Supplementary Data (Revilla-i-Domingo et al.)

Supplementary Materials and Methods

Mice

The following mice were maintained on the C57BL/6 background: *Pax5*$^{+/-}$ (Urbánek et al, 1994), *Pax5*$^{fl/fl}$ (Horcher et al, 2001), *Pax5*$^{Bio/Bio}$ (McManus et al, 2011), *Rag2*$^{-/-}$ (Shinkai et al, 1992), Eβ$^{-/-}$ (Bouvier et al, 1996), *Rosa26*$^{BirA/BirA}$ (Driegen et al, 2005), transgenic *Vav*-Cre (de Boer et al, 2003) and transgenic *Cd23*-Cre (Kwon et al, 2008) mice. All animal experiments were carried out according to valid project licenses, which were approved and regularly controlled by the Austrian Veterinary Authorities.

Flow cytometry

Mice at the age of 4-5 or 6-10 weeks were used for FACS analyses of early lymphoid progenitors and mature B cells, respectively. Single cell suspensions of bone marrow and lymph nodes were prepared and unspecific antibody binding was inhibited by incubation with CD16/CD32 Fc block (BD Biosciences). Bone marrow and lymph node cells were stained for flow cytometry with the following antibodies: anti-B220/CD45R (RA3-6B2), CD3ε (145-2C11), CD4 (GK1.5), CD5 (53-7.3), CD8a (2.43), CD11b/Mac1 (M1/70), CD11c (HL3), CD19 (1D3), CD21 (7G6), CD23 (B3B4), CD25/IL-2Rα (PC61), CD43 (S7), CD49b (DX5), CD117/c-Kit (ACK4), CD127/IL-7Rα (A7R34), CD135/Flt3 (A2F10.1), Gr1 (RB6-8C5), IgD (1.19), IgM^a (Igh-6a/DS-1), IgM^b (AF6-78), Ly6C (6C3), Ly6D (49H4.3), NK1.1 (PK136), Sca1/Ly6A (D7), TCRβ (H57-597) and Ter119 (TER119) antibodies.

FACS and MACS sorting

FACS sorting was used to isolate pro-B and mature B cells with a purity of > 95% for RNA sequencing (Supplementary Figures S5A, S7B and S8A). The purity of the sorted ALPs and BLPs, which were used for the replica #1 RNA-seq experiment, was 94% and 99%, respectively (Supplementary Figure S7A), while the purity of sorted ALPs and BLPs used for the replica #2 experiment was 85%. For isolating ALPs and BLPs, lineage-positive cells (Lin$^+$) cells were first depleted from the bone marrow by MACS sorting using the following lineage (Lin) marker antibodies: CD19, CD3ε, CD4, CD8a, DX5, Gr1, Ly6C, Mac1 and Ter119. Subsequently, ALPs (Lin$^-$B220$^-$IL-7Rα$^+$Flt3$^+$Sca1^{lo}c-Kit^{lo}Ly6D$^-$) and BLPs (Lin$^-$B220$^-$IL-7Rα$^+$ Flt3$^+$Sca1^{lo}c-Kit^{lo}Ly6D$^+$) were sorted with a FACS Aria machine (Becton Dickinson; Supplementary Figure S7A). Wild-type and *Rag2*$^{-/-}$ pro-B cells were sorted from the bone marrow as CD19$^+$B220$^+$c-Kit$^+$ CD25$^-$IgM$^-$ cells (Supplementary Figure S7B). Short-term cultured *Pax5*$^{fl/fl}$ pro-B cells were sorted as CD19$^+$B220$^+$ cells and *Pax5*$^{Δ/Δ}$ pro-B cells as CD19$^-$B220$^+$ cells (Supplementary Figure S5A). Wild-type mature B cells were purified from lymph nodes by MACS depletion of non-B cells with anti-PE beads after staining with PE-labeled TCRβ, CD4, CD8a, Mac1, Gr1, DX5 and Ly6C antibodies (Supplementary Figure S3A). *Pax5*$^{Δ/Δ}$ mature B cells were FACS-

sorted as $B220^{+}IgD^{lo/-}CD25^{+}TCR\beta^{-}$ cells (Horcher et al, 2001) from the lymph nodes of *Cd23*-Cre $Pax5^{fl/fl}$ or *Cd23*-Cre $Pax5^{fl/-}$ mice (Supplementary Figure S8A).

In vitro culture of lymphoid progenitors

$Pax5^{-/-}$, $Pax5^{\Delta/\Delta}$, $Rag2^{-/-}$, $Pax5^{Bio/Bio}$ $Rag2^{-/-}$ and wild-type pro-B cells were cultured on OP9 cells in IL-7-containing IMDM as described (Nutt et al, 1997).

Cap analysis of gene expression (CAGE)

RNA was isolated from short-term cultured and sorted $Rag2^{-/-}$ pro-B cells (see Supplementary Figure S5A) as well as from sorted mature B cells from wild-type lymph nodes (see Supplementary Figure S3A). Total RNA (~50 μg) of these two cell types was sent to ImaGenes (Berlin) for the prepartion of a CAGE library, which was subsequently analyzed by Solexa deep sequencing.

