Supplemental Materials

Three-step transcriptional priming that drives the commitment of B cell progenitors toward B cells.

Tomohiro Miyai, Junichiro Takano, Takaho A. Endo, Eiryo Kawakami, Yasutoshi Agata, Yasutaka Motomura, Masato Kubo, Yukie Kashima, Yutaka Suzuki, Hiroshi Kawamoto, and Tomokatsu Ikawa*

*Correspondence author.
E-mail: tomokatsu.ikawa@riken.jp

Supplemental Materials & Methods

Antibodies, growth factors and flow cytometry. The following antibodies were purchased from BD Pharmingen or BD Horizon: c-kit (2B8), Sca-1 (D7), Flk2 (A2F10.1), CD34 (RAM34), IL-7Ra (SB/199), CD24 (M1/69), BP-1 (6C3), CD43 (S7), erythroid lineage cells (TER119), Mac-1 (M1/70), Gr-1 (RB6-8C5), CD11c (HL3), B220 (RA3-6B2), CD8 (53-6.7), CD4 (H129.19), NK1.1 (PK136), CD3ε (145-2C11), and CD19 (1D3). IgM (II/41), IgD (11-26C) and human CD25 (M-A251). TER119, Mac-1, Gr-1, B220, CD19, NK1.1, CD3ε, CD4, and CD8 were used as Lin markers. For single cell RNA-seq analysis, CD11c, CD25, IgM, IgD antibodies were also added to Lin markers whereas CD4 was excluded. Recombinant murine IL-3, IL-7, stem cell factor (SCF), Flt-3 ligand (Flt-3L), thrombopoietin (Tpo) were purchased from R&D. 6-10 wk old mice were used for all experiments. Mice were sacrificed and single cell suspensions of BM, spleen and thymus were prepared. Cell number was counted, then ~1×10^7 cells were stained with combination of antibodies in FACS buffer (1× PBS with 2% FCS and 0.05% NaN_3). Cellularity is assessed with FACSAria II or FACSCanto II (Beckton Dickinson).

Isolation of hematopoietic progenitors and transduction of retrovirus. Single cell suspensions of FL cells from 14 dpc or BM cells of B6.CD45.1 mice were prepared as described previously(Ikawa et al. 2004). Cells were then incubated with purified monoclonal antibodies specific for lineage markers (Lin: TER119, Mac-1, Gr-1, B220

and Thy1.2) for 20 min on ice. Lin^+ cells were depleted with Dynabeads Sheep anti-Rat IgG (Invitrogen) according to the manufacturer's protocol. LSK cells or HSCs from FL and adult BM were isolated by FACS Aria II. Retroviral vector pMCs-IRES-GFP containing human ID3 cDNA fused with modified human estrogen receptor ER^{T2} (pMCs-hId3-ER^{T2}-IG) were generated as previously described. shRNA against various targets were prepared by Open Biosystems LMP system (MSCV-LTRmiR30-PIG, Thermo Scientific). The vector of the coding sequence of GFP was replaced by human CD25 gene (a gift from Dr. C. Murre). Vector construction of specific targets was performed according to the manufacturer's protocols, and target sequences are shown in Table S1. Retrovirus was generated by transfection of pMCs-hId3-ER^{T2}-IG or shRNA construct plasmid together with pantropic envelope protein VSV-G plasmid into Plat E packaging cell line using FuGENE 6 Transfection Reagent (Promega). Viral supernatant was collected at day 2 and 3 after transfection.

PCR analysis for genomic rearrangement of IgH locus. Genomic DNA is extracted by DNeasy Blood and Tissue Kit (Qiagen). PCR is performed with KAPATaq Extra HotStart ReadyMix with dye (KAPA Biosystems). Primer sequences are shown in Supplementary Table 2.

ChIP and ChIP-seq. Detailed ChIP protocol was described elsewhere(Ikawa et al. 2006). Cells were fixed with 1% formaldehyde solution for 10 (for methylated histone) or 20 (for TFs) minutes at 25°C. Soluble chromatin was immunoprecipitated with

normal rabbit IgG (as negative control, Santa Cruz Biotechnology, #sc-2027), anti-H3K4me3 (Millipore, #07-473), anti-H3K27me3 (Millipore, #07-449), anti-H3K9me3 (Medical & Biological Laboratories, #MABI0308), anti-E2A (Santa Cruz Biotechnology, #sc-349X). For quantification, precipitated and input DNA were purified with MinElute PCR Purification Kit (Qiagen) and performed quantitative PCR with SYBR Premix Ex Taq (TaKaRa). Primer sequences are shown in Table S2. For sequencing, purified DNA were fragmented with DNA shearing system M220 (Covaris) then libraries were prepared with NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs). Procedures for quality check of libraries and sequencing were corresponded to RNA-seq.

Data analysis. Sequences obtained in RNA-seq experiments were aligned on mouse genome (NCBI build 37, mm9) using tophat2 (version 2.1.0)(Kim et al. 2013) and assigned to mouse genes using cufflinks (version 2.1.1)(Trapnell et al. 2010) with gene annotation provided by iGenomes. Fold changes and p-values of gene expression in B cell progenitors of Eμ-Nfil3 and WT mice was performed using cuffdiff (version 2.1.1)(Trapnell et al. 2013). Expression similarity among experimental replicates was estimated using Pearson correlation of RPKM values of all genes. In the PCA plot, we used 4290 representative genes that had different expression (difference of maximum and minimum FPKM values is more than twice) among time points with less than 0.01 q-value using ANOVA. To extract the TF (E2A, EBF1 and PAX5) target genes, previously deposited ChIP-seq records (Accession numbers of Gene Expression

Omnibus are as follows; GSM546523 for E2A (Lin et al. 2010), SRX018714 for EBF1 (Treiber et al. 2010) and SRX148793 for PAX5 (Revilla-i-Domingo et al. 2012)) were used. Alignment on mouse genome (mm9) was performed with bowtie2-2.2.2 (Langmead and Salzberg 2012). Peaks were called with MACS2 (Zhang et al. 2008) following peak annotation with ChIPseeker package (Yu et al. 2015). Based on the obtained gene sets of respective TF targets, temporal expression signatures of TF target or non-target genes were calculated among highly variable genes during time-course. ChIP-Seq reads were aligned on mouse genome (mm9) using bowtie 2.1.0 and compared with gene locations Mapping to reference genome is performed with Integrative Genomics Viewer (https://www.broadinstitute.org/igv). To see epigenetic changes of differently expressed genes before and after differentiation, we evaluated expression changes between 0 hr and 144 hr samples using DESeq2 package(Love et al. 2014) and collected significant genes (q-value < 0.01) expressed more than 2 times (410 genes) or less than 0.5 times (304 genes). RPKM around TSS (-5kb to +2kb, 100bp window size) was subtracted between two time points and represented.

PCA, heat maps, correlation plots, k-means clustering, GO analysis, ChIP-seq subtraction diagrams were analyzed using in-house computer programs developed by Takaho A Endo, RIKEN IMS.

Statistics. All experiments were performed at least 2 times. The two-tailed Student's t-test was used to analyze differences between two groups.

References for Supplemental Materials & Methods

Ikawa T, Kawamoto H, Goldrath AW, Murre C. 2006. E proteins and Notch signaling cooperate to promote T cell lineage specification and commitment. *J Exp Med* **203**: 1329–1342.

