Supplemental Materials

Supplemental Methods

Supplemental Figures 1 to 7

Supplemental Figure legends 1 to 7

Supplemental Table 1 to 8

Supplemental Methods

Vector constructions for GFP reporter transgenic mice

Genomic regions D, E, F, G, H and I were PCR amplified from Ik-BAC-3 vector and cloned as *SacII/SacI* fragments into B-p-GFP, downstream of the *EGFP* reporter. For the J-B-p-GFP construct, genomic region J was PCR amplified from Ik-BAC-3 vector and introduced into B-p-GFP upstream of B, as J is located 5' of the B region in the endogenous *Ikzf1* locus. *SstI* (*SacI* isoschizomer) sites within the D and E genomic regions of B-p-GFP-D and B-p-GFP-E, respectively, were trashed by *SacI* partial digestion and re-ligation, preserving the unique 3' *SacI* site for excision of the transgene from the vector backbone prior to microinjection. Primers for amplifying each genomic region used in the GFP transgenic constructs are provided in Supplemental Table 4. B-p-GFP-DI and B-p-GFP-DIII were generated using ~500 bp fragments from the 5' (DI) and 3' (DIII) ends of the 1.7 kb D region.

The *Ikzf1* mini-cassette (Ik-MC2), containing all putative regulatory regions (J, D, E, F, G, H and I) cloned into B-p-GFP, was generated as follows. First, intermediate construct #1 was generated, consisting of the D, E, and F regions (PCR-amplified from Ik-BAC-3 (for F), B-p-GFP-D (for D) or B-p-GFP-E (for E) with appropriate restriction sites at the 5' and 3' ends) cloned into pBlueScript vector. Separately, regions G, H and I (all PCR-amplified from Ik-BAC-3 with 5' and 3' restriction sites) were cloned into pBlueScript to generate intermediate construct #2. The G-H-I fragment was excised from intermediate construct #2 and ligated into intermediate construct #1, downstream of region F. Finally, the D-E-F-G-H-I fragment was cut out and inserted into the *SacII/SacI*

site of J-B-p-GFP to generate Ik-MC2. All genomic regions were PCR amplified using the same annealing sequences and templates used in the GFP transgenic constructs.

Constructs for recombination templates for BAC Engineering

The DNA recombination template for targeting the human *CD2* (hCD2) reporter gene into Ik-BAC-3 was excised from IkAB/pQS1, a construct derived from the pQS1 plasmid (kind gift from Dr. M. Busslinger). The pQS1 construct contains the following sequences, listed from 5' to 3', subcloned into the pSP64 vector: (1) the *IRES* from encephalomyocarditis virus, (2) a cDNA encoding the extracellular and transmembrane domains of *CD2* with SV40 poly (A) sequence, and (3) a cDNA encoding neomycin (kanamycin resistance) flanked by *frt* sites to enable Flp-mediated deletion of the neomycin (*neo*) gene in bacteria (Q. Sun and M. Busslinger, unpublished data). IkAB/pQS1 was constructed by inserting into pQS1 (1) a 180 bp 5' homology arm (IkA) at the XmaI/EcoRI site upstream of the *IRES* sequence, and (2) a 301 bp 3' homology arm (IkB^1) at the *NsiI/BspEI* site downstream of *frt-neo-frt*. IkA consists of *Ikzf1* genomic sequence that begins 180 bp upstream of and disrupts the translation start site in exon 2 by terminating at "A" of the ATG. IkB^1 consists of *Ikzf1* genomic sequence that begins 60 bp downstream of exon 2 and extends 300 bases downstream.

The recombination template for replacing the endogenous pBeloBAC11 *loxP* site with an ampicillin resistance gene (*amp^R*) was PCR amplified from the pTamp plasmid (kind gift from N. Copeland) using the indicated primers. The pTamp

plasmid contains the *amp* gene (from pEGFP-1, Stratagene) along with 920 bp of 5', and 370 bp of 3', pBeloBAC11 vector sequence flanking the *loxP* site.

To generate the DNA recombination template for knock-in of the *loxP*-flanked D region into Ik-BAC-hCD2, the D-LoxP/pBS construct was created by insertion of the following sequences, listed from 5' to 3', into the pBluescript II KS(-) vector: (1) 5' homology arm (UpD[2]) consisting of the 321 bases immediately upstream of the *Ikzf1* D region, (2) D region[3] (1.7 kb) flanked by *loxP* sites, (3) *neo* cDNA flanked by *frt* sites, and (4) 3' homology arm (DownD) consisting of the 328 bases immediately downstream of the *Ikzf1* D region.

For electroporations, recombination templates excised from plasmids were purified from 0.8-1% agarose gel fragments using the PerfectPrep Gel Cleanup kit (Eppendorf) and resuspended in sterile water. PCR-generated recombination templates were treated with *DpnI* restriction enzyme (to eliminate residual template vector from the PCR reaction), purified using the QIAQuick PCR Purification kit (Qiagen) and resuspended in sterile water. The primer sequences and more detailed strategy to generate these constructs are available upon request.

(Notes)

[1] IkB was amplified from Ik-BAC-3 using primers 5'-cta gca agg ctg cca taa aag aaa-3' and 5'-cac gtg gca agt gct ctg ac-3', which contain *NsiI* and *BspEI* restriction sites, respectively, at their 5' ends.

[2] UpD was PCR amplified from Ik-BAC-3 using primers 5'gag ctt tcg agt tca cac agg and 5'ct cct ctt tgt ttc tct cag ta, which contain *XbaI* and *SpeI* cleavage sites , respectively, at their 5' ends. The primer with the 5' *SpeI* site also contains the 34-bp canonical *loxP* sequence (5'ata act tcg tat aat gta tgc tat acg aag tta t) immediately upstream of the annealing sequence.

[3] The D region was amplified from Ik-BAC-3 using 5'gga gtc caa gat taa aag gcc (with SpeI site at 5' end) and 5'tc att att aca cat tgg ggt gcc (with *EcoRI* site and *loxP* sequence at 5' end).

Generation of BAC Transgenic Mice

The original BAC DNA (Ik-BAC-3) was obtained from Genome Systems, Inc. (St. Louis, MO) as a pBeloBAC 11 vector-based clone in DH10b *E. coli* cells. The pBeloBAC11 vector contains a chloramphenicol resistance (Cm^R) selection marker. To confirm that this BAC encompasses the entire *Ikzf1* locus, PCR reactions were performed using primer sets that amplify sequences from the 5' end (J region), middle (exon 4) and 3' end (exon 8) of the *Ikzf1* gene. *Ikzf3* (Aiolos gene)-specific primers were used in the negative control PCR reaction. Restriction mapping revealed that Ik-BAC-3 originates ~67 kb upstream of untranslated exon 1a and extends ~20 kb downstream of translated exon 8.