Genome-wide mapping of DNase I hypersensitive sites

Pro-B cells for DNase I hypersensitive site mapping were prepared by short-term culturing of $B220^{+}$ bone marrow cells from $Rag2^{-/-}$ and $Pax5^{-/-}$ $Rag2^{-/-}$ mice in the presence of OP9 cells and IL-7. Mature B cells were directly isolated from the lymph nodes of $E\beta^{-/-}$ mice lacking mature T cells (Bouvier et al, 1996; Supplementary Figure S1D). Live pro-B cells and mature B cells were enriched by density gradient purification using Lympholyte M (Cedarlane Laboratories) prior to DNase I treatment, which was carried out on permeabilized cells as previously described with minor modifications (Cappabianca et al, 1999). Briefly, cells were washed with ice cold ψ buffer (11 mM phosphate buffer, pH 7.4, 108 mM KCl, 22 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM DTT) and then treated with various concentrations of DNase I (Worthington Biochemicals, Lakewood, NJ) in ψ buffer containing 0.2% Nonidet P-40 for 6 minutes at 20°C. DNase I treatment was terminated by adding an equal volume of 2x lysis buffer (50 mM Tris-HCl, pH 8, 20 mM EDTA pH 8, 1% SDS, 500 μg/ml proteinase K) followed by overnight incubation at 55°C. After RNase A digestion, the DNA was isolated by phenol-chloroform extraction and isopropanol precipitation. DNA fragments between 150 bp and 250 bp were isolated by sucrose gradient centrifugation as previously described with minor modifications (Sabo et al, 2006). Briefly, the DNA solution in STE buffer (1 M NaCl, 20 mM Tris-HCl, pH 8, 5 mM EDTA) was overlaid onto a discontinuous sucrose gradient (10-30%) prior to centrifugation for 16-24 hours at 25°C at 30,000 rpm in a swinging SWTi-40 rotor in a Beckman ultracentrifuge. Fractions of 250 μl were collected from the top of the gradient and tested for efficient enrichment of DNase I 'double-hit' fragments released from DNase I hypersensitive (DHS) sites by quantitative PCR amplifying a DHS site at the *Cd19* enhancer or *Tbp* promoter compared to an inactive intergenic region on chromosome 1 (control). Fractions enriched for DHS sequences were pooled and subjected to Solexa deep sequencing.

ChIP-seq analysis of histone modifications

Pro-B cells isolated from the bone marrow of *Rag2*$^{-/-}$ mice were expanded *in vitro* for 4-5 days on OP9 cells in the presence of IL-7, and mature B cell were purified from lymph nodes by MACS-sorting prior to chromatin immunoprecipitation (ChIP) with histone modification-specific antibodies as described in detail (Schebesta et al, 2007). Rabbit polyclonal antibodies recognizing the following histone tail modifications were used for ChIP analysis: H3K4me1 (ab8895) and H3K4me3 (ab8580) from Abcam, H3K4me2 (07-030) and H3K9ac (07-352) from Upstate Biotechnology and H3K27me3 from T. Jenuwein (MPI Freiburg). The precipitated DNA was quantified by real-time PCR analysis with SYBR green, which was carried out on a MyiQ instrument (Bio-Rad) as described (Decker et al, 2009). The ChIP-precipitated DNA (≥ 5 ng) was submitted to Solexa deep sequencing.

Bio-ChIP-seq analysis of Pax5 binding

Pro-B cells isolated from the bone marrow of *Pax5*$^{Bio/Bio}$ *Rag2*$^{-/-}$ mice were expanded *in vitro* for 4-5 days on OP9 cells in the presence of IL-7, and mature *Pax5*$^{Bio/Bio}$ B cell were purified from lymph nodes by MACS-sorting. About 5x10^7 pro-B cells or 1x10^7 mature B cells were used for chromatin precipitation by streptavidin pulldown (Bio-ChIP), which has recently been described in detail (Ebert et al, 2011). The precipitated genomic DNA was quantified by real-time PCR, and 5-10 ng of ChIP-precipitated DNA was submitted to Solexa deep sequencing.

cDNA preparation for RNA-sequencing

Total RNA was isolated from ex vivo sorted or short-term cultured cells using the RNeasy Plus Mini Kit (Qiagen) as described in the manufacturer's manual. This kit allows efficient elimination of genomic DNA. RNA quality and concentration was assessed using the RNA 6000 Nano kit (Agilent) in a 2100 Bioanalyzer (Agilent) following manufacturer's instructions. Low yields (≤ 1 μg) of total RNA were obtained from the following sorted cell populations: 70-85 ng from 4-5x10^5 *Pax5*$^{\Delta/\Delta}$ mature B cells, 200-500 ng from 1.2-5x10^5 ALPs or BLPs, ~300 ng from 3x10^5 wild-type pro-B cells and ~1 μg from 7-8x10^5 *Rag2*$^{-/-}$ pro-B cells. Higher yields (> 1 μg) of total RNA were obtained from following cell types: 1-2 μg from 3-5x10^6 wild-type mature B cells and 2-4 μg from 6-9x10^5 short-term cultured wild-type or *Pax5*$^{\Delta/\Delta}$ pro-B cells. Samples with a higher yield of total RNA (> 1 μg) were processed for deep sequencing as described (Mortazavi et al, 2008) with modifications that preserved strand specificity (Parkhomchuk et al, 2009). Modifications that required less starting RNA but did not preserve strand specificity were used for samples with lower yields of total RNA (≤ 1 μg). We confirmed that both protocols (with or without strand-specificity) gave similar gene expression values (with a Pearson correlation of 0.92), when applied to the same RNA sample (data not shown).

mRNA was obtained by subjecting total RNA to one or two rounds of poly(A) selection using the Dynabeads mRNA purification kit (Invitrogen) following the manufacturer's instructions. The mRNA was fragmented by addition of 5× fragmentation buffer (200 mM Tris acetate, pH 8.2, 500 mM potassium acetate, 150 mM magnesium acetate) followed by heating at 94^{o}C for 3 min. Fragmented mRNA was cleaned by following the RNA clean-up protocol of the

RNeasy Mini Kit (Qiagen). The size distribution was assessed using the RNA 6000 Pico kit (Agilent) in a 2100 Bioanalyzer (Agilent) to make sure that the peak of RNA fragmentation was between 200 and 500 nucleotides.