Ikawa T, Kawamoto H, Wright LYT, Murre C. 2004. Long-Term Cultured E2A-Deficient Hematopoietic Progenitor Cells Are Pluripotent. *Immunity* **20**: 349–360.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Lin YC, Jhunjhunwala S, Benner C, Heinz S, Welinder E, Mansson R, Sigvardsson M, Hagman J, Espinoza CA, Dutkowski J, et al. 2010. A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat Immunol* **11**: 635–643.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Revilla-i-Domingo R, Bilic I, Vilagos B, Tagoh H, Ebert A, Tamir IM, Smeenk L, Trupke J, Sommer A, Jaritz M, et al. 2012. The B-cell identity factor Pax5 regulates distinct transcriptional programmes in early and late B lymphopoiesis.

EMBO J **31**: 3130–3146.

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31**: 46–53.

Trapnell C, Williams B a, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511–515.

Treiber T, Mandel EM, Pott S, Györy I, Firner S, Liu ET, Grosschedl R. 2010. Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin. *Immunity* **32**: 714–725.

Yu G, Wang LG, He QY. 2015. ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**: 2382–2383.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.

Supplemental Figures

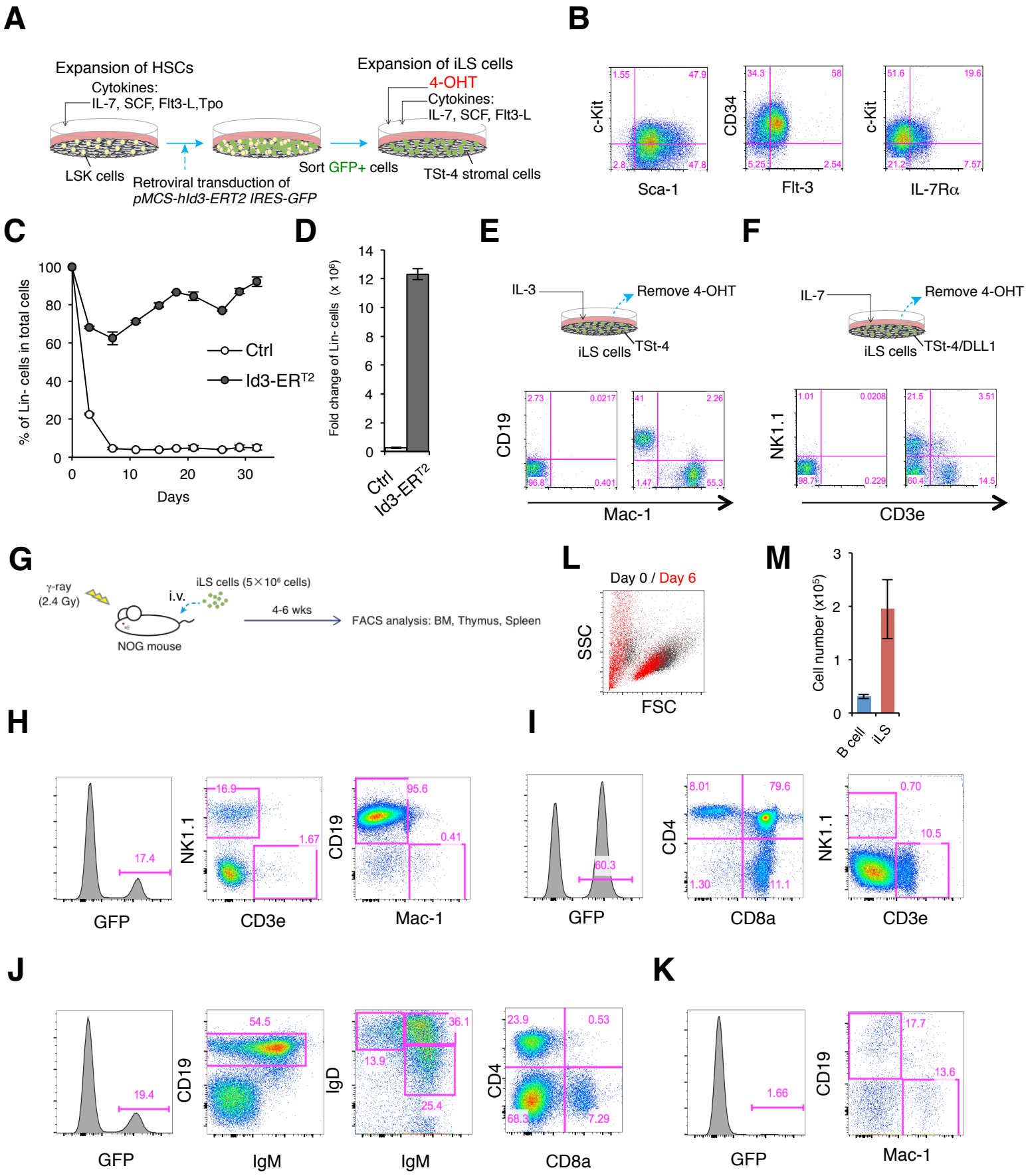

Figure S1. Establishment and characterization of iLS cells. (A) Scheme for establishment of iLS cells. (B) Flow cytometric profiles of Id3-ER^{T2} transduced cells. (C) Proportion of lineage negative (Lin$^-$: CD19$^-$Mac-1$^-$NK1.1$^-$) populations of LSK(control) or Id3-ER^{T2} transduced cells. Values represents mean ± SD in a representative of two independent experiments.

(Continue to next page)

(D) Number of Lin^- cells after 1 month culture in LSK or Id3-ER^{T2} transduced cells. Values represents mean ± SD in a representative of two independent experiments. (E) Myeloid lineage potential of iLS cells *in vitro*. iLS cells were cultured in the presence of 10 ng/ml IL-3 on TSt-4 stromal cells for 10 days (*upper*). FACS profiles before (*lower, left*) and after (*lower, right*) the culture are shown. (F) T and NK lineage potential of iLS cells *in vitro*. iLS cells were cultured in the presence of 5 ng/ml IL-7 on TSt-4/DLL1 stromal cells for 3 wks (*upper*). FACS profiles of before (*lower, left*) and after (*lower, right*) the culture are shown. (G-K) Hematopoietic reconstitution potential of iLS cells. iLS cells were transferred into γ-ray-irradiated NOG mice, then primary and secondary lymphoid organs (BM: H and K, thymus: I and spleen: J) were analyzed at 4 (H,I) or 6 (J,K) wks after transfer. (L) Forward Scatter (FSC) vs Side Scatter (SSC) profiles of iLS cells before (day 0) and after (day 6) the culture for B cell induction. (M) Comparison of cell proliferation between iLS cells and iLS-derived B cells. 1 x 10^4 of iLS cells and iLS derived-B cells (at day 6 of the culture) were harvested and further cultured for 3 days. The numbers of GFP^+ cells were analyzed by FACS.

