$ mice and used for adoptive transfer experiments as above. Briefly, splenocytes from SW_{HEL} mice (CD45.2) that were either *Irf4*$^{+/+}$ or *Irf4*$^{-/-}$ were prepared and the frequency of HEL binding cells was determined via a two step flow cytometric method (see above). For CFSE labeling, the splenocytes were washed with PBS and loaded with 1.0 μM CFSE (Invitrogen) prior to adoptive transfer. 0.5 million HEL+ cells were injected into wildtype CD45.1 hosts in combination with 200 million SRBC (Lampire Biologicals) conjugated to HEL^{2X} (prepared fresh). 4.5 days post immunization, splenocytes were isolated and labeled in order to sort donor HEL binding cells using anti-CD45.1, CD45.2, B220, and HEL + anti-HEL (HyHEL9). Cells were sorted on a FACSAria (BD).

Irf4-inducible SW_{HEL} experiments: SW_{HEL} mice were crossed to Irf4-inducible mice either on the *Irf4*$^{+/+}$ or *Irf4*$^{-/-}$ background and used for adoptive transfer experiments as above except that B cells were first enriched by a negative selection based bead method (Sciammas et al., 2011). When *Irf4*$^{+/+}$ Irf4-inducible SW_{HEL} B cells were used, 50,000 HEL^{+} cells were transferred and when *Irf4*$^{-/-}$ Irf4-inducible SW_{HEL} B cells were used, 500,000 HEL^{+} cells were transferred along with HEL conjugated SRBC. Doxycycline (DOX, 0.5mg/mL) was administered in the drinking water that was sweetened with sucrose (1% w/v). Recipient mice were fed DOX or control water for 1-2 days prior to transfer of Irf4-inducible B cells and water was changed every two days.

RNA preparation, microarray, and RT-PCR

For microarray: Total RNA was prepared from triplicate cell samples using Trizol. 250 ng were used to generate cRNA for hybridization to Affymetrix mouse 430A chips, using standard procedures. Signal intensities were normalized using D-Chip using the PM-only model and the output was used to quantify differential gene expression between groups.

For RT-qPCR: Total RNA from immunized SW_{HEL} B cells was prepared by sorting 5,000 CD45.2+, B220+, HEL+, CD45.1- cells directly into RLT buffer from the RNeasy Micro Kit (Qiagen). RNA was isolated using the manufacturer's recommendations and cDNA was prepared using SuperScript II (Invitrogen). 500 cell equivalents was used for each SYBR green based quantitative PCR analysis using 2X master mix (Clontech). Each primer pair was analyzed in duplicate reactions. Oct1 values were used to normalize expression levels of the other transcripts.

Electro-mobility shift assays, 293T cell transfection, Nuclear Extracts, and in vitro translation

293T cells were transfected with expression constructs for *Irf4* and PU.1 or a non-insert control using Fugene (Roche) according to manufacturer's recommendations. Nuclear extracts were then prepared as previously described and used to prepare protein – DNA complexes. Reactions were separated on non-denaturing PAGE and the radioactive signal was imaged and quantitated from a Phosphorimager (Molecular Dynamics). In some cases (Fig. S4J and K) the source of the protein for preparing protein – DNA complexes was derived from *In*

vitro Translation. In this case, the T7 phage polymerase based, coupled transcription and translation kit (Promega) was used. The amount of material used in the EMSA was standardized in control reactions.