Fragmented mRNA from samples with higher RNA yields (> 1 μg) was used as template for the first-strand cDNA synthesis reaction with random hexamers, using the Superscript III First-Strand Synthesis System for RT-PCR (Invitrogen) according to the manufacturer's manual (with 1.5 μl of 50 ng/μl random hexamers in a total reaction volume of 30 μl). dNTPs were subsequently removed by purifying the first-strand cDNA synthesis reaction with a Mini Quick Spin Column for DNA (Roche) according to the manufacturer's instructions. The second-strand cDNA synthesis reaction was prepared in a total volume of 210 μl, containing the purified first-strand reaction, 1.5 μl of 50 ng/μl random hexamers (to restore the random hexamers removed during purification), 42 μl of 5x second-strand buffer (Invitrogen), 0.42 μl of 100 mM dATP, dCTP, dGTP and dUTP, 5.6 μl of 10 U/μl *E. coli* DNA polymerase I (Invitrogen), 1.4 μl of 10 U/μl *E. coli* DNA ligase (Invitrogen), 1.4 μl of 2 U/μl RNase H. The reaction was incubated at 16°C for 2 h. The reason for using dUTP and not dTTP in the second-strand reaction was to allow elimination of the second strand during library preparation (see below), thereby preserving strand specificity (Parkhomchuk et al, 2009).

Fragmented mRNA from samples with lower RNA yields (≤ 1 μg) was used as template for first-strand cDNA synthesis in a volume of 40 μl using random hexamers and the Superscript VILO cDNA synthesis kit (Invitrogen) according to the manufacturer's instructions. The second-strand cDNA synthesis reaction was prepared in a total volume of 400 μl, containing the first-strand reaction, 80 μl of 5x second-strand buffer (Invitrogen), 0.2 μl of 100 mM dATP, dCTP, dGTP and dTTP, 10.7 μl of 10 U/μl *E. coli* DNA polymerase I (Invitrogen), 2.7 μl of 10 U/μl *E. coli* DNA ligase (Invitrogen), 2.7 μl of 2 U/μl RNase H. The reaction was incubated at 16°C for 2 h.

cDNA was cleaned using the MinElute Reaction Cleanup kit (Qiagen) following manufacturer's instructions, eluting with 20 μl of elution buffer (10 mM Tris-Cl pH 8.5). The cDNA concentration was measured using the Quant-iT Picogreen dsDNA Assay (Invitrogen) in a NanoDrop Fluorospectrophotometer ND-3300 according to the manufacturer's instructions. We obtained the following yields of cDNA from the different cell types: 3-9 ng from ALPs or BLPs, ~10-25 ng from *Rag2*$^{-/-}$ or wild-type pro-B cells, ~3 ng from *Pax5*$^{\Delta/\Delta}$ mature B cells, 10-20 ng from short-term cultured wild-type or *Pax5*$^{\Delta/\Delta}$ pro-B cells and ~3 ng from wild-type mature B cells. About 3-5 ng of cDNA were used for Solexa deep sequencing as described below.

Solexa deep sequencing

About 5 ng of ChIP-precipitated DNA, DNA excised from DHS sites or cDNA prepared from mRNA was used as starting material for generating single-end or paired-end sequencing libraries as described by Illumina's ChIP Sequencing sample preparation protocol. DNA fragments of the following sizes were selected for the different experiments: 150–250 bp for DHS site mapping, 200–350 bp for ChIP-seq and 150–700 bp for RNA-seq. For strand-specific RNA-sequencing,

the uridines present in one cDNA strand were digested at 37°C for 30 min in 57 µl of 1x TE buffer pH 8 with 5 units of uracil-N-glycosylase (New England Biolabs) as described (Parkhomchuk et al, 2009) followed by PCR amplification. Completed libraries were quantified using the Agilent Bioanalyzer dsDNA 1000 assay kit and Agilent QPCR NGS library quantification kit. Cluster generation and sequencing was carried out by using the Illumina/Solexa Genome Analyzer (GA) II and IIx systems according to the manufacturer's guidelines. Sequencing of the paired-end libraries (DHS site mapping) yielded a read length of 76 bp for each end of a DNA fragment, whereas a read length of 36 bp was obtained with single-end libaries (ChIP-seq and RNA-seq).

Sequence alignment

Sequence reads that passed the Illumina quality filtering were considered for alignment. In case of RNA-seq experiments, reads corresponding to mouse ribosomal RNAs (BK000964.1) and non-polyadenylated replication histone mRNAs were removed. The remaining reads were aligned against the mouse genome assembly version of July 2007 (NCBI37/mm9) using the Bowtie program version 12.5 (Langmead et al, 2009), allowing up to two mismatches and ignoring any read that would map more than once in the genome. In detail, the alignment was done using the following parameters: –m 1 –v 2 –best –strata –tryhard. RNA-Seq cleaning was performed using the following settings: –v 3 –k 1 --tryhard --chunkmbs 256.

Database of RefSeq-annotated genes

The peak-to-gene assignment and calculation of RNA expression values were all based on the RefSeq database, which was downloaded from UCSC on April 29th, 2010. The annotation of three additional transcripts (*Ebf1*, *Igll1* and *Pax5*) of interest for our study was added, which resulted in a final list of 26,814 transcripts. For peak-to-gene assignment, overlapping transcripts (on the same strand) were merged to single gene models for each transcript set using the JNOMICS software (I. Tamir, unpublished), provided that the transcripts precisely overlapped at least for one exon. Furthermore, the elimination of genes present on genomic contigs (so-called 'random' chromosomes), which were not incorporated into the genomic sequences of the respective chromosome, resulted in 21,428 gene models. In addition, double entries (i.e. exactly the same gene at two different genomic locations), that would be missed during the alignment procedure due to elimination of repeated sequence reads, were not considered for the calculation of RNA expression values from RNA-seq data, which resulted in 21,025 gene models.

Peak calling of ChIP-seq data

Peaks were called using the MACS program version 1.3.6.1 (Zhang et al, 2008) with default parameters, a read length of 36 or 76, a genome size of 2,654,911,517 bp (mm9) and the appropriate input control sample. In cases where the read shifting model could not be built with the default mfold parameter (32), the mfold cutoff was lowered such that sufficient high-quality peaks (100) could be found to build a model. For DHS and chromatin experiments, the read shifting model building was turned off (-nomodel). Peaks were further filtered for p-values of <

10^{-10}. This cutoff is more conservative than the default p-value (10^{-5}), and efficiently removed false positive (unique) peaks of technical replicas (M. Jaritz, data not shown).