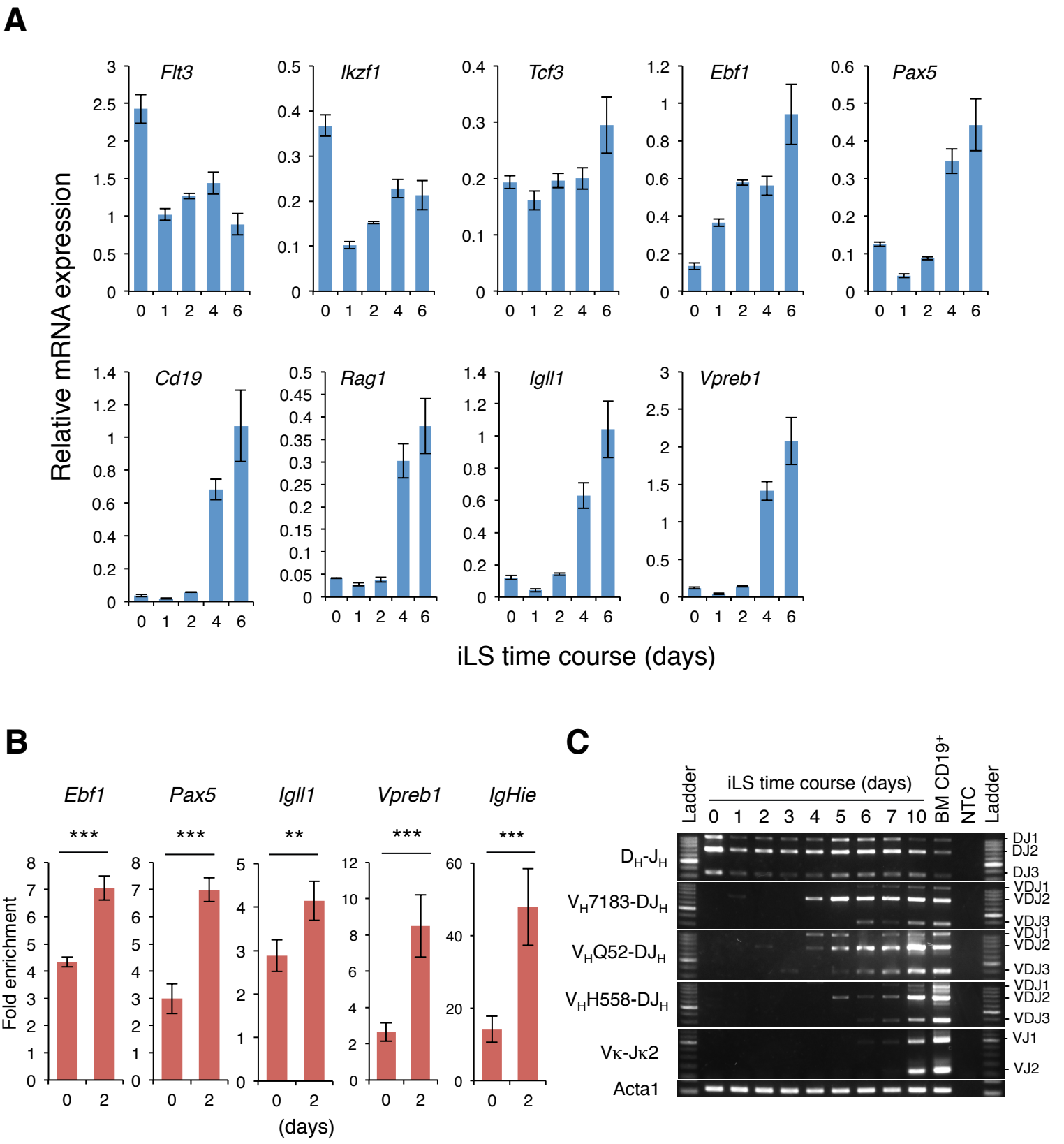

Figure S2. Induction of iLS cells into B lineage cells. (A) qRT-PCR analysis of TFs (*Tcf3*, *Ebf1*, *Pax5*), *Cd79a* (*mb-1*), *Igll1* (15) for B lineage-associated genes, and *Flt3* and *Cd34* for stem cell-associated genes. Values represent means ± SD in representative of two independent experiments. (B) ChIP-qPCR analysis for E2A occupancy at the indicated loci. Values represent means ± SD in representative of three independent experiments. ** p<0.01, *** p<0.001 (C) VDJ rearrangement in Igh locus during B cell differentiation. Acta1 is shown as a loading control.

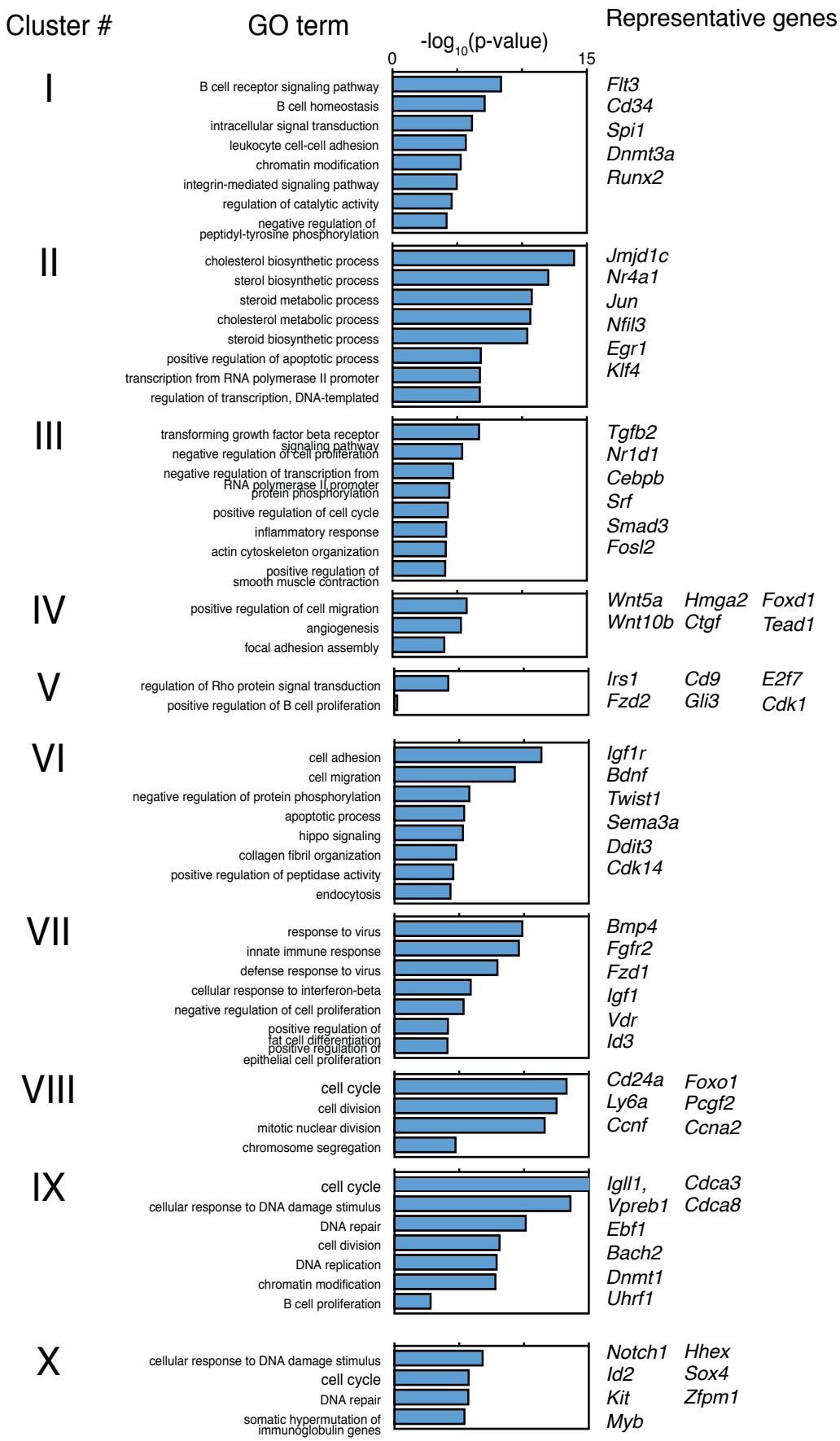

Figure S3. Characteristic gene ontologies of gene clusters in Figure 1F. Statistical significances of gene enrichment are indicated as the bar graphs of their p-values.