BAC modifications were performed using EL250 cells (kind gift from Dr. N. Copeland), a DH10b-based *E. coli* strain adapted for highly efficient homologous recombination of linear targeting cassettes into BAC DNA. EL250 cells use a defective l phage to supply temperature-regulated recombinogenic functions

(encoded by *exo*, *bet* and *gam*) and arabinose-inducible *flpe* recombinase activity (for *Frt*-site-mediated BAC modifications). Most templates for homologous recombination contained at their 3' end a *neo* gene flanked by *frt* sites, which enabled selection of BAC recombinants by kanamycin resistance (Kn^R) and subsequent excision of *neo* from the targeted BAC by induction of *flpe* in the EL250 cells. EL250 cells containing Ik-BAC-3 were selected on LB/chloramphenicol (20 μg/ml) plates.

For BAC engineering, a single colony was used to inoculate a 2 to 5 ml overnight culture, which was then diluted 1:50 in LB medium (containing 20 μg/ml chloramphenicol) and grown at 32ºC to an OD_{600} = 0.7-0.8. Ten-milliliter aliquots were induced for Beta, Exo and Gam expression by shifting the cells to 42ºC in a shaking (200 rpm/min) water bath for 15 minutes, followed by chilling in an ice bath slurry for 10-20 minutes. A control (uninduced) aliquot was not subjected to the 42ºC induction but instead maintained on ice for 20 minutes. For BAC transformations, the 42ºC step was omitted. In order to make them electrocompetent for uptake of the targeting template, the cells were centrifuged at 4º C for 8 min at 5500g and washed twice in 10 ml of ice-cold sterile water. For the final wash, cells were resuspended in 1 ml water and spun down in a 4ºC microfuge for 20 sec. Cells were then resuspended in 50-100 μl of ice-cold water for electroporation. The linear DNA recombination template (100-300 ng, diluted in sterile water) was electroporated into 50 μl of ice-cold competent cells in 0.1 cm cuvettes (Bio-Rad) using a Bio-Rad Gene Pulser set to 1.8 kV, 25 mF, and 200 ohms. Immediately after electroporation, 300 μl of SOC medium was added to the cells,

which were then incubated at 32ºC for 1.5 hr with shaking (225 rpm) and spread onto LB plates containing chloramphenicol and kanamycin (20 μg/ml each). The double selection was used to select for colonies in which the *neo*-containing targeting template had successfully integrated into the BAC (Fig.4A). Cm^R Kn^R colonies were subjected to PCR analysis to verify the site-specific integration of the targeting cassette.

Elimination of the *frt*-flanked *neo* gene from the integrated *CD2* cassette was accomplished by arabinose-induced activation of *flpe* in the EL250 cells (Fig.4A). Overnight cultures from single Cm^R Km^R colonies were diluted 50-fold in LB medium containing 20 μg/ml chloramphenicol and grown to an OD_{600} = 0.5, at which point *flpe* expression was induced by incubating the cultures in 0.1% L-arabinose for one hour. Cells were then diluted 10-fold in LB medium, grown for an additional hour and spread onto LB/chloramphenicol plates (20 μg/ml). Loss of *neo* in Cm^R colonies was confirmed by replating the cells onto LB/kanamycin (20 μg/ml) plates and by performing PCR and Southern analysis of the *neo*-deleted region.

Using these techniques, the *CD2* reporter gene (IRES-hCD2-poly (A)) was targeted into Ik-BAC-3 to replace the first translated *Ikzf1* exon (exon 2) (Fig. 4A). The resulting modified BAC (Ik-BAC-hCD2) was purified from a 500-ml culture using a Nucleobond AX 500 Column (Clontech), resuspended in microinjection buffer (10 mM Tris-HCl (pH 7.5) / 0.1 mM EDTA (pH 8.0) / 100 mM NaCl, freshly supplemented with 30 μM spermine and 70 μM spermidine) and injected into fertilized (C57BL/6 x C3H) F_1 oocytes by standard methods. Transgenic founders were identified by PCR using primers to the *CD2* region, followed by Southern

analysis using a 0.5 kb probe from the *IRES* sequence upstream of the *CD2* cDNA in the BAC transgene. For copy number determination, Southern blotting was performed using the UpIkA probe from a region upstream of the 5' homology arm in the *CD2* targeting cassette. UpIkA recognizes both wild-type (11.9 kb) and transgenic (4.1 kb) alleles of EcoRI-digested genomic tail DNA.

To generate the Ik-BAC-lDl transgenic construct, a *loxP*-flanked D (lDl) fragment was targeted into Ik-BAC-hCD2 by homologous recombination in EL250 cells. Like the recombination template used for targeting the *CD2* reporter into the BAC, the lDl cassette included an *frt*-flanked *neo* gene, which marked successful BAC recombinants and was subsequently removed from the BAC by *flpe* activation. It is important to note that lDl insertion was preceded by a recombination step involving the removal of the endogenous *loxP* site in the pBeloBAC11 vector which, if present, would interfere with the intended *Cre*-mediated deletion of region D in Ik-BAC-lDl transgenic mice. The pBeloBAC11 *loxP* site was replaced by an ampicillin resistance gene (*amp*) using a recombination template containing the *amp* gene (from pEGFP-1, Stratagene) along with 54 bp of 5', and 200 bp of 3', pBeloBAC11 vector sequence flanking the *loxP* site.