Chromatin precipitation for PCR and massively parallel sequencing

ChIP-qPCR was performed as previously described. Chromatin from *ex vivo* sorted GC B cells, stimulated B1-8i, *Irf4*$^{+/+}$, or *Irf4*$^{-/-}$ B cells was isolated and then sonicated in a Biorupter (Diagenode Inc.) to obtain DNA fragments ranging in size from 100 to 500 bp. 25ug of chromatin fragments were immunoprecipitated using control IgG (Santa Cruz, sc-2027), anti-IRF4 (Santa Cruz, sc-6059), anti-PU.1 (Santa Cruz, sc-352), anti-H3K4me1 (Abcam, ab8895), and H3K27Ac (Abcam, ab4729), or anti-Batf (prepared in house). After reversal of formaldehyde crosslinks, specific DNA sequences were assessed by quantitative real-time PCR. Primers used for PCR are listed in Table S4. For ChIPseq, stimulated B1-8i B cells were used to prepare sheared chromatin, as above. 200ug of chromatin fragments were immunoprecipitated using anti-IRF4 (Santa Cruz, sc-6059) or anti-PU.1 (Santa Cruz, sc-352) and DNA libraries were then prepared using ChIPseq sample prep kit (Illumina, 1003473). Input DNA was processed in parallel to obtain a "background" sample for enrichment processing (see below). Libraries were amplified for 12-14 cycles using Illumina oligos and product ranging in size between 200-350 bp was gel-purified prior to sequencing. DNA was sequenced using the Illumina Genome Analyzer II and the Solexa data processing pipeline was used to process the images into individual sequences which were then subsequently aligned to the mm9 build of the mouse genome using ELAND software. Sequences that exhibited more than two mismatches per sequence tag or that mapped to multiple genomic locations were eliminated before the next step of peak finding (see below).

Computational analysis

Peak finding: ChIP-seq peaks were called using the QuEST software tool (ref). Seqeunces from input chromatin was used for background control and represented a concatenation of several experiments, with a total of 30.8 M uniquely aligned reads. We ran QuEST with a threshold of 15 and performed a post-peak calling analysis to obtain a false discovery rate (FDR) of 1%. Briefly, we compared the ratio of the number of psuedo-ChIP peaks (called from the input) to the number of ChIP peaks at successively higher thresholds. This ratio is the FDR, and we chose a threshold that obtained an FDR of 1%; for the PU.1 data sets, the threshold of 15 resulted in an FDR of < 1%, so 15 was used.

A final set of peaks for each TF was obtained by combining replicate ChIP-seq experiments, two replicates for PU.1 and three for IRF4. Final PU.1 peaks were defined as those peaks that were coincident between both replicates. Final IRF4 peaks were defined as those peaks that were coincident between at least two of the three replicates. These final sets were used to define coincident peak groups between PU.1 and IRF4 at different time points in Fig. 4A and B. Peaks are considered coincident if their peak maxima are within 200 bp; analyses constructing coincident and non-coincident peak groups were carried out using in-house software. Comparison

of peak positions with genome annotations were performed using the coordinates of UCSC genes (NCBI37/mm9). Schematic representations of peaks in relation to genome annotations were generated using the Integrated Genome Browser (IGB) v6.5.

DNA Motif and TFBS analysis: We used the MEME software tool (Bailey and Elkan, 1994) to search for over-represented motifs in four coincident peak groups: IRF4 (AND PU.1) cobound regions at Day 1 and Day 3, and IRF4 (not PU.1) regions at Day 1 and Day 3 (see Fig. 4A and B). We obtained the top 10 motifs. The EICE motif (Fig. 4A) was the top hit for both cobound peak sets. For the IRF4 (not PU.1) peak sets, the ISRE motif was the tenth and first hit and the AP-1 motif was the third and seventh hit at Day 1 and Day 3, respectively.