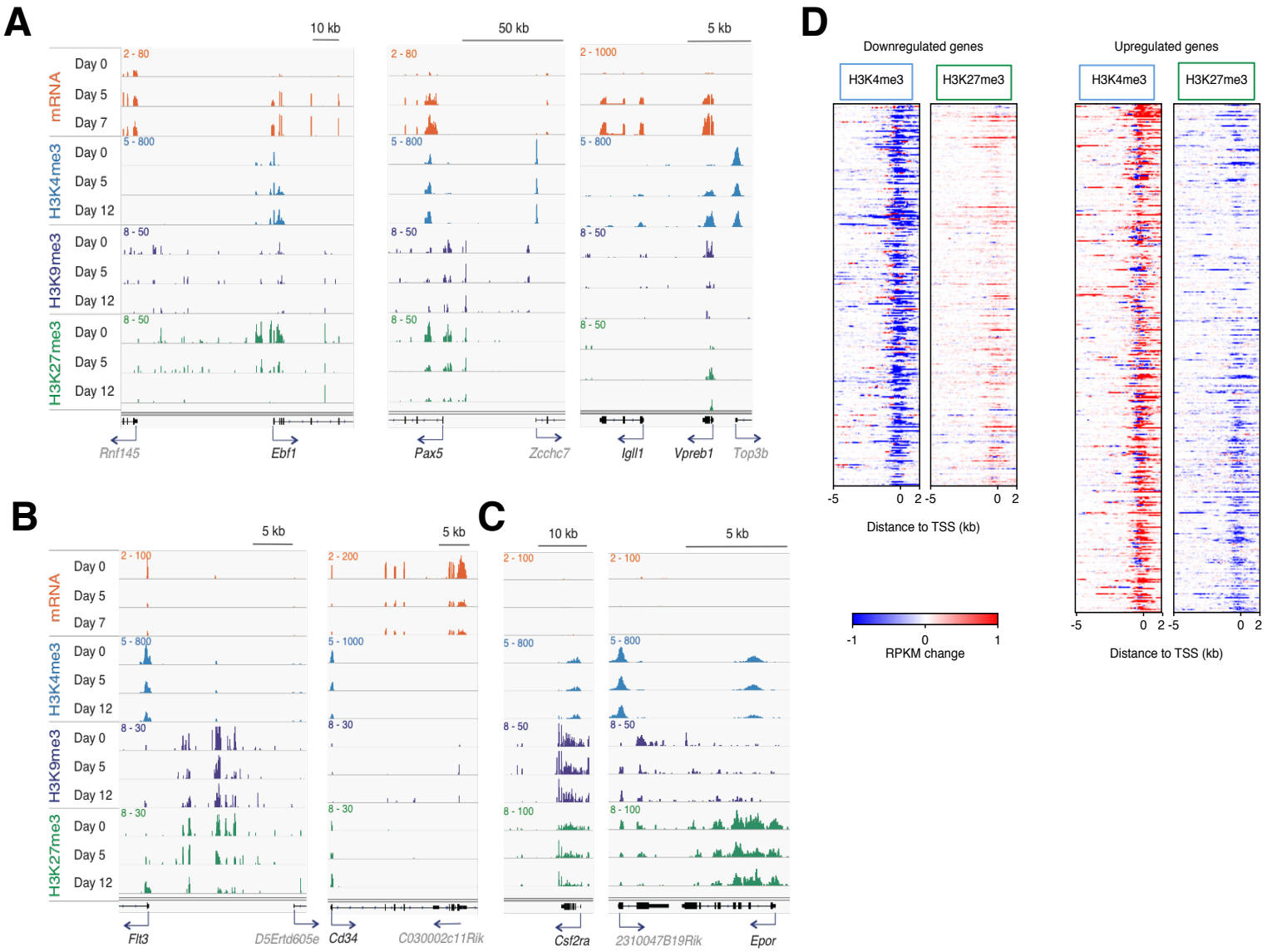

Figure S4. Histone modification during B cell lineage commitment. (A-C) ChIP-sequencing analysis using H3K4me3 (light blue), H3K9me3 (dark blue) and H3K27me3 (green) –specific antibodies of iLS cells differentiating into B cells. Loci of B cell-associated genes: *Ebf1*, *Pax5*, *Igll1* and *Vpreb1* (A), stem cell-associated genes: *Flt3* and *CD34* (B) and other lineage-associated genes: *Csf2ra* and *Epor* (C) are shown, respectively. (D) Alterations of H3K4me3 and H3K27me3 occupancy around TSS of down-regulated and up-regulated (over 2-fold difference with high significance ($q < 0.01$) between day 0 and day 6) gene loci in iLS cells differentiating into B cells.