Ik-BAC-ΔD was generated by *Cre*-mediated recombination in the bacterial line EL350. The deletion of the D region was confirmed by PCR using primers flanking the deleted region as well as by the southern blotting of the BAC plasmid using a 0.3 kb probe corresponding to the 5' homology arm (UpD) of the D-loxP/pBS targeting construct. The copy numbers of the transgene was analyzed by the genomic southern blotting of the tail DNA using the UpD probe. This probe

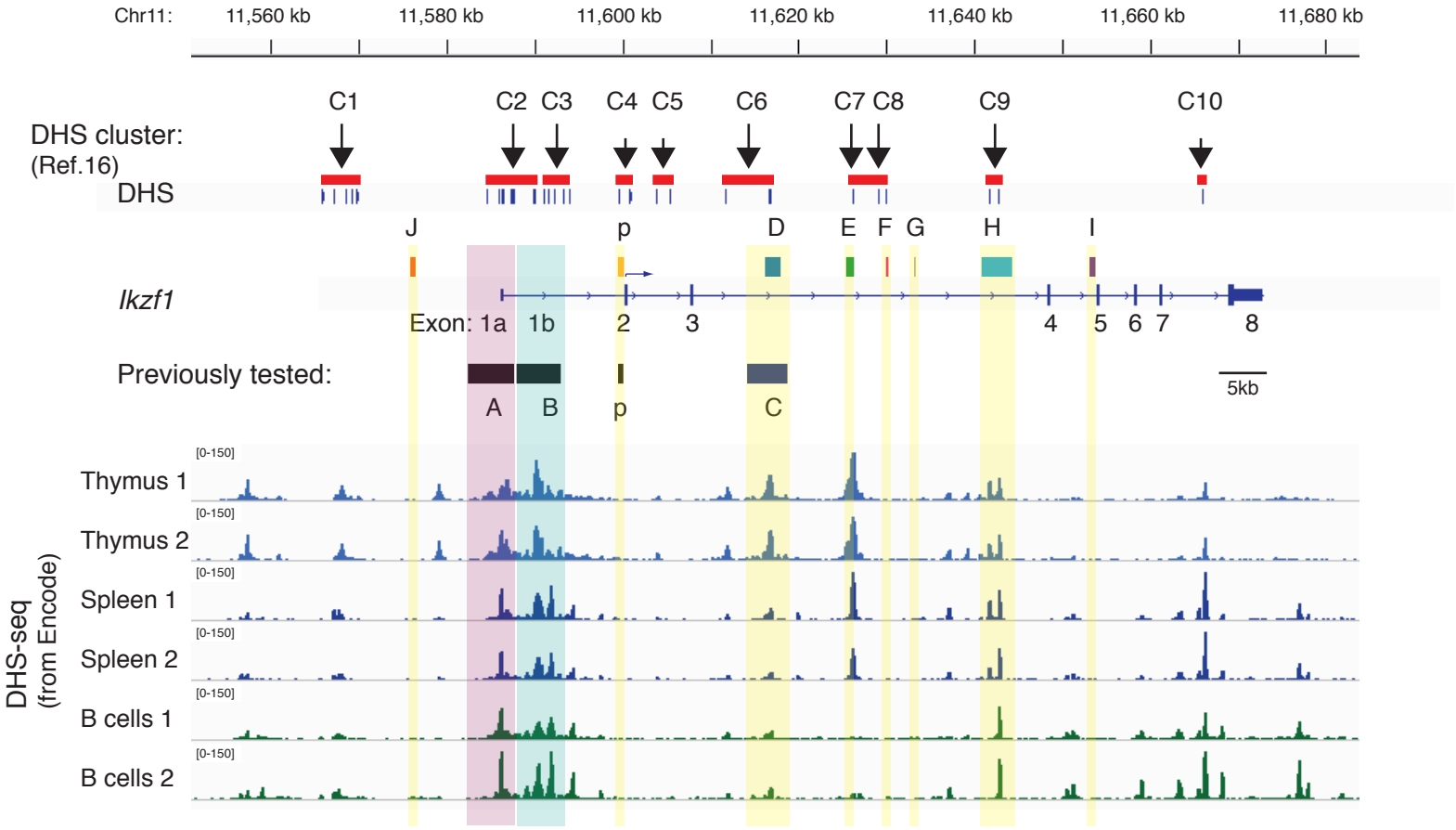
effectively differentiates between wild-type, undeleted transgenic and deleted transgenic alleles in *EcoRI* digested genomic fragments.

Flow Cytometric analysis

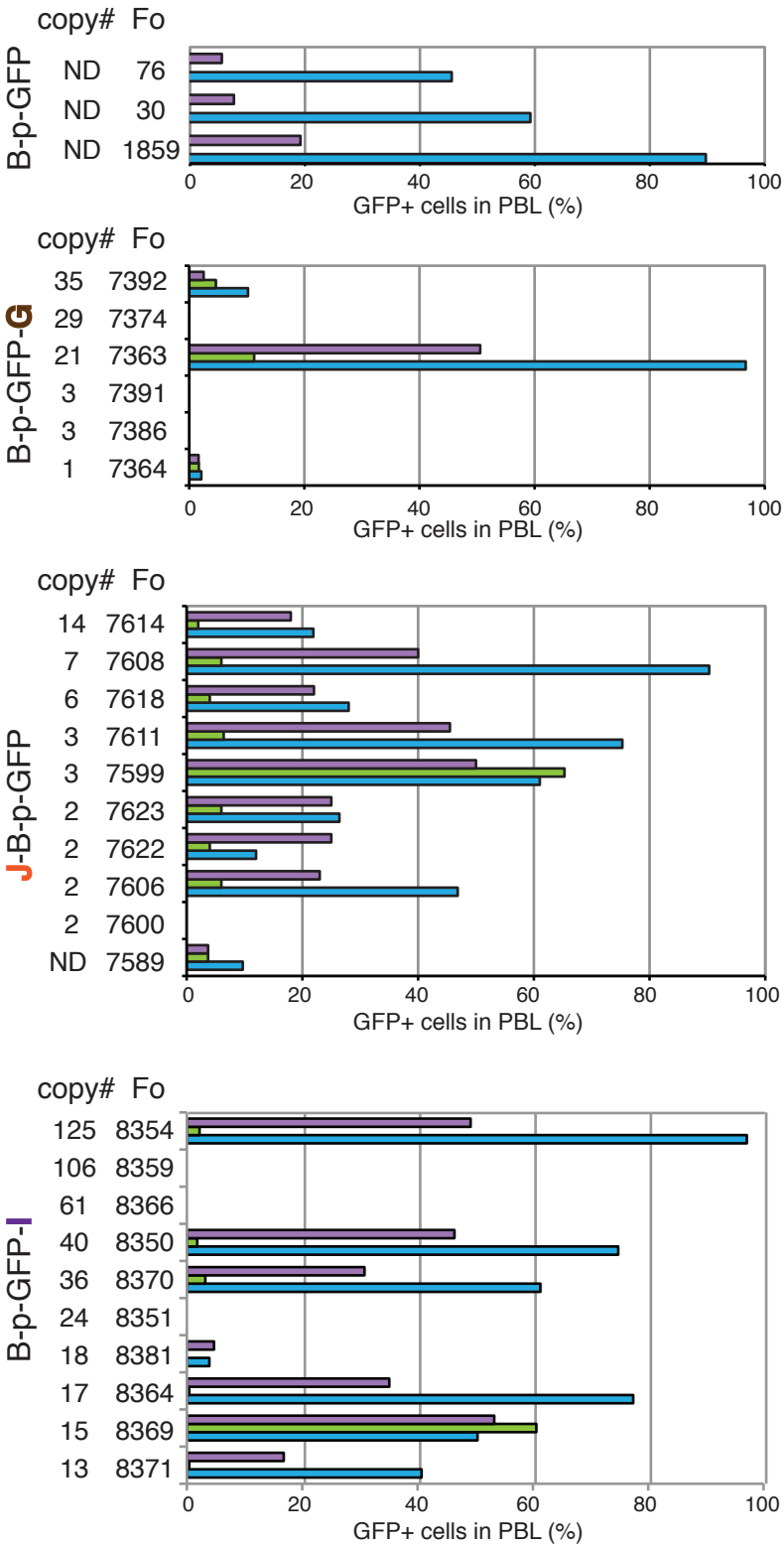
Flow cytometric analysis was performed using fluorescent antibodies from BD, Pharmingen, Caltag or eBiosciences that are specific for the following surface antigens: B220 (RA3-6B2), CD4 (RM4-5), CD8a (53-6.7), TCRβ (H57-597), Mac-1 (M1/70), Ter119 (Ter-119), CD71, Gr-1 (RB6-8C5), c-Kit (2B8, ACK2), Sca-1, Flt3 (A2F10.1) and hCD2 (RPA-2.10). Antibody stained cells were analyzed on FACScan, FACS Calibur, FACSCanto or Fortessa flow cytometer (BD). Data analysis was performed using FlowJo software (Tree Star, Inc.). Cell sorting was preformed on a Moflo (Dako Cytomation).