Expectation Maximization of Binding and Expression Profiles (EMBER): We use Expectation Maximization of Binding and Expression pRofiles (EMBER) (Maienschein-Cline et al., 2011) to infer the genes targeted by each peak. In the first step of EMBER, discrete expression profiles are made, briefly described here. Ten "behavior dimensions" were defined by comparing expression levels at different pairs of DNA microarray conditions. These pairs are (1) WT Day 0 to WT Day 1; (2) WT Day 1 to WT Day 3; (3) WT Day 0 to WT Day 3; (4) Day 0 *Irf4*$^{-/-}$ to WT; (5) Day 1, *Irf4*$^{-/-}$ to WT; (6) Day 3, *Irf4*$^{-/-}$ to WT; (7) Day 1, *Irf4*$^{-/-}$ to Irf4-inducible (*Irf4*$^{-/-}$)+low Dox; (8) Day 1, *Irf4*$^{-/-}$ to Irf4-inducible (*Irf4*$^{-/-}$)+high dox; (9) Day 3, *Irf4*$^{-/-}$ to Irf4-inducible (*Irf4*$^{-/-}$)+low Dox; (10) Day 3, *Irf4*$^{-/-}$ to Irf4-inducible (*Irf4*$^{-/-}$)+high dox. Behavior dimensions 5-10 are plotted in Fig. 6A. Overall, the contribution of the full expression pattern with all 10 behavior dimensions is given in Fig. S6. For each behavior dimension, the change in expression level for each probe set is discretely classified into one of five categories, denoted in the scale bar. Using μ as mean expression and σ as standard deviation, we define a large upregulation (++) as $3(\sigma_1 + \sigma_2) < \mu_2 - \mu_1$ (which implies $\mu_2 > \mu_1$ as well, indicating an increase in expression). A small up-regulation (+) is $(\sigma_1 + \sigma_2) < \mu_2 - \mu_1 \leq 3(\sigma_1 + \sigma_2)$. No change (0) is $-(\sigma_1 + \sigma_2) \leq \mu_2 - \mu_1 \leq (\sigma_1 + \sigma_2)$. Down-regulation classifications (-- and -) are defined analogously to up-regulation classifications by reversing the order of σ_2 and σ_1.

In the second step of EMBER, potential targets are determined for each peak by considering all probe sets lying within 100 kbp of the peak; Affymetrix coordinates for the probe sets were used here, and each peak set from coincident peak analysis is run separately. In the third step of EMBER, an unsupervised machine learning algorithm (expectation maximization) is used to find an over-represented set of expression profiles, called the expression pattern, among the potential targets for each peak set. After convergence to an expression pattern, gene targets can be determined by scoring each probe set's expression profile against the expression pattern. There are two results of EMBER: the expression pattern, and a list of target genes. Expression patterns for four coincident peak groups are shown in Fig. 6A, and gene ontology analysis for the target genes is shown in Fig. S6. Expression pattern images are prepared as described previously (Maienschein-Cline et al., 2011).

In our analysis, roughly half of all genome targeting events (compare to the number of peaks in Fig. 4A and 4B) scored above the threshold. Even in the absence of experimental limitations, functional peaks might not be recognized as such because the associated gene expression changes too loosely conformed to the major behavior patterns. In addition, the method cannot detect target regions on transcriptional units that are not represented by probes on the microarray (non-coding RNAs and microRNAs) and effects at a range longer than the selected range (100 kbp).

Comparison of IRF4 (and PU.1) and IRF4 (not PU.1) target genes with GC B cell and plasma cell expression signatures: We obtained DNA microarray data (GEO database, NCBI) for Naïve B cells, Germinal Center B cells (GC), and Plasma cells (PC) (Luckey et al., 2006). We defined three additional behavior dimensions, Naïve to GC, Naïve to PC, and GC to PC, and discretized the behavior of each gene in these dimensions, following the procedure in our EMBER description. Using these new dimensions, we defined three new gene groups, shown in Fig. S6A, B and C: GC only genes, which were ++ or -- in the Naïve to GC dimension but not in the Naïve to PC dimension, PC only genes, which were ++ or -- in Naïve to PC but not Naïve to GC, and GC+PC genes, which were ++ or -- in both Naïve to GC and Naïve to PC. For each gene group and for the two Day 3 peak groups discussed, we computed the EMBER log-odds matrix for the three new behavior dimensions. We also quantify the over-representation of each gene group among the EMBER targets for a peak group by a log-ratio. The ratio compares the percentage of potential targets (i.e., genes within 100 kbp of a peak) belonging to the gene group that are also EMBER targets to the overall percentage of potential targets that are EMBER targets. A *p*-value for each log-ratio is computed from the binomial distribution, based on the null hypothesis that genes from a particular gene group (e.g., GC only) are chosen as EMBER targets at a rate equal to the overall percentage of potential targets that are EMBER targets.