We observed that the MACS program called a fraction of Pax5 peaks even at a p-value of < 10^{-10} as significant, although they were very small. When normalizing the sum of read counts in a given peak to the total number of reads in all peaks, these small peaks could be readily identified in a density diagram (Supplementary Figure S4A). We defined this normalized value as RPKMpL (*r*eads *p*er 1 *k*b of peak width per *m*illion reads in peaks *p*er total *l*ength of all peaks). By comparing the peak sets of pro-B and mature B cells (first Bio-ChIP-seq replicas; Supplementary Table S1), we defined a cutoff of 0.7 RPKMpL that removed only 0.5% (112) of all Pax5 peaks (20,725) present in the pro-B cell data set. In contrast, the cutoff of 0.7 RPKMpL eliminated 7,226 small peaks (32%) of 22,694 Pax5 peaks identified in mature B cells. The same procedure did not remove any Pax5 peaks from the second Bio-ChIP-seq replicas of pro-B and mature B cells (Supplementary Table S1).

Identification of CAGE islands and regulatory regions

For each strand, all overlapping or consecutive sequence reads were joined to a CAGE tag cluster using the JNOMICS program (I. Tamir, unpublished). CAGE tag clusters were extended by 200 base pairs on both sides, and joined if overlapping. *R*eads *p*er CAGE tag cluster per *m*illion mapped sequence reads (RPMs) were determined using custom-made Perl scripts. CAGE tag clusters coinciding with annotated transcription end sites (TESs; Carninci et al, 2006) were excluded from further analysis. All other CAGE tag clusters with an RPM value of > 2.6 (Supplementary Figure S1B) identified transcription start sites (TSSs) of expressed genes (Carninci et al, 2006). The threshold of 2.6 RPM was chosen to include 95% of the TSSs of all expressed transcripts with an RPKM value of > 1.

Regulatory regions were defined in the genome of pro-B and mature-B cells by overlapping the corresponding DHS peaks with CAGE tag clusters (RPM > 2.6), using the MULTOVL program (A. Aszodi, unpublished). Distal elements were defined as DHS peaks that did not overlap with CAGE tag clusters (RPM > 2.6). Active promoters were defined as DHS peaks that overlapped with CAGE tag cluster (RPM > 2.6). The most active genes often contained several CAGE tag cluster, all of which overlapped with one single DHS peak. Therefore the number of DHS peaks overlapping with CAGE tag clusters rather than the number of CAGE tag clusters coinciding with DHS peaks was used for determining the number of active promoters (Figure 1A,B).

Peak-to-gene assignment

Peaks were assigned to genes in a stepwise manner by prioritizing genes containing peaks in their promoter and/or gene body. For this, peaks overlapping with the promoter (-2.5 kb to +2.5 kb relative to TSS) or gene body (+2.5 kb to TES) were first assigned to the corresponding gene. Second, all other peaks within a specified region upstream of the TSS or downstream of the TES were assigned to the gene containing peaks in the promoter or gene body. The region was specified to extend up to the promoter (-2.5 kb of TSS) or TES (+2.5 kb) of the nearby gene on

either DNA strand with a maximal region size of 50 kb. If two genes were present within 50 kb of a peak and were not prioritized by the presence of peaks in their promoter or gene body, the peak was assigned to the closest gene. Finally, all non-assigned peaks were classified as intergenic. This assignment procedure was implemented as a custom-made Shell script available from the authors.

Analysis of RNA-seq data

The number of reads aligned to the gene models in each RNA-sequencing experiment was first normalized by using the edgeR program (Robinson et al, 2010). Read numbers from the RNA-sequencing experiment corresponding to replica #1 of the *Rag2*$^{-/-}$ pro-B cell sample was used as reference for normalization of all other RNA-sequencing experiments. Normalized expression values were then calculated as RPKMs (normalized *r*ead number *p*er *k*ilo base pairs of exon model length per *m*illion of reads aligned to all gene models; Mortazavi et al, 2008), using customized Perl scripts.

For calculations of gene expression changes between samples, the exon model length is irrelevant. Therefore, for calculations, representation and statistical analyses of global gene expression changes, RPMs (normalized *r*ead number *p*er *m*illion of reads aligned to all gene models) were used. Differential expression significances [-$\log_{10}$(p-value)] were calculated using the edgeR program to identify differentially expressed genes between ALPs and BLPs, BLPs and *Rag2*$^{-/-}$ pro-B cells, *Rag2*$^{-/-}$ pro-B cells and mature B cells, short-term cultured *Rag2*$^{-/-}$ and *Pax5*$^{\Delta/\Delta}$ pro-B cells, and between mature B cells and *Pax5*$^{\Delta/\Delta}$ mature B cells. A common dispersion of 0.006 (calculated by comparing the two replicates of *Rag2*$^{-/-}$ pro-B cells using edgeR) was used for all differential expression significance calculations. This ensured that the same statistical model was applied for all significance calculations. The calculated significances were used to create bins of genes with different degrees of differential expression. The first bin (containing the genes least differentially expressed) included genes with a significance < 5 (black bars in Figures 4D and 6C). Other bins included genes with significance between 5 and 20 (grey bars), 20 and 47, 47 and 83 and above 83 (with the latter 3 bins corresponding to the blue bars in Figures 4D and 6C). As the dispersion was kept constant for all RNA-seq analyses, the significance is essentially a function of the fold change. A more intuitive definition of the bins can thus be formulated as follows: for a gene expressed at 15 RPMs in the lower expressing cell type, the significances defining the bins (5, 20, 47 and 83) represent 2-, 4-, 8-, and 16-fold changes, respectively.

At least two replica experiments were performed for each cell type, and these replicas always correlated well (Pearson correlation of > 0.95). Replica #1 of each RNA-sequencing experiment was always compared to replica #1 of other RNA-sequencing experiments for analyses of global gene expression changes. The reported results were confirmed by analyzing the second (#2) replica experiments.