Identification of transcription factor binding sites

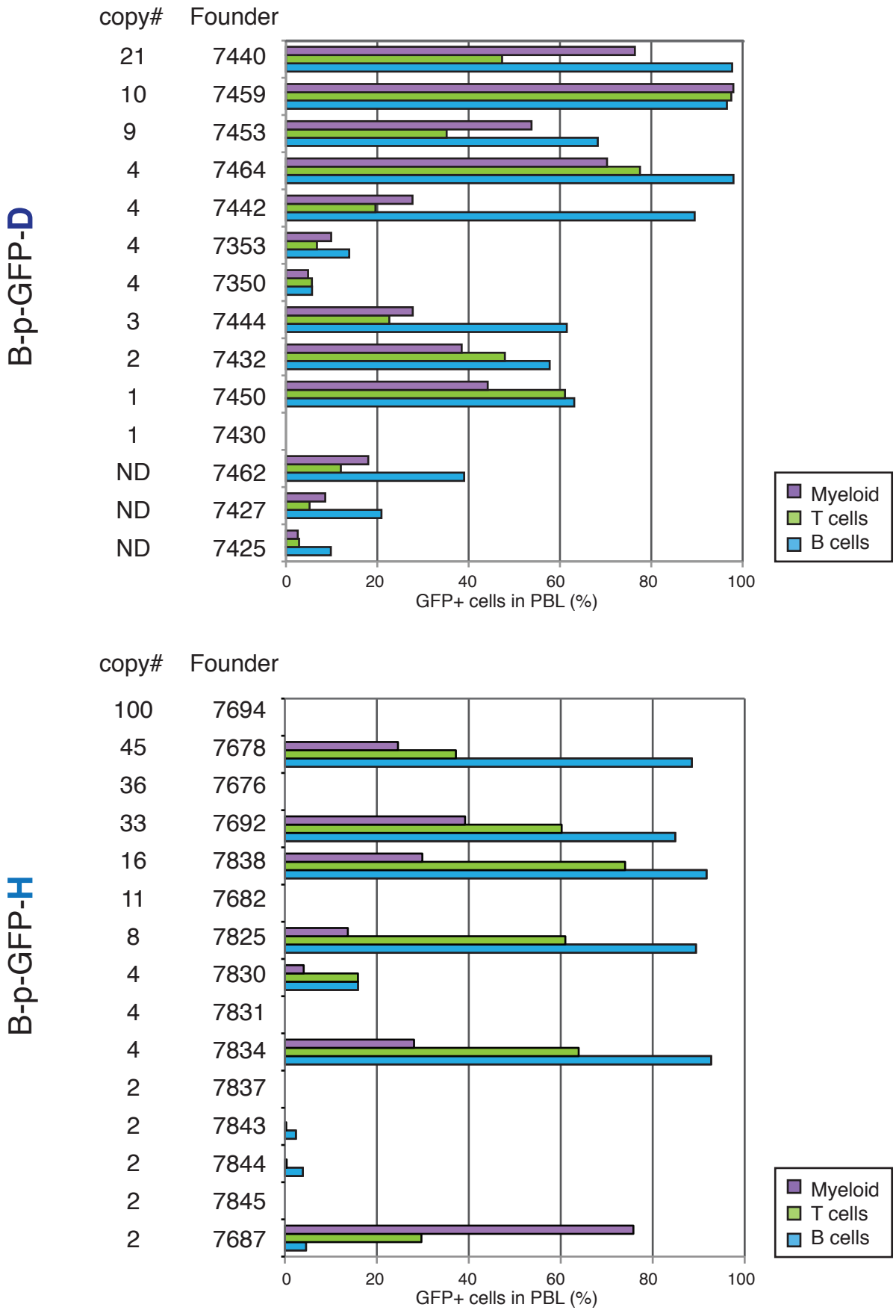
The conserved genomic regions with potential enhancer activity were visualized and analyzed with UCSC Genome Browser as well as MegaBlast. TRANSFAC was employed to identify the TFs with significant binding sites on the extracted conserved genomic sequences.