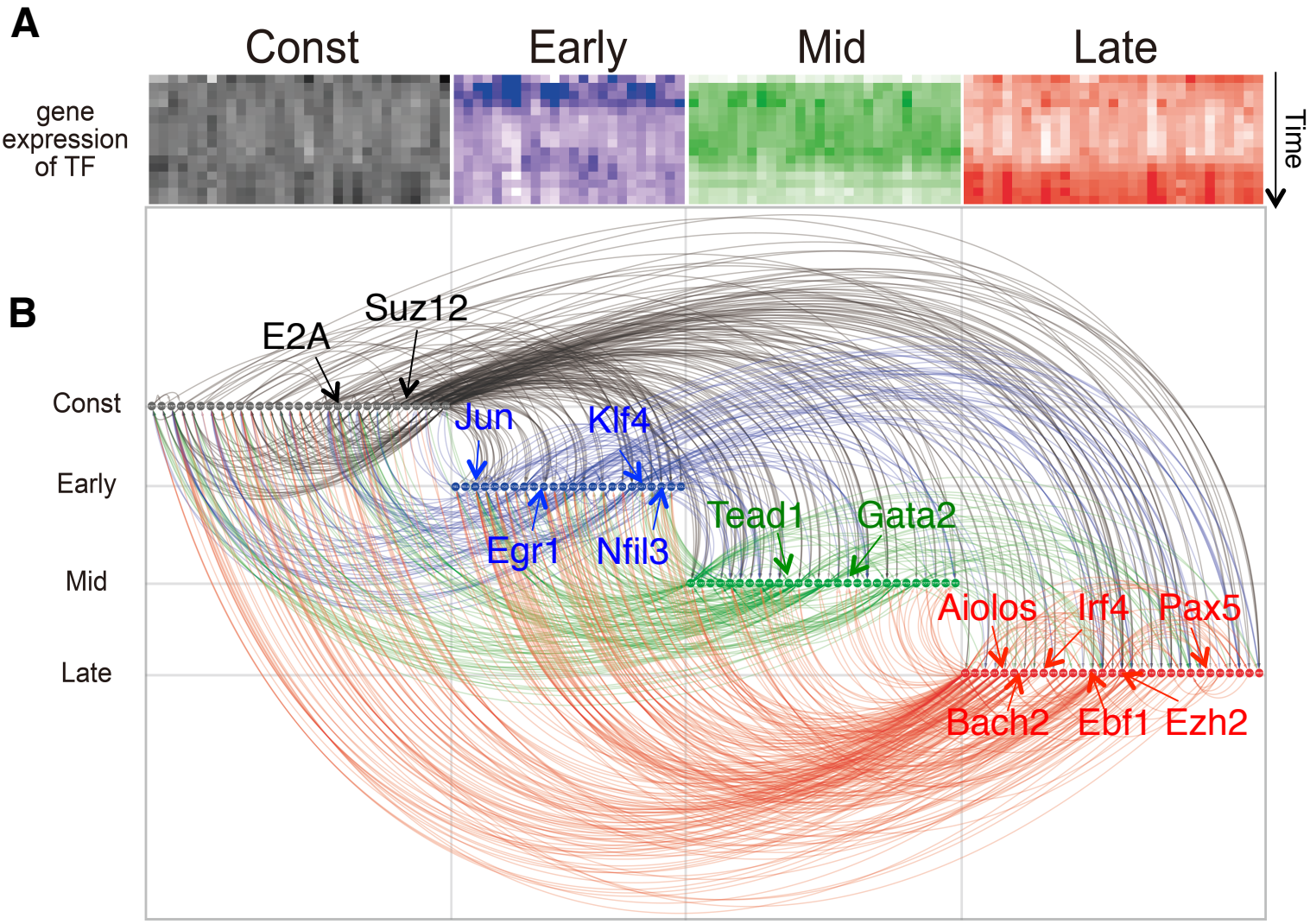

Figure S5. Direct connection of TFs in the TF network during B cell commitment. (A) Gene expression profiles of 115 TFs (33 constant, 27 early, 23 mid and 32 late) which comprise the transcriptional network. Time phases are parallel to Figure 3C. Each line represents the expression pattern of the factor shown in the node of network (B). Details are shown in Supplementary Table S4. (B) Inter-regulation between time phase-specific TFs. Frequently associated interactions (detected in over 70% of ChIP-seq datasets) are represented as an edge. Regulatory connections of individual genes are listed in Supplementary Table S5.

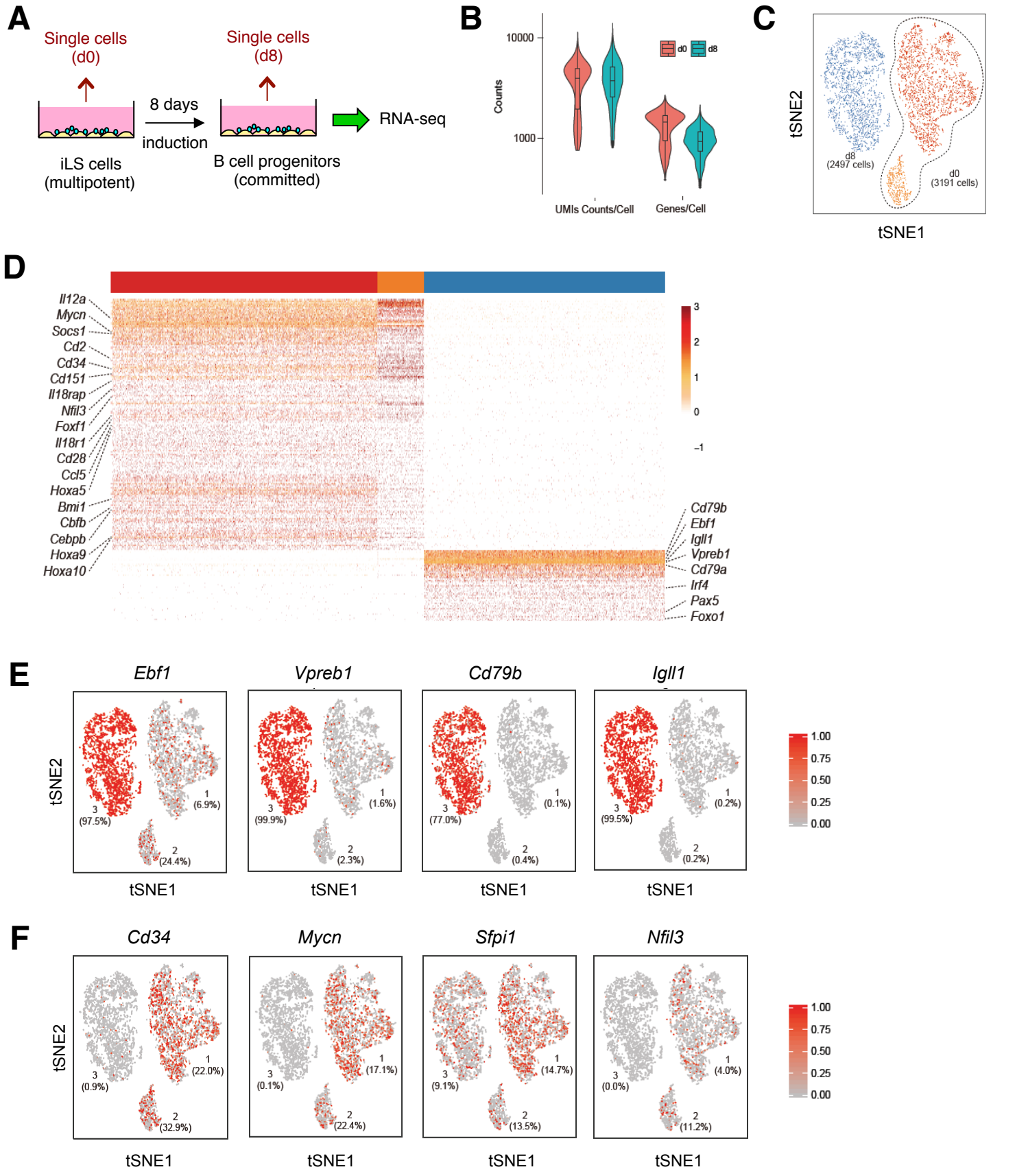

Figure S6. Digital single cell RNA-sequencing analysis during B cell commitment. (A) Scheme for single cell RNA-seq analysis using iLS cells. (B) Distribution of number of UMI counts (left) and genes (right) detected per about 3000 iLS cells. (C) t-SNE projection of multipotent (d0) and B-committed (d8) progenitors isolated from iLS cells. (D) Hierarchical clustering of gene expression data. The representative genes enriched in each cluster are displayed in the heat map, showing normalized gene expression on a $\log_2$ scale from white to red (low to high). (E-F) tSNE projection of iLS cells, with each cell colored based on their normalized expression of *Ebf1*, *Vpreb1*, *Cd79a*, *Igll1*, *Cd34*, *Mycn*, *Sfpi1* and *Nfil3*.