A comparison of biological replicas for several RNA-sequencing experiments showed that about 10% of the genes with 0 RPKMs in one replica had > 0 RPKMs in the other replica. Of these genes, > 95% had a RPKM value < 1 in the replica with > 0 RPKMs (see Supplementary

Figure S1A). Therefore we considered all genes with a RPKM value of > 1 as being expressed, and genes with a RPKM value < 1 as not being expressed. According to this definition of expressed genes, our depth of sequencing was sufficient to reach saturation for the number of expressed genes in all RNA-sequencing experiments.

De novo motif discovery

For de novo discovery of the Pax5-binding motif (Figure 3C), peaks were called for pro-B and mature B cell tracks using the SISSRs program (Jothi et al, 2008) with a p-value threshold of 0.05. Only SISSRs peaks overlapping with MACS peaks that are common between pro-B and mature B cells were used for de novo motif discovery. Peaks were ranked by the number of reads, and sequences of the top 500 peaks were retrieved from the ENSEMBL mouse 47 genome database. De novo motif discovery was done using the MEME program (Bailey et al, 2009) with the following parameters: -dna -revcomp -minw 6 -maxw 20 -nmotifs 6 -mod zoops -maxsites 500. The motif with the highest E-value (1.1e-431) was selected and further used to scan the entire peak dataset (pro-B and mature B unique and common peaks) using the MotifLocator program (Thijs et al, 2002) with default parameters and precompiled background file. Finally, the sequences of the Pax5 peaks were shuffled by maintaining the dinucleotide composition (Emboss ShuffleSeq) to identify the random occurrence of the motif. This de novo motif discovery procedure could only detect the Pax5 recognition sequence and an ETS factor-binding motif at the Pax5 peaks.

Motif analysis of Pax5-binding regions

Motifs enriched in pro-B or mature B cells (Supplementary Figure S10C) were identified by analyzing the following Pax5 peak sets; unique Pax5 peaks of pro-B cell, unique Pax5 peaks of mature B cell and common Pax5 peaks (Figure 3B) as well as Pax5 peaks associated with activated Pax5 target genes in pro-B cells (Figure 4D) or mature B cells (Figure 6C). The peaks were sorted by MACS tag counts, and a nucleotide sequence fragment ranging 100 bp upstream and downstream of the peak center (200 bp in total) was extracted for each peak from the repeat-masked version of the NCBI37 (mm9) genome using the MEME Suite program (Bailey et al, 2009). Sequence motif databases were downloaded from JASPAR (Bryne et al, 2008), TRANSFAC (version 2010.2; Matys et al, 2006) and Chen et al (2008) or were taken from our own ChIP-seq analyses (unpublished data). Motifs were searched in the different Pax5 peak sets using the MotifLocator program (Thijs et al, 2002) with default values and precompiled mouse background frequencies downloaded from the MotifLocator web site (http://homes.esat.kuleuven.be/~bioi_marchal/MotifSuite/motiflocator.php). In a post-processing step, the number and percentage of peak sequences that contained at least one motif was calculated for each motif in the different Pax5 peak sets. The motifs were sorted according to the difference of their percentages in the pro-B and mature B cell unique peak sets, whereby motifs found in more than 50% of any peak set (unique and common) were discarded. Motifs of transcription factors that are not expressed in B lymphocytes according to the ImmGen database (Heng & Painter, 2008) were also eliminated.

Supplementary Tables and Figures

Table S1. Description of all Solexa sequencing experiments generated for this study.

Table S2. Genomic coordinates (mm9) of all promoters and distal elements identified in pro-B cells. The data of this table correspond to Figure 1A. The bed file can be uploaded to any genome browser to visualize the identified promoters and distal elements.

Table S3. Genomic coordinates (mm9) of all promoters and distal elements identified in mature B cells. The data of this table correspond to Figure 1B. The bed file can be uploaded to any genome browser to visualize the identified promoters and distal elements.

Table S4. Genomic coordinates (mm9) of the different distal elements identified in pro-B cells. The data of this table correspond to Figure 2A-C. The bed file can be uploaded to any genome browser to visualize the distal elements of the five different classes.

Table S5. Genomic coordinates (mm9) of the different distal elements identified in mature B cells. The data of this table correspond to Figure S3B-C. The bed file can be uploaded to any genome browser to visualize the distal elements of the five different classes.

Table S6. Genomic coordinates (mm9) of the Pax5 peaks identified in pro-B and mature B cells. The data of this table correspond to Figure 3.

Table S7. RNA-seq data of regulated Pax5 target genes identified in pro-B and mature B cells. The normalized expression of each gene is shown as RPKM value. The genes belonging to the three different significance classes are indicated by different colors: red (significance > 82.9), green (47.0 – 82.9) and white (20.4 – 47.0). The data of this table correspond to Figures 4D and 6B.

Table S8. RNA-seq data of genes regulated during the developmental transitions analyzed. The normalized expression of each gene is shown as RPKM value. The genes belonging to the three different significance classes are indicated by different colors: red (significance > 82.9), green (47.0 – 82.9) and white (20.4 – 47.0). The data of this table correspond to Figures 5A,B and S9A,B.

Table S9. Oligonucleotides used for ChIP-qPCR analysis (Figure S6C).