Gene ontology analysis: The EMBER-assigned targets for each peak set were analyzed for enrichment of gene ontology terms using **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (**DAVID)** v6.7 (Huang et al., 2009a; Huang et al., 2009b). The background gene list used to compute statistical enrichment consisted of all mouse genes represented in Affymetrix arrays. Enrichment of each gene cluster was ranked by DAVID and the corresponding p-value is reported. The top 3 ontology clusters corresponding to Molecular Functions (MF), Cellular Component (CC) and Biological Process (BP) are reported.

DNaseI genomic region comparison: We obtained the narrowPeak DNase-seq processed data for CD19+ and CD43- B cells, as well as fibroblasts (as a negative control), from ENCODE (http://genome.ucsc.edu/ENCODE/). For each of our four peak groups (Fig. 4A and B), and each DNase-seq replicate, we calculated the percentage of peaks with the center of a DNaseI region within 200 bp of the peak. We then averaged the percentage overlap across the replicates and combined the results for CD19+ and CD43- B cells. We also calculated the same percentage for the peaks with EMBER-selected targets (those peaks are

indicated in boxes in Fig. 6A). We found a fairly uniform increase of about 3 percentage points in the overlap when peaks were selected by EMBER (results not shown), which is consistent with the observation that EMBER selects target genes preferentially for functional binding sites.

References

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings / International Conference on Intelligent Systems for Molecular Biology ; ISMB *2*, 28-36.

Huang, D., W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1-13.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Kehry, M.R., and Castle, B.E. (1994). Regulation of CD40 ligand expression and use of recombinant CD40 ligand for studying B cell growth and differentiation. Semin Immunol *6*, 287-294.

Kwon, H., Thierry-Mieg, D., Thierry-Mieg, J., Kim, H.P., Oh, J., Tunyaplin, C., Carotta, S., Donovan, C.E., Goldman, M.L., Tailor, P., *et al.* (2009). Analysis of interleukin-21-induced Prdm1 gene regulation reveals functional cooperation of STAT3 and IRF4 transcription factors. Immunity *31*, 941-952.

Luckey, C.J., Bhattacharya, D., Goldrath, A.W., Weissman, I.L., Benoist, C., and Mathis, D. (2006). Memory T and memory B cells share a transcriptional program of self-renewal with long-term hematopoietic stem cells. Proc Natl Acad Sci U S A *103*, 3304-3309.

Maienschein-Cline, M., Zhou, J., White, K.P., Sciammas, R., and Dinner, A.R. (2011). Discovering Transcription Factor Regulatory Targets Using Gene Expression and Binding Data. Bioinformatics.

Mittrucker, H.W., Matsuyama, T., Grossman, A., Kundig, T.M., Potter, J., Shahinian, A., Wakeham, A., Patterson, B., Ohashi, P.S., and Mak, T.W. (1997). Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function. Science *275*, 540-543.

Phan, T.G., Amesbury, M., Gardam, S., Crosbie, J., Hasbold, J., Hodgkin, P.D., Basten, A., and Brink, R. (2003). B cell receptor-independent stimuli trigger immunoglobulin (Ig) class switch recombination and production of IgG autoantibodies by anergic self-reactive B cells. J Exp Med *197*, 845-860.

Sciammas, R., Li, Y., Warmflash, A., Song, Y., Dinner, A.R., and Singh, H. (2011). An incoherent regulatory network architecture that orchestrates B cell diversification in response to antigen signaling. Mol Syst Biol *7*, 495.

Sonoda, E., PewznerJung, Y., Schwers, S., Taki, S., Jung, S., Eilat, D., and Rajewsky, K. (1997). B cell development under the condition of allelic inclusion. Immunity *6*, 225-233.

Valouev, A., Johnson, D.S., Sundquist, A., Medina, C., Anton, E., Batzoglou, S., Myers, R.M., and Sidow, A. (2008). Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods *5*, 829-834.