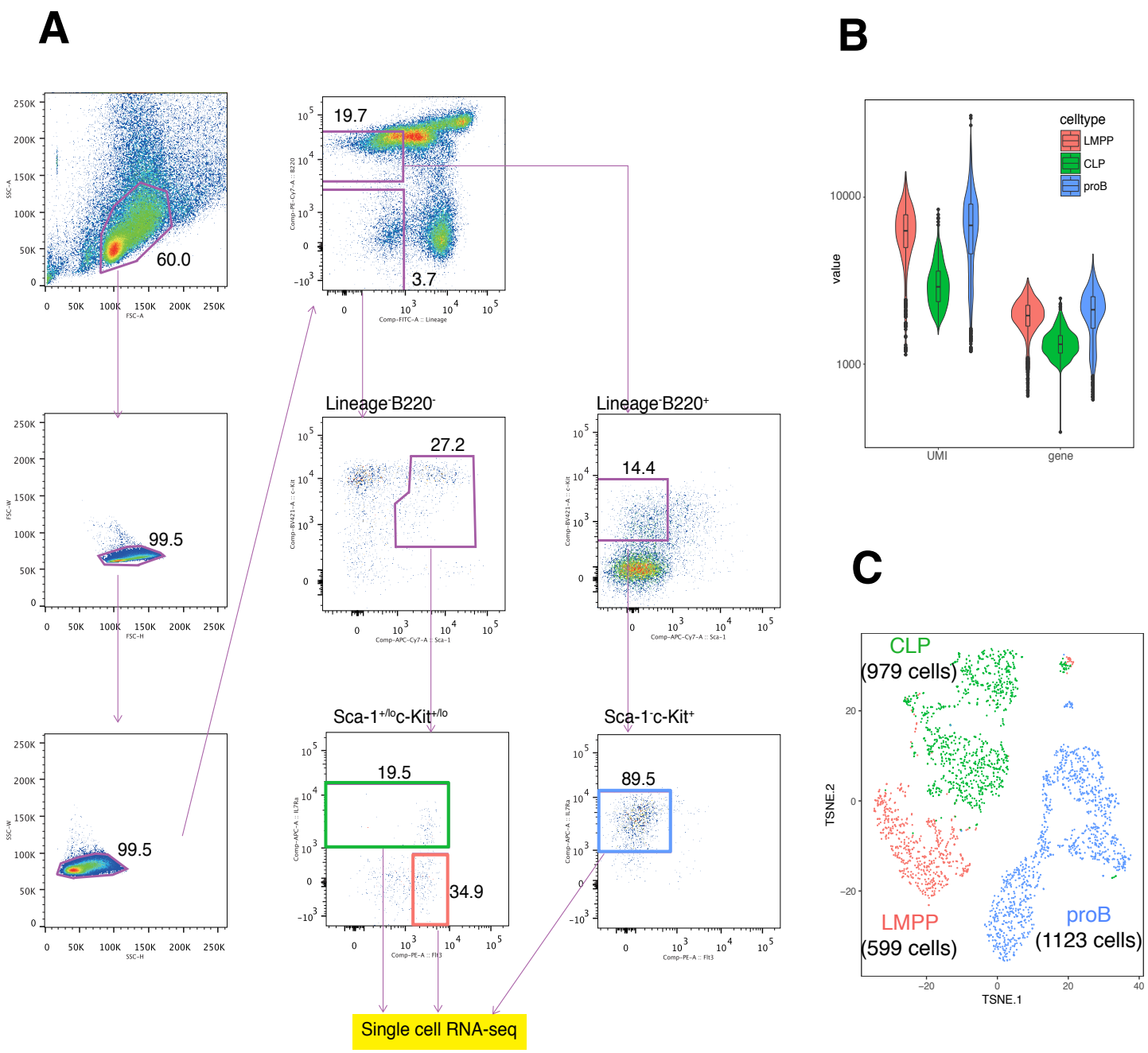

Figure S7. Single cell RNA sequencing analysis using BM-derived B cell progenitors. (A) Gating strategies for isolation of LMPP, CLP and pro-B cells. Lineage antibody cocktail contains anti-CD3e, CD8, CD11b, CD11c, CD25, IgM, IgD, TER119, NK1.1 and Gr-1 antibodies. (B) Distribution of number of UMI counts (left) and genes (right) in each cell types. (C) t-SNE projection of LMPP, CLP and pro-B cells.

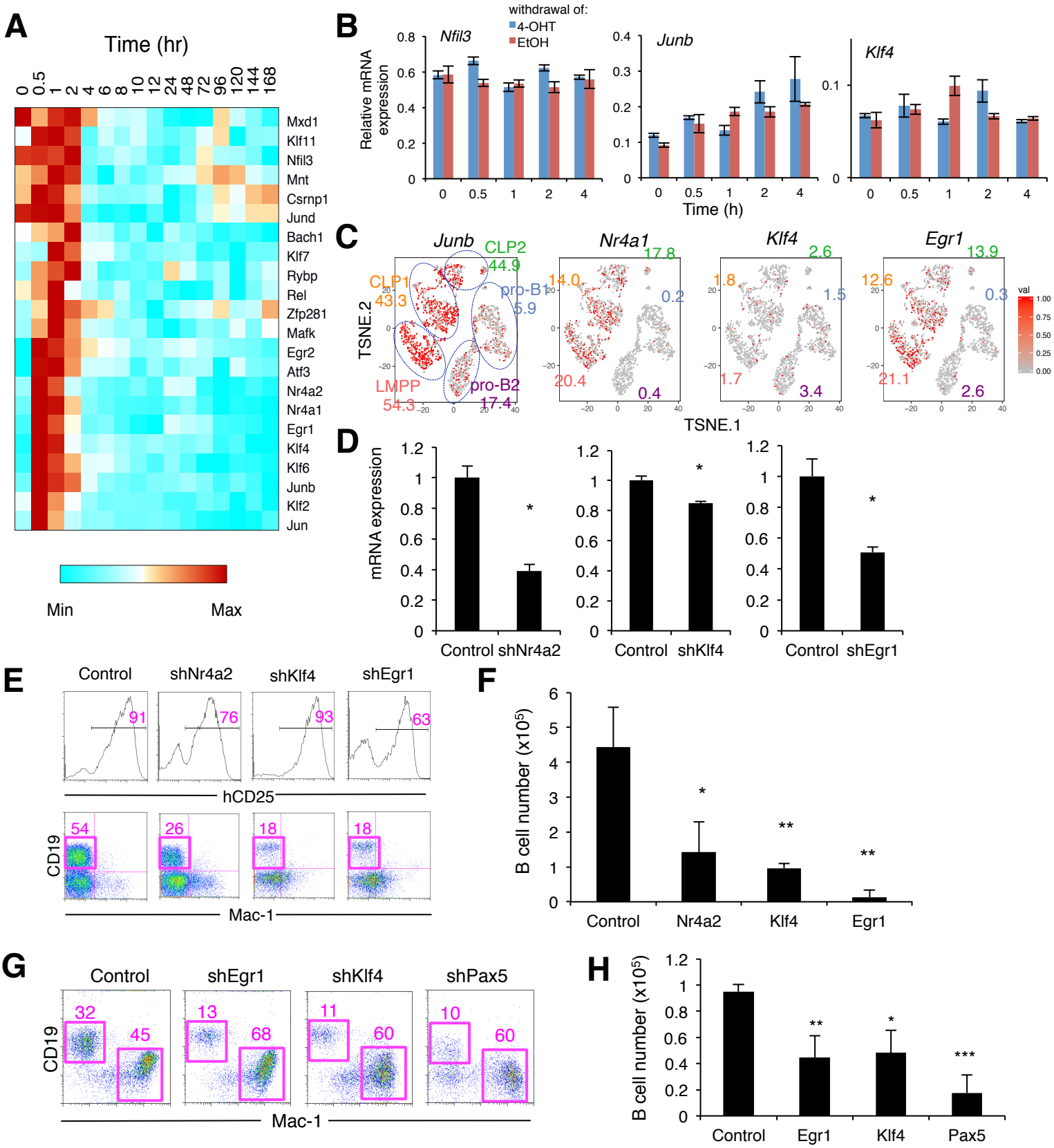

Figure S8. Early responding TFs are essential for B cell differentiation. (A) Gene expression profiles of transcription factors among cluster II (1st wave) genes in Figure 1F. (B) qRT-PCR analysis of indicated genes in Id3-transduced progenitors after withdrawal of 4-OHT or EtOH (as vehicle control) at an early time point. Values represent mean ± SD in representative of two independent experiments. (C) tSNE projection of LMPP, CLP and pro-B cells in BM, with each cell colored based on their normalized expression of *Junb*, *Nr4a1*, *Egr1* and *Klf4*. The numbers indicate the frequency of expressing cells in each cluster. (D-H) shRNA-mediated KD of indicated genes using iLS cells (D-F) or BM LSK cells (G, H). KD efficiency (D), flow cytometric profiles (E, G) and B cell numbers (F, H) are shown. shRNA-transduced cells (hCD25$^+$) are gated for Mac-1 vs CD19 profiles. Values represent mean ± SD in three independent experiments. * $p<0.05$, ** $p<0.01$, *** $p<0.001$. compared to control.