Supplementary Table S1: Description of all Solexa sequencing experiments generated for this study (GSE38046)

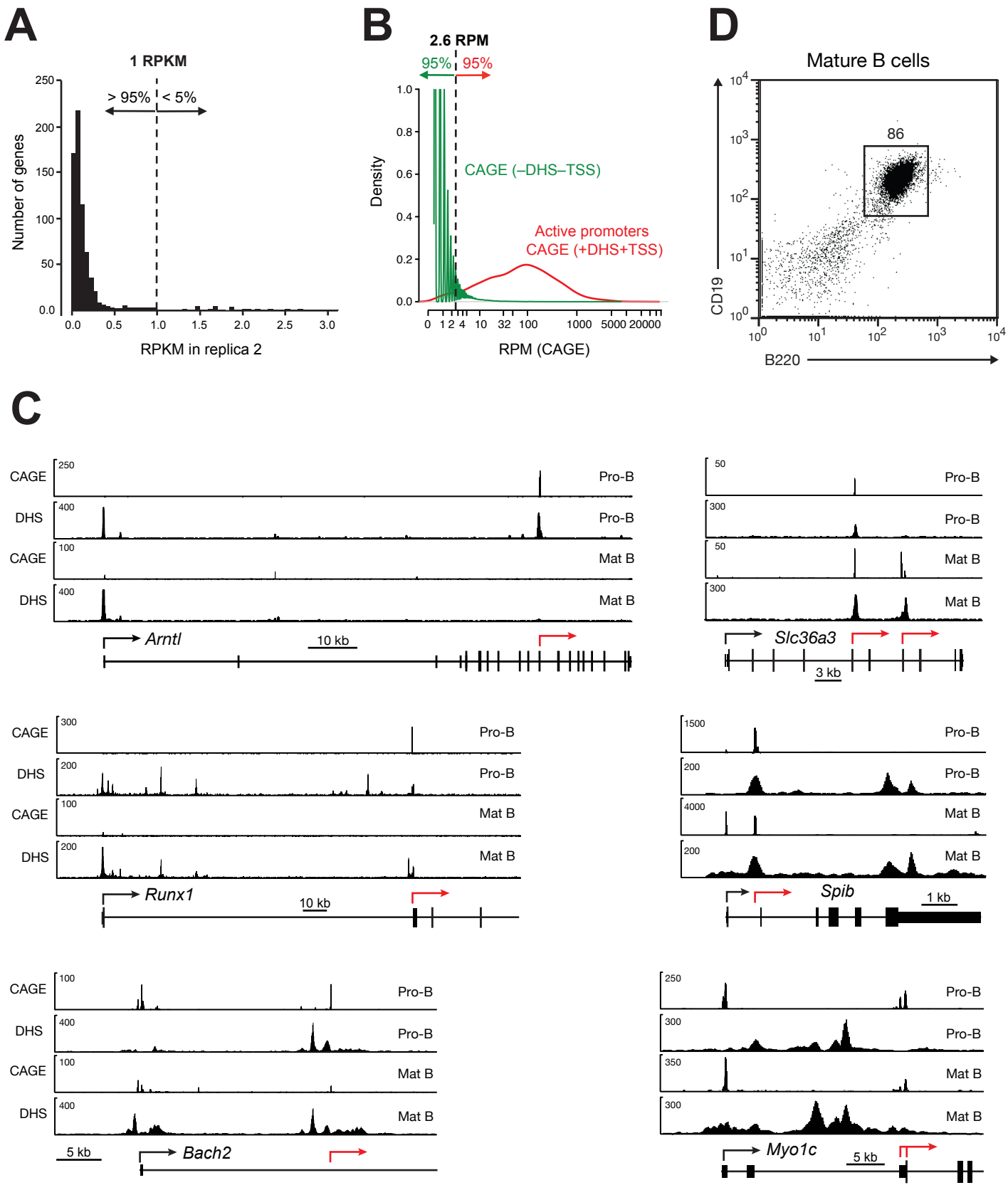
Exp. #	Sample name	Experiment	Genotype	Cell type	Isolation method	Replica	Solexa protocol	Mapped reads (x10^6)	MACS peak #
8485	RNA-seq ALP	RNA-seq	WT	ALP	Ex vivo	1	36 bp - single read	11.1	NA
8558	RNA-seq ALP	RNA-seq	WT	ALP	Ex vivo	2	36 bp - single read	7.8	NA
8486	RNA-seq BLP	RNA-seq	WT	BLP	Ex vivo	1	36 bp - single read	10.9	NA
8559	RNA-seq BLP	RNA-seq	WT	BLP	Ex vivo	2	36 bp - single read	6.7	NA
8817	RNA-seq pro-B	RNA-seq	Rag2(−/−)	Pro-B	Ex vivo	1	36 bp - single read	13.4	NA
8818	RNA-seq pro-B	RNA-seq	Rag2(−/−)	Pro-B	Ex vivo	2	36 bp - single read	10.9	NA
8819	RNA-seq pro-B	RNA-seq	WT	Pro-B	Ex vivo	1	36 bp - single read	6.9	NA
8276	RNA-seq mature B	RNA-seq	WT	Mature B	Ex vivo	1	36 bp - single read	20.9	NA
8159	RNA-seq mature B	RNA-seq	WT	Mature B	Ex vivo	2	36 bp - single read	12.5	NA
8275	RNA-seq pro-B	RNA-seq	Rag2(−/−)	Pro-B	4-5 day culture	1	36 bp - single read	24.5	NA
8403	RNA-seq pro-B	RNA-seq	WT = Pax5(fl/fl)	Pro-B	4-5 day culture	1	36 bp - single read	6.5	NA
8404	RNA-seq pro-B	RNA-seq	WT = Pax5(fl/fl)	Pro-B	4-5 day culture	2	36 bp - single read	6.7	NA
8402	RNA-seq Pax5-deleted pro-B	RNA-seq	Pax5(Δ/Δ) = Vav-Cre Pax5(fl/fl)	Pro-B	4-5 day culture	1	36 bp - single read	9.2	NA
8405	RNA-seq Pax5-deleted pro-B	RNA-seq	Pax5(Δ/Δ) = Vav-Cre Pax5(fl/fl)	Pro-B	4-5 day culture	2	36 bp - single read	6.8	NA
8743	RNA-seq Pax5-deleted mature B	RNA-seq	Pax5(Δ/Δ) = Cd23-Cre Pax5(fl/fl)	Mature B	Ex vivo	1	36 bp - single read	11.2	NA
8742	RNA-seq Pax5-deleted mature B	RNA-seq	Pax5(Δ/−) = Cd23-Cre Pax5(fl/−)	Mature B	Ex vivo	2	36 bp - single read	16.2	NA
8093	Pax5 Bio-ChIP-seq pro-B	Bio-ChIP-seq	Pax5(Bio/Bio) Rag2(−/−)	Pro-B	4-5 day culture	1	36 bp - single read	36.9	20,725
8417	Pax5 Bio-ChIP-seq pro-B	Bio-ChIP-seq	Pax5(Bio/Bio) Rag2(−/−)	Pro-B	4-5 day culture	2	36 bp - single read	20.0	24,585
8097	Pax5 Bio-ChIP-seq mature B	Bio-ChIP-seq	Pax5(Bio/Bio)	Mature B	Ex vivo	1	36 bp - single read	20.0	15,468
8090	Pax5 Bio-ChIP-seq mature B	Bio-ChIP-seq	Pax5(Bio/Bio)	Mature B	Ex vivo	2	36 bp - single read	16.5	11,760
8095	Control Bio-ChIP-seq mature B	Bio-ChIP-seq	Rosa26(BirA/BirA)	Mature B	Ex vivo	1	36 bp - single read	2.1	46
8666	H3K4me1 ChIP-seq pro-B	ChIP-seq	Rag2(−/−)	Pro-B	4-5 day culture	1	36 bp - single read	28.1	37,037
8114	H3K4me2 ChIP-seq pro-B	ChIP-seq	Rag2(−/−)	Pro-B	4-5 day culture	1	36 bp - single read	10.4	31,376
8109	H3K4me2 ChIP-seq pro-B	ChIP-seq	Rag2(−/−)	Pro-B	4-5 day culture	2	36 bp - single read	9.5	31,434
8115	H3K4me3 ChIP-seq pro-B	ChIP-seq	Rag2(−/−)	Pro-B	4-5 day culture	1	36 bp - single read	9.9	18,613
8110	H3K4me3 ChIP-seq pro-B	ChIP-seq	Rag2(−/−)	Pro-B	4-5 day culture	2	36 bp - single read	8.3	19,126
8113	H3K9ac ChIP-seq pro-B	ChIP-seq	Rag2(−/−)	Pro-B	4-5 day culture	1	36 bp - single read	8.9	18,806
8108	H3K9ac ChIP-seq pro-B	ChIP-seq	Rag2(−/−)	Pro-B	4-5 day culture	2	36 bp - single read	8.1	17,734
8111	H3K27me3 ChIP-seq pro-B	ChIP-seq	Rag2(−/−)	Pro-B	4-5 day culture	1	36 bp - single read	15.3	15,939
8116	H3K27me3 ChIP-seq pro-B	ChIP-seq	Rag2(−/−)	Pro-B	4-5 day culture	2	36 bp - single read	9.0	13,184
8069	H3K4me2 ChIP-seq mature B	ChIP-seq	WT	Mature B	Ex vivo	1	36 bp - single read	7.4	24,927
8088	H3K4me3 ChIP-seq mature B	ChIP-seq	WT	Mature B	Ex vivo	1	36 bp - single read	9.9	15,405
8070	H3K4me3 ChIP-seq mature B	ChIP-seq	WT	Mature B	Ex vivo	2	36 bp - single read	5.8	16,979
8059	H3K9ac ChIP-seq mature B	ChIP-seq	WT	Mature B	Ex vivo	1	36 bp - single read	3.9	14,461
8086	H3K9ac ChIP-seq mature B	ChIP-seq	WT	Mature B	Ex vivo	2	36 bp - single read	4.6	12,825
8080	H3K27me3 ChIP-seq mature B	ChIP-seq	WT	Mature B	Ex vivo	1	32 bp*- single read	9.5	17,608
8300	CAGE-seq pro-B	CAGE-seq	Rag2(−/−)	Pro-B	4-5 day culture	1	27 bp - single read	5.8	NA
8300	CAGE-seq mature B	CAGE-seq	WT	Mature B	Ex vivo	1	27 bp - single read	6.7	NA
8439	DHS-seq pro-B	DHS-seq	Rag2(−/−)	Pro-B	4-5 day culture	1	76 bp - paired end	27.3	38,462
8335	DHS-seq pro-B	DHS-seq	Rag2(−/−)	Pro-B	4-5 day culture	2	36 bp - single read	15.6	14,405
8953	DHS-seq Pax5-deleted pro-B	DHS-seq	Pax5(−/−) Rag2(−/−)	Pro-B	3-4 week culture	1	76 bp - paired end	76.0	38,576
8438	DHS-seq Pax5-deleted pro-B	DHS-seq	Pax5(−/−)	Pro-B	3-4 week culture	2	76 bp - paired end	32.2	16,916
8440	DHS-seq mature B	DHS-seq	Tcrb Eβ(−/−)	Mature B	Ex vivo	1	76 bp - paired end	28.7	32,688
8630	DHS-seq mature B	DHS-seq	Tcrb Eβ(−/−)	Mature B	Ex vivo	2	76 bp - paired end	57.1	40,805