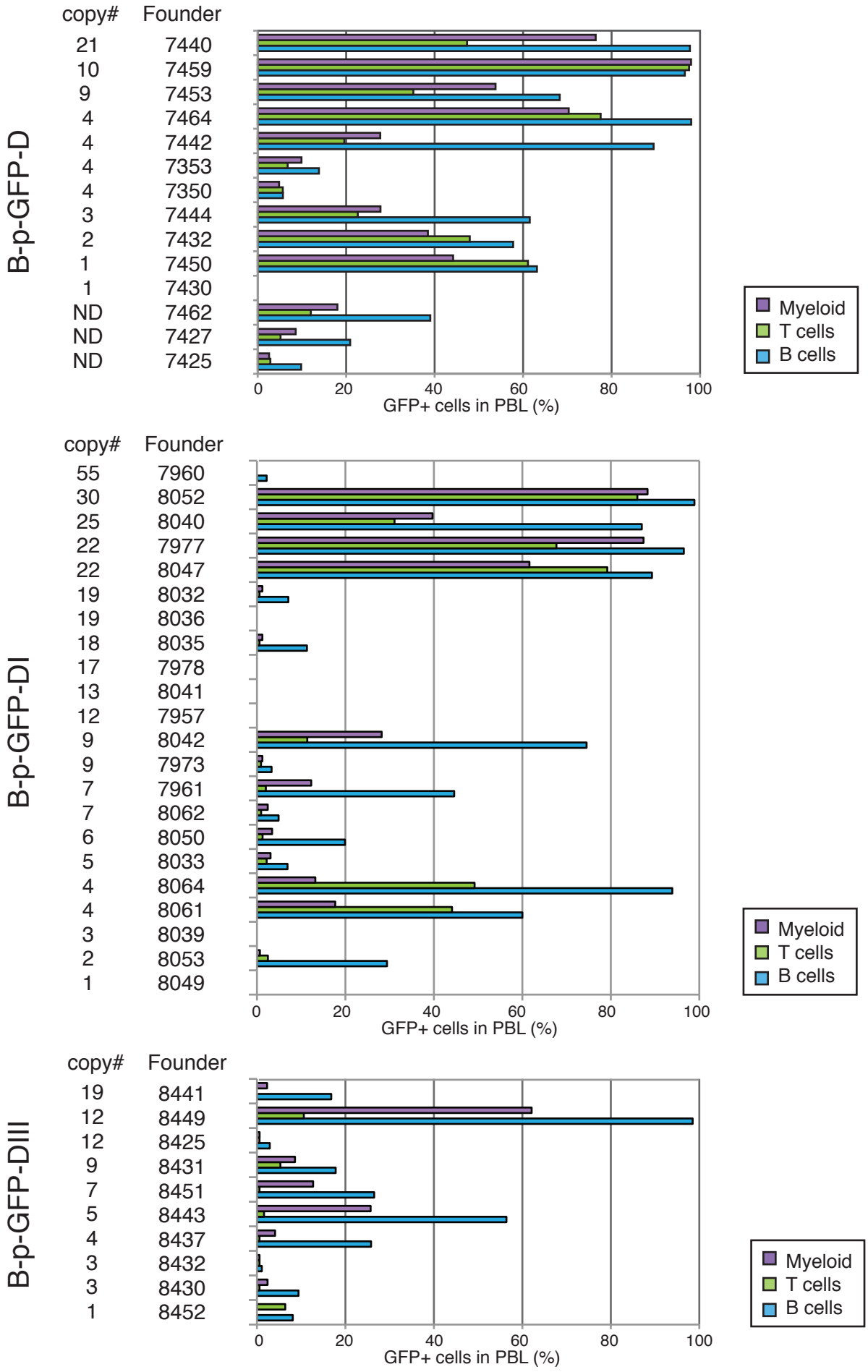
Supplemental Figure 1



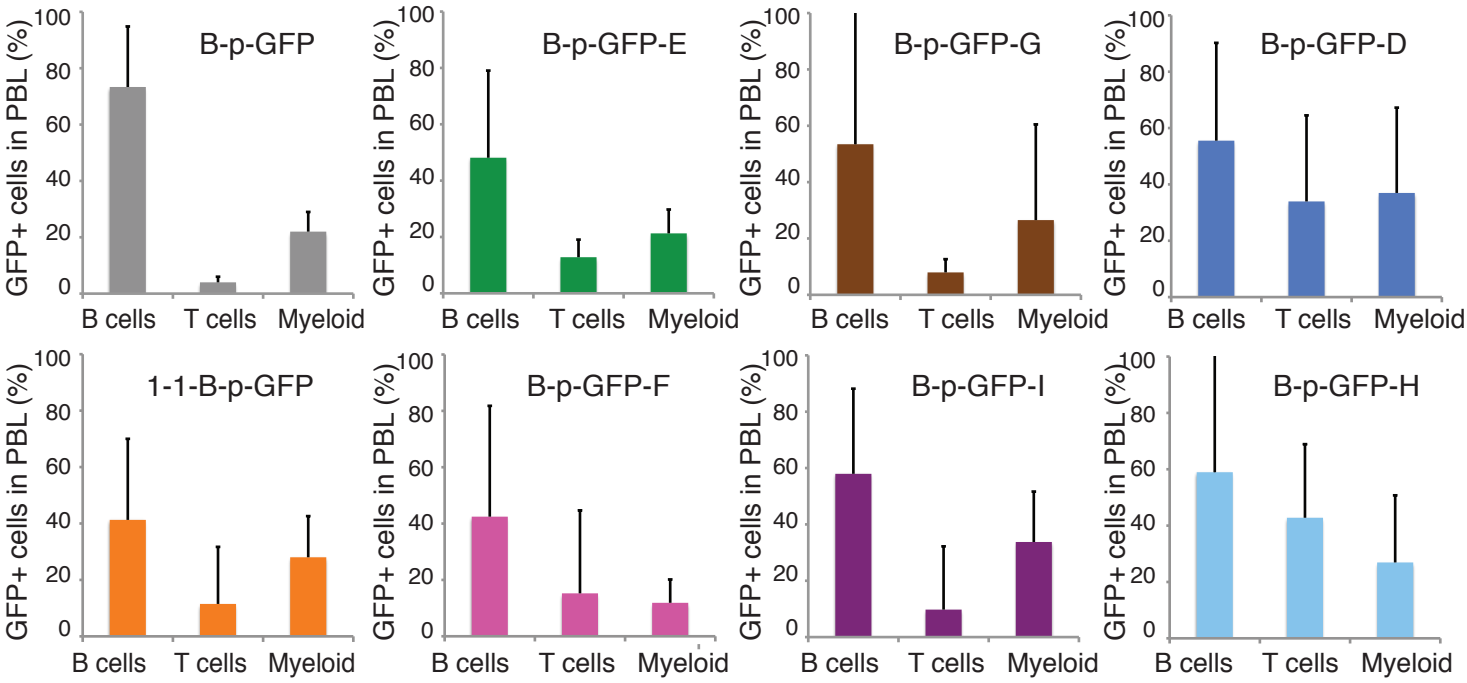
Supplemental FIgure 2



Supplemental Figure 3



Spplemental Figure 4



Supplemental Figure 5

D1

```
                    Irf4  Tcf3, Tcf7
M: TTTGGTCCCAGTGTTATTTTCTGCTTTGTTCTCTGTGTGTTGAAAGA
   |||||||||||| |||||||||| |||||| | ||||||||||||
H: TTTGGTCCCAGTTTTATTTTCTGCCTTGTTCCATCTGTGTTGAAAGA
```