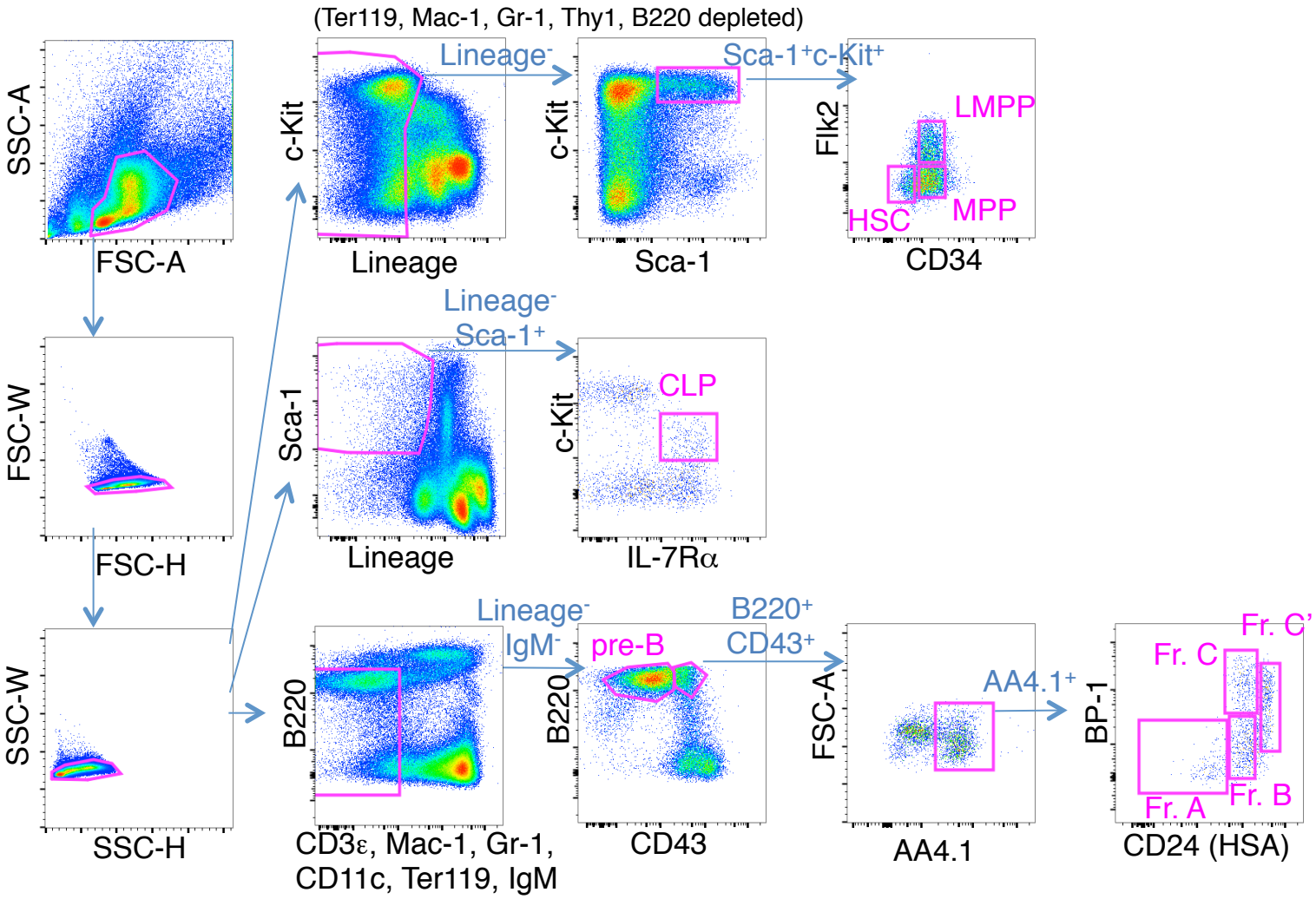

Figure S9. Gating strategies for isolation of hematopoietic populations in BM. Cell debris and doublets were excluded by width/height plot of forward and side scatter. The lineage cocktail contains CD3ε, CD11b, TER-119, B220, Gr-1, CD8α and NK1.1.