*, Read length was originally 36 bp, but the last 4 bp were eliminated, because the last 4 sequencing cycles produce more errors than usual.
NA, not applicable Replica 1 – used for generating the data shown in the figures Replica 2 – used for confirmation of the data shown

Supplementary Table S9: Oligonucleotides used for ChIP-qPCR analysis

Gene	Forward	Reverse	Used in
Alpl (G)	ACGCCAGATAGCCATGAGTC	CTGCCCTGCTTCTACGACAT	Fig. S6C
Emb (U)	TCTCTCCGTGACACTGTGGT	TCAGGGCAGCCTGTAGAAAC	Fig. S6C
Emb (G)	CTCCTGTCTCAGCTTCCAAAG	TGCAAAACACCCATGACTTC	Fig. S6C
Itgb3 (P)	AGCTGCAGGTGTGGTATTGA	TGCTTTCCTCCTGGTTTCAT	Fig. S6C
Itgb3 (G)	GCACTAGGCCTGTGGTTAAGA	AAGGTGACTCGCTTGGTCCT	Fig. S6C
Cd19 (I)	CCCTCTTCTCATTCGTTTTCCA	CCAGGAAAGAATTTGAGAAAAATCA	Fig. S6C

G - Gene body
U - Upstream region
P - Promoter
I - Intergenic region



Supplementary Figure S1 (Revilla et al.)