D2

```
                  Tcf7
M: CAACTGACATTCCGTTAACTTTCCTTCTGCATGGGGTTTCAGAAAGAAATGG
   |||||||||||| |||||||||||||||| ||||||||||||||||||||||
H: CAACTGACATTTCGTTAACTTTCCTTCTACATGGGGTTTCAGAAAGAAATGG
```

D3

```
     Runx1/3          E-box          Gata1                Ets/Ikzf1
M: GCTGTGGTTGCCAGTTGCCACCTGTCCTCCTATCTGGTTAGGACATTTTTCTGATTTGGG
   |||||||||||||||||||||||||||| |||||||||||||| || ||||||||||||||
H: GCTGTGGTTGCCAGTTGCCACCTGTCCCCCTATCTGGTTAGGGCA-TTTTCTGATTTGGG
        Satb1
M: AAGTATAAATATAACCCAG
   |||| ||||||||||||||
H: AAGTGTAAATATAACCCAG
```

D4

```
                   Satb1
          Ets/Ikzf1 Cebpa/Cebpb     Irf3
M: AACAGCTTCCCTTCCTTTTATGCAAAAGTAG-AAGGTTTAA
   |||||||||||| |||||||||||||||||| |||| ||||
H: AACAGCTTCCCCTCCTTTTATGCAAAAGTAGAAAGGCTTAA
```

D5

```
                 Ets/Ikzf1
M: GCTGCCAGGAGAGTACTTCCCCTACAGGGGATGTGGCTGCAGGCTGTCGATTTCCACTGAAGTCACC
   |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||| |||||
H: GCTGCCAGGAGAGTACTTCCCCCACAGGGGATGTGGCTGCAGGCTGTCGATTTCCACTGAAATCACC
```

D6

```
           Irf1/Irf7
M: GTTACCCTAGAAGTGAAACCCTGCTGAAATGAGAAAAAGGCTTTTATTACAGCCCTGCAG
   |||||||||||||||||| |||||||||||||||| ||| |||||| | || ||||||
H: GTTACCCTAGAAGTGAAATCCTGCTGAAATGAGAAAAGTGCTGTTATTATATTTCTACAG
                       Stat5a
M: GCAGCCATATGATTATTGTTATTATTATTCTTAGAACCTGGCA
   ||||| ||||||| ||| |||||||| |||||| ||||||||||
H: GCAGCTATATGATCATTTTTATTATGATTCTTGGAACCTGGCA
```

D7

```
                        Satb1
                   Hoxa9   Mef2c
M: CTCCACTTGAACTTGGTGGCAATAAAAATAAGAATC
   ||||| |||||||||||||||||||| ||||||||
H: CTCCAGTTGAACTTGGTGGCAATAAAATTAAGAATC
```

Supplemental Figure 6

H1

```
                                                                  Ikzf2
M: CTTTCTTTTTGCCTTCTCTATTGCAGTTGCATATGGGGCTGATGGCTTTAGGGATTTTCA
   |||||||||||||||| |||||||||| |||||||||||||||| |||||||||||| ||
H: CTTTCTTTTTGCCTTCTTTATTGCAGTTACATATGGGGCTGATGACTTTAGGGATTTCCA
        Ets/Ikzf1                                                 Tal1:Tcf3
M: TGCAATAATTCCCAAATCTTTCTCTCGTAAGTATATGCTTTGCTTCTAGAAAACAGATGTTCA
   |||||||||||||||||||||||||||||||||||||||| |||||||| ||||||| | |
H: TGCAATAATTCCCAAATCTTTCTCTCGTAAGTATATGCCTTGCTTCTGGAAAACAAAAGCATG
```

H2

```
                                                  Runx1
M: AACAACAACCAAACTGTTCTGGGCCAATATCGCCACCTTGTGGTCATG
   |||| |||||||||||||||||||||||| |||||||||||||||
H: AACAGCAACCAAACTGTTCTGGGCCAATATCACCACCTTGTGGTCATG
```

H3

```
                                Runx1
M: ATGTAGCATTTCACTGAAGGAATACCACAGCTAAACAGTTC
   ||||| |||||||||||||| ||||||||||||||| |||||||
H: ATGTACCATTTCACTGAAGCAATACCACAGCTAAGCAGTTC
```

Supplemental Figure 7

Supplemental Figure legends

Supplemental Figure 1. DHS alignment on *Ikzf1* locus. Mapping of DHS obtained from our previous study {Kaufmann, 2003 #351} is shown. Each DHS is indicated by a short vertical bar and DHS clusters are indicated by red bars above the locus. DHS peaks on thymus, spleen and B cells were mapped below the locus for comparison. DHS-seq data used for this analysis were obtained from the ENCODE project {Consortium, 2011 #3826} (Supplemental Table 8).

Supplemental Figure 2. Activities of the *Ikzf1* enhancers in peripheral blood leukocytes (PBL). The percentage of GFP^+ myeloid cells ($Mac\text{-}1^+$, purple bars), T cells ($TCR\beta^+$, green bars) and B cells ($B220^+$, blue bars) in the PBL was determined for each founder animal by flow cytometry. Copy numbers are noted beside each founder line. N.D., not determined. (**A**), B-p-GFP; (**B**), B-p-GFP-G; (**C**), J-B-p-GFP; (**D**), B-p-GFP-I; (**E**), B-p-GFP-E; (**F**). For the B-p-GFP reporter lines, only data from GFP-expressing founders are shown (i.e. the 24 non-expressing founders were omitted).

Supplemental Figure 3. Activities of the *Ikzf1* enhancer D and H in PBL . GFP expression in peripheral blood subsets of the B-p-GFP-D (**A**) and B-p-GFP-H (**B**) transgenic reporter lines are shown as indicated. The percentage of GFP^+ myeloid cells ($Mac\text{-}1^+$, purple bars), T cells ($TCR\beta^+$, green bars) and B cells ($B220^+$, blue bars) in the PBL was determined for each founder line by flow cytometry. Copy numbers are noted beside each founder line. N.D., not determined.