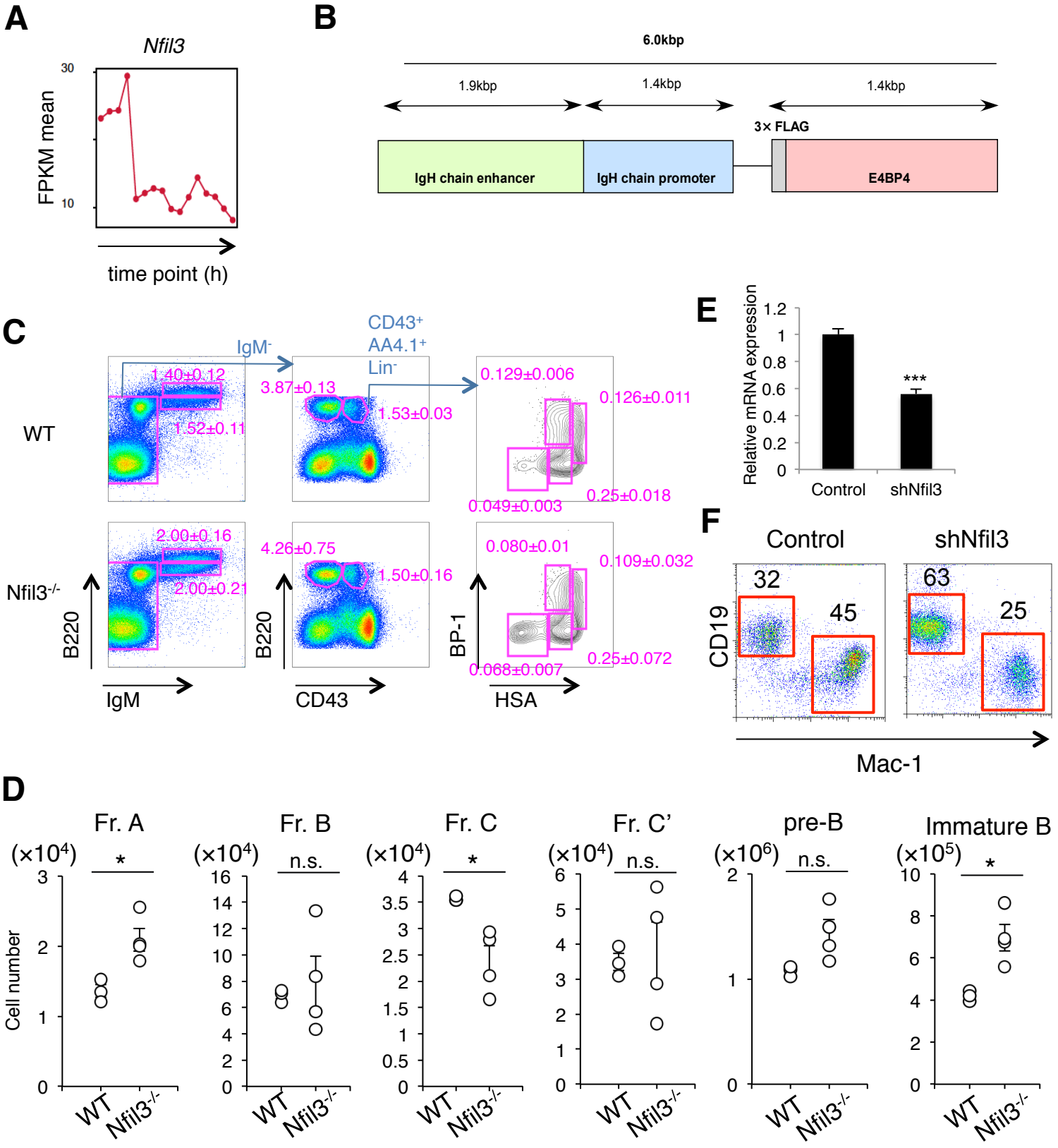

Figure S10. Critical role of *Nfil3* in B lineage commitment. (A) Expression pattern of *Nfil3* in iLS cells differentiating into B cells by RNA-seq analysis. (B) Schematics of the Eμ-Nfil3 vector. (C, D) FACS profiles (C) and actual cell number (D) of B cell progenitor populations in wild type (WT) and *Nfil3*$^{-/-}$ mice. Values represent mean ± SD in a representative of three independent experiments. * p<0.05. (E, F) shRNA-mediated KD of *Nfil3* in BM LSK cells differentiating into B cells. Cells were cultured on TSt-4 stromal cells for 6 days after shRNA induction. KD efficiency (E) and FACS profiles (F) are shown. *** p<0.001.

Supplemental Tables

Table S1. Sequences of shRNA oligonucleotides. Nucleotides shown in magenta are gene-specific shRNA sequences.

Target	Sequence
Cbx2	TGCTGTTGACAGTGAGCGAAGACACTCTATGCTTTGCAAGTAGTGAAGCCACAG ATGTACTTGCAAAGCATAGAGTGTCTCTGCCTACTGCCTCGGA
Uhrf1	TGCTGTTGACAGTGAGCGCGGCCACACACTCTTCGATTATTAGTGAAGCCACAG ATGTAATAATCGAAGAGTGTGTGGCCATGCCTACTGCCTCGGA
Mysm1	TGCTGTTGACAGTGAGCGAATAGATAATGTGAGAAGGAAGTAGTGAAGCCACAG ATGTACTTCCTTCTCACATTATCTATGTGCCTACTGCCTCGGA
Pax5	TGCTGTTGACAGTGAGCGAGCAGAGCGAGTCTGTGACAATTAGTGAAGCCACAG ATGTAATTGTCACAGACTCGCTCTGCCTGCCTACTGCCTCGGA
Nr4a2	TGCTGTTGACAGTGAGCGACCTGTCACTCTTCTCCTTTAATAGTGAAGCCACAGA TGTATTAAAGGAGAAGAGTGACAGGCTGCCTACTGCCTCGGA
Klf4	TGCTGTTGACAGTGAGCGCTCTGCCTTGCTGATTGTCTATTAGTGAAGCCACAGA TGTAATAGACAATCAGCAAGGCAGATTGCCTACTGCCTCGGA
Egr1	TGCTGTTGACAGTGAGCGCATTCTTGATGTGAAGATAATTTAGTGAAGCCACAGA TGTAAATTATCTTCACATCAAGAATATGCCTACTGCCTCGGA
Nfil3	TGCTGTTGACAGTGAGCGACTGCTCTCCCTGAAATTAAAGTAGTGAAGCCACAG ATGTACTTTAATTTCAGGGAGAGCAGCTGCCTACTGCCTCGGA

Table S2. Primer sequences for qRT-PCR, ChIP-qPCR and rearrangement PCR.

Primer sequences for qRT-PCR

	Forward	Reverse
Flt3	AGTTCACCAAAATGTTCACG	ACTTCTTCCAGGTCCAAGAG
Ikzf1	GAGCTGGACTTGTGTCTTTG	TTTCACACCTTGGTGAACTC
Tcf3	CAGAACTGGAAACAAAGCAG	TGGTGGCTCTGAAGTAGAAG
Ebf1	TGGGTTACAGGTCATATTCG	GAACTGCTTGGACTTGTACG
Pax5	CATTCGGACAAAAGTACAGC	GATGCCACTGATGGAGTATG
Cd19	CAGTGATGGGACTAGCAGAC	GTAGTGTTGCCAGAAACTCG
Rag1	GAAGCTTCTGGCTCAGTCTACATCT	ACCCTCATAGCGCTGCAGGTT
Igll1	GTTCTAATGGGATGCTAGGC	AGCGTCCTTCTCTTATCAGG
Vpreb1	GAGTGGGAAGGAGAAAAGTC	CCTTCCCATACCAGACTAGC
Nr4a2	ATGCTGGTGCTAGTGTCTGC	TGGATCCCCAGCTTTCTCAAG
Klf4	GCACACCTGCGAACTCACAC	GTTTGCGGTAGTGCCTGGTC
Egr1	ACGACAGCAGTCCCATCTACTCGG	GGACTCGACAGGGCAAGCATATGG
Nfil3	GAACTCTGCCTTAGCTGAGGT	ATTCCCGTTTTCTCCGACACG
Ikzf3	GACACGTGCCCTATGACAACAGCAG	GCATGCGTAGTTGCAGAGGTGACAC
Bach2	TGGATGGTTGACTTTGTTTC	AGGGTTTCACTTTTGCCTAC
Irf4	TGGGTAAACACTCCTGACAC	AGTGTTCCTCAGCCTACCTC
Jun	TTGCCCCAACAGATCCCGGT	TGCGGTTCCTCATGCGCTTC
Cbx2	CAAACACAACAGCTGGGAGC	GGTCTCTTGCCTCTCTTCCG
Mysm1	ATAGGACAGTCAGCCTCCTC	AAAAGCTAGGTTGAGGGCTA
Uhrf1	TGAAGTACTGGCCAGAGAGA	CAGGGTACTGCATAGTGAGC
Dnmt1	ATGGTGTTGTCTACCGACTG	CAGAATACTTGCGGTAGTGC

Primer sequences for ChIP-qPCR

	Forward	Reverse
Ebf1	CCACGCTCCTTCCAGTTTAG	ATCCTTCCGCCTTATCACAG
Pax5	CCTGTACTAGCACAATTGTCAACC	CTGAATGGGACCTAGATACTGTAGTG
Igll1	GGGTTAAGACAGGCAGCTGTGAG	CAAACCCCAGGCTGTCTCTAGTT
Vpreb1	TGCCAAGCTGGCCATGTGAACAC	GATGTTCCTCTACCATATGTGAG
IgHie	GGTGGGGCTGGACAGAGTGTT	TCAGAACCAGAACACCTGCAG

Primer sequences for rearrangement PCR

DHL	MTTTTTGTSAAGGGATCTACTACTGTG
VH7183	GAASAMCCTGTWCCTGCAAATGASC
VHQ52	ACTGARCATCASCAAGGACAAYTCC
VHJ558	CARCACAGCCTWCATGCARCTCARC
JH3	CTCACAAGAGTCCGATAGACCCTGG
mVkdeg-F1	GSTTCAGTGGCAGTGGRTCWGGRAC
mJk2-R1	TTTCCAGCTTGGTCCCCCCTCCGAA
Acta1-F	GGCATCGTGTTGGATTCTG
Acta1-R	CACGAAGGAATAGCCACGC