Figure S1. Cutoff values for CAGE and RNA-seq analyses and identification of alternative promoters.

(**A**) Definition of a cutoff value of > 1 RPKM for the identification of expressed gene by RNA-sequencing. RPKM values were compared in two biological replicas of short-term cultured pro-B cells. Of 8306 genes that had 0 RPKMs in replica #1, 682 had a RPKM value > 0 in replica #2. The histogram shows the distribution of RPKMs for these 682 genes in replica #2. More than 95% of these genes had a RPKM value < 1. (**B**) Definition of the cutoff value of > 2.6 RPM for CAGE tag clusters of expressed genes. The cutoff for removing CAGE tag clusters with background sequence reads was defined by plotting the RPM values (x-axis) of two different classes of CAGE tag clusters together with their Gaussian kernel density estimation (y-axis). The CAGE tag clusters (139,253) shown in green do not overlap with any DHS peak or annotated TSS or TES. In contrast, the CAGE tag clusters (2,595) shown in red overlap with DHS peaks and annotated TSSs of expressed genes that were characterized by a normalized expression value of > 1 RPKM, as determined by RNA-sequencing (A). The cutoff was set at an RPM value of > 2.6, which retained 95% of the active promoters (red), but excluded 95% of all CAGE tag clusters without DHS sites (green), which were considered to be background. RPM - *R*eads *p*er CAGE tag cluster per *m*illion mapped sequence reads. (**C**) Identification of alternative promoters in pro-B cells and mature B cells by CAGE and DHS site analyses. Black arrows indicate RefSeq-annotated promoters and red arrows newly identified alternative promoters of the indicated genes. (**D**) Purity of mature B cells used for DHS site mapping. Mature B cells were directly isolated from lymph nodes of $E\beta^{-/-}$ mice lacking mature T cells (Bouvier et al, 1996) followed by enrichment of live cells by sucrose gradient purification. The purity was determined by FACS analysis prior to density gradient purification.

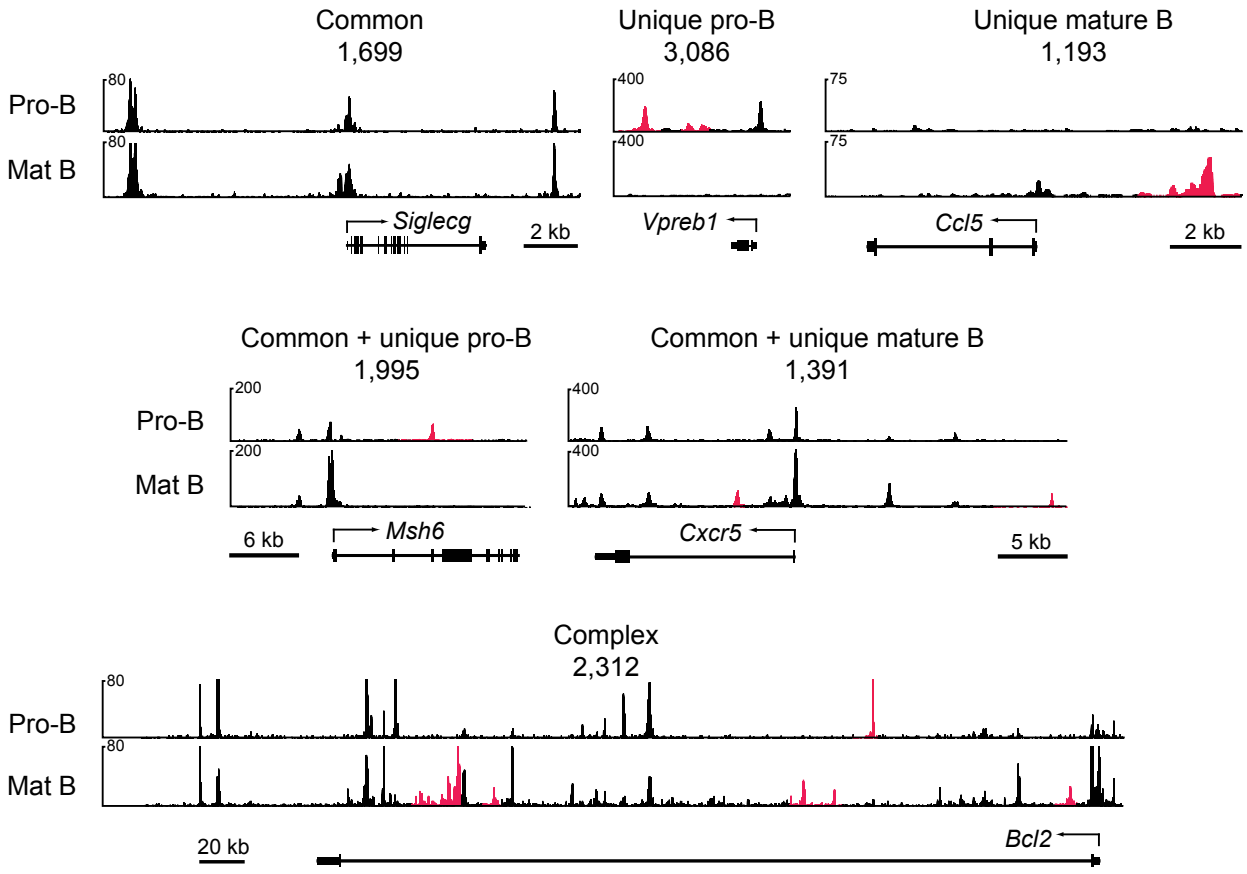