Supplemental Figure 4. Cell type specificity of enhancer D sub-regions. The percentage of GFP$^+$ myeloid cells (Mac-1$^+$, purple bars), T cells (TCRβ$^+$, green bars) and B cells (B220$^+$, blue bars) was determined for each founder animal by flow cytometry. Copy numbers are noted beside each founder line. N.D., not determined.

Supplemental Figure 5. Cell type specificity of the *Ikzf1* enhancers

Lineage-specific GFP expression in PBL subsets from from founders generated by *Ikzf1*enhancer-based reporters. The average percentage of GFP$^+$ cells within peripheral blood B cells (B), T cells (T) and Myeloid cells (My) is shown. Error bars (Standard deviation) indicate variegation of GFP expression among the different founder lines made by each enhancer-based reporter.

Supplemental Figure 6. Mouse–human sequence alignments and transcription factor binding motifs on D sub-regions.

Sequence alignments of seven mouse–human homologous sub-regions (D1 to D7) in D are shown (M; mouse, H; human). Motifs for the key transcription factors that were identified by in silico analyses (Supplemental Table 7) and ChIP-seq (Fig. 7) were highlighted on each sub-region. A factor and the corresponding motif are marked with the same color such as blue, bolded and underlined or boxed. Some factors on Supplemental Table 7 are not highlighted because they were found in DI or DIII but not in D1 to D7.

Supplemental Figure 7. Mouse–human sequence alignments and transcription factor binding motifs on H sub-regions.

Sequence alignments of three mouse–human homologous sub-regions (H1 to H3) in H are shown (M; mouse, H; human). Motifs for the key transcription factors that were identified by in silico analyses (Supplemental Table 7) and ChIP-seq (Fig. 7) were highlighted on each sub-region. A factor and the corresponding motif are marked with the same color such as blue, bolded and underlined or boxed.

Supplemental Table 1. Predicted functions of cis regions based upon GFP reporter expression data. Tg lines with GFP expression more than >1% in PB were considered as GFP+ Founders (F_0).

Transgenic construct	Proportion of GFP$^+$ founders (GFP $^+F_0$/all F_0)	Chi squre test (X^2)**	P-value**	Average GFP expression in PB in all F_0 (%)	P-value**	Average GFP expression in PB in GFP$^+$ F_0(%)	P-value**	MFI of all F_0	P-value***	MFI in PB of GFP$^+$ F_0	P-value***	Expression profile (average % of GFP expression among GFP$^+$ F_0)			Predicted function(s) of *cis* region
												B	T	M	
B-p-GFP*	3/27 (11%)	N/A	N/A	4.5	N/A	40.2	N/A	ND	ND	ND	ND	++++ (73.3)	+/– (4.0)	++ (22.0)	**promoter**
J-B-p-GFP	9/10 (90%)	20.7	5.30E-06	19.0	1.14E-02	21.1	2.08E-01	162.0	1.83E-02	244.5	2.59E-04	+++ (41.3)	+ (11.5)	++ (28.0)	**chromatin opening, moderate enhancer**
B-p-GFP-D	13/14 (93%)	25.9	3.61E-07	39.4	1.23E-03	42.4	9.16E-01	1023.1	N/A	1567.3	N/A	++++ (55.5)	+++ (34.0)	+++ (37.0)	**chromatin opening, strong enhancer, T cell specificity**
B-p-GFP-DI	17/22 (77%)	22.0	2.77E-06	21.1	1.97E-02	27.2	4.90E-01	345.5	5.20E-01	1282.1	9.25E-01	+++ (42.9)	++ (22.3)	++ (21.3)	**chromatin opening, moderate enhancer, T cell specificity**
B-p-GFP-DIII	10/10 (100%)	25.3	4.90E-07	15.1	5.81E-02	15.1	4.55E-02	511.7	1.14E-01	511.7	6.52E-02	++ (26.3)	+/– (1.9)	+ (12.5)	**chromatin opening, weak enhancer**
B-p-GFP-E	15/17 (88%)	25.7	4.10E-07	18.9	6.29E-04	21.4	6.82E-02	303.3	1.39E-01	504.8	2.47E-01	+++ (45.1)	+ (12.0)	++ (20.0)	**chromatin opening, moderate enhancer**
B-p-GFP-F	9/12 (75%)	15.9	6.61E-05	15.6	1.61E-01	20.8	2.74E-01	269.7	1.62E-01	1097.8	5.01E-01	+++ (42.5)	+ (15.2)	+ (11.9)	**chromatin opening, moderate enhancer**
B-p-GFP-G	2/6 (33%)	1.9	1.70E-01	10.3	4.27E-01	30.7	7.34E-01	N/A	N/A	N/A	N/A	++++ (53.4)	+/– (8.0)	++ (26.5)	**no significant function in hemopoietic system**
B-p-GFP-H	8/15 (53%)	8.9	2.86E-03	25.6	5.64E-02	47.8	7.04E-01	63.8	2.69E-03	550.4	3.71E-02	++++ (59.0)	+++ (42.8)	++ (27.0)	**weak chrom. opening, moderate enhancer, T cell specificity**
B-p-GFP-I	7/10 (70%)	12.8	3.41E-04	22.0	1.51E-04	31.1	4.79E-01	136.7	5.00E-01	620.7	6.34E-02	++++ (57.9)	+/– (9.8)	+++ (33.8)	**chromatin opening, moderate enhancer**

*Reported in Kaufmann et al (2006)
**Compared to B-p-GFP line
***Compared to B-p-GFP-D line

0 to10 %	+/–
10 to 20 %	+
20 to 30 %	++
30 to 50 %	+++
50 to 100 %	++++

Supplemental Table 2. Predicted functions of D-subregions based upon GFP reporter expression data. Tg lines with GFP expression more than >1% in PB were considered as GFP+ Founders (F_0).

Transgenic construct	Proportion of GFP$^+$ founders (GFP $^+F_0$/all F_0)	Chi squre test (X^2)***	P-value***	Average GFP expression in PB in all F_0 (%)	P-value***	Average GFP expression in PB in GFP$^+$ F_0 (%)	P-value***	MFI of all F_0	P-value***	MFI in PB of GFP$^+$ F_0	P-value***	Expression profile (average % of GFP expression among GFP$^+$ F_0)			Predicted function(s) of *cis* region
												B	T	M	
B-p-GFP-D	13/14 (93%)	25.89	N/A	39.4	N/A	42.4	N/A	1023.1	N/A	1567.3	N/A	++++ (55.5)	+++ (34.0)	+++ (37.0)	**chromatin opening, strong enhancer, T cell specificity**
B-p-GFP-DI	17/22 (77%)	1.50	2.21E-01	21.1	9.09E-02	27.2	2.01E-01	345.5	5.20E-01	1282.1	9.25E-01	+++ (42.9)	++ (22.3)	++ (21.3)	**chromatin opening, moderate enhancer, T cell specificity**
B-p-GFP-DIII	10/10 (100%)	0.75	3.88E-01	15.1	2.81E-02	15.1	1.68E-02	511.7	1.14E-01	511.7	6.52E-02	++ (26.3)	+/– (1.9)	+ (12.5)	**chromatin opening, weak enhancer**

***Compared to B-p-GFP-D line

0 to10 %	+/–
10 to 20 %	+
20 to 30 %	++
30 to 50 %	+++
50 to 100 %	++++

Supplemental Table 3. The sequences of mouse-human conserved sub-regions in D.

Sub-regions	Sequences (5' to 3')	Length (bp)	Homology (%)
D1	TTTGGTCCCAGTGTTATTTTCTGCTTTGTTCTCTGTGTGTTGAAAG	46	89
D2	CAACTGACATTCCGTTAACTTTCCTTCTGCATGGGGTTTCAGAAAGAAATGG	52	96
D3	GCTGTGGTTGCCAGTTGCCACCTGTCCTCCTATCTGGTTAGGACATTTTTCTGATTTGGGAAGTATAAATATAACCCAG	78	94
D4	AACAGCTTCCCTTCCTTTTATGCAAAAGTAGAAAGGTTTAA	40	93
D5	GCTGCCAGGAGAGTACTTCCCCTACAGGGGATGTGGCTGCAGGCTGTCGATTTCCACTGAAGTCACC	66	96
D6	GTTACCCTAGAAGTGAAACCCTGCTGAAATGAGAAAAAGGCTTTTATTACAGCCCTGCAGGCAGCCATATGATTATTGTTATTATTATTCTTAGAACCTGGCA	103	86
D7	CTCCACTTGAACTTGGTGGCAATAAAAATAAGAATC	37	94

Supplemental Table 4. The sequences of mouse-human conserved sub-regions in H.

Sub-regions	Sequences (5' to 3')	Length (bp)	% Homology
H1	CTTTCTTTTTGCCTTCTCTATTGCAGTTGCATATGGGGCTGATGGCTTTAGGGATTTTCATGCAATAATTCCCAAATCTTTCTCTCGTAAGTATATGCTTTGCTTCTAGAAAACA	115	95
H2	AACAACAACCAAACTGTTCTGGGCCAATATCGCCACCTTGTGGTCATG	48	96
H3	ATGTAGCATTTCACTGAAGGAATACCACAGCTAAACAGTTC	41	93

Supplemental Table 5. Oligo sequences used to generate transgenic reporter constructs

Transgenic construct	Forward	Reverse
J-B-p-GFP	agc tcg taa gaa ctt gat c	agc ctc cca cag tg
B-p-GFP-D	gga gtc caa gat taa aag gcc	tca tta tta cac att ggg gtg cc
B-p-GFP-E	ggg gcc ctg cag ctg	tct gtg cac aga tag ctt aga
B-p-GFP-F	ggc cta tat ttg tgt gtc tgt	tct ggc ata gca gct acg
B-p-GFP-G	gcc aac acc tgg acc tg	tct aag tcc tag ggt gct c
B-p-GFP-H	gca ctg tct tct gca ctc	tct gac aga ttt cta ttc aac ca
B-p-GFP-I	ttc cct ctg gag gac	agc atg gta agc act gag a.

Supplemental Table 6. Genomic coodinates of DHS determined in Kaufman et al (2003).

DHS cluster	Fragment used to determine DHS	5' (mm9)	3' (mm9)	DHS a	DHS b	DHS c	DHS d	DHS e	DHS f	DHS g
C1	BamHI-BamHI	11563780	11573791	11570191	11569591	11568991	11567491	11566291	NA	NA
C2	BamHI-EcoRI	11582772	11588395	11587995	11587795	11587595	11586795	11586595	11586195	11584895
C3	EcoRI-EcoRI	11588395	11593297	11590195	11590395	11591295	11591795	11592495	NA	NA
C4	EcoRI-EcoRI	11593297	11597381	11594181	11593581	NA	NA	NA	NA	NA
C5	BamHI-BamHI	11597639	11609618	11599839	11601139	11604139	11605639	NA	NA	NA
C6	BamHI-BamHI	11609618	11619977	11617077	11616877	11611977	NA	NA	NA	NA
C7/8	EcoRI-EcoRI	11618861	11632719	11630119	11629219	11626319	NA	NA	NA	NA
C9	XbaI-XbaI	11641030	11644300	11642900	11641800	NA	NA	NA	NA	NA
C10	BamHI-BamHI	11655232	11669484	11665984	NA	NA	NA	NA	NA	NA

Supplemental Table 7. Representative of factors whose binding motifs were found in conserved regions. Unique sites at enhancer D (D minus H) identify a network of factors expressed in the HSC and in early lympho-myeloid progenitors whereas shared binding sites (Common D and H) identify factors expressed highly at intermediate stages of lymphocyte differentiation. This latter group of binding sites and their cognate factors map at the DI sub-region (DI minus DIII) of enhancer D.

D minus H	Common D and H	DI minus DIII
Tcf7	*Runx1*	*Tcf3*
Runx3	*Cbfb*	*Tcf7*
E-box	*Ets family*	*Runx3*
Gata1	*Ikzf1*	*Runx1*
Cebpa	*Tcf3*	*Cbfb*
Cebpb	*Tal1*	*E-box*
Mycn		*Gata1*
Pax5		*Cebpa*
Irf1,3,4,7		*Cebpb*
Hoxa9		*Mycn*
Stat5a		*Irf3*
Mef2c		
Satb1		

Supplemental Table 8. List of ENCODE data used in the study.

ChIP-Seq	**GEO_Accession**	**DCC_Accession**
P300	GSM912909	wgEncodeEM00195
H3K27Ac	GSM1000103	wgEncodeEM002474
Ets-1	GSM1003774	wgEncodeEM002778
c-Myc		wgEncodeEM001951
Gata-1	GSM923575	wgEncodeEM002349
Gata-2		wgEncodeEM002356
Tal1	GSM923582	wgEncodeEM002359
DHS-seq	**GEO_Accession**	**DCC_Accession**
Thymus	GSM1014185	wgEncodeEM003395
Spleen	GSM1014182	wgEncodeEM003394
B cells (CD19+)	GSM1014190	wgEncodeEM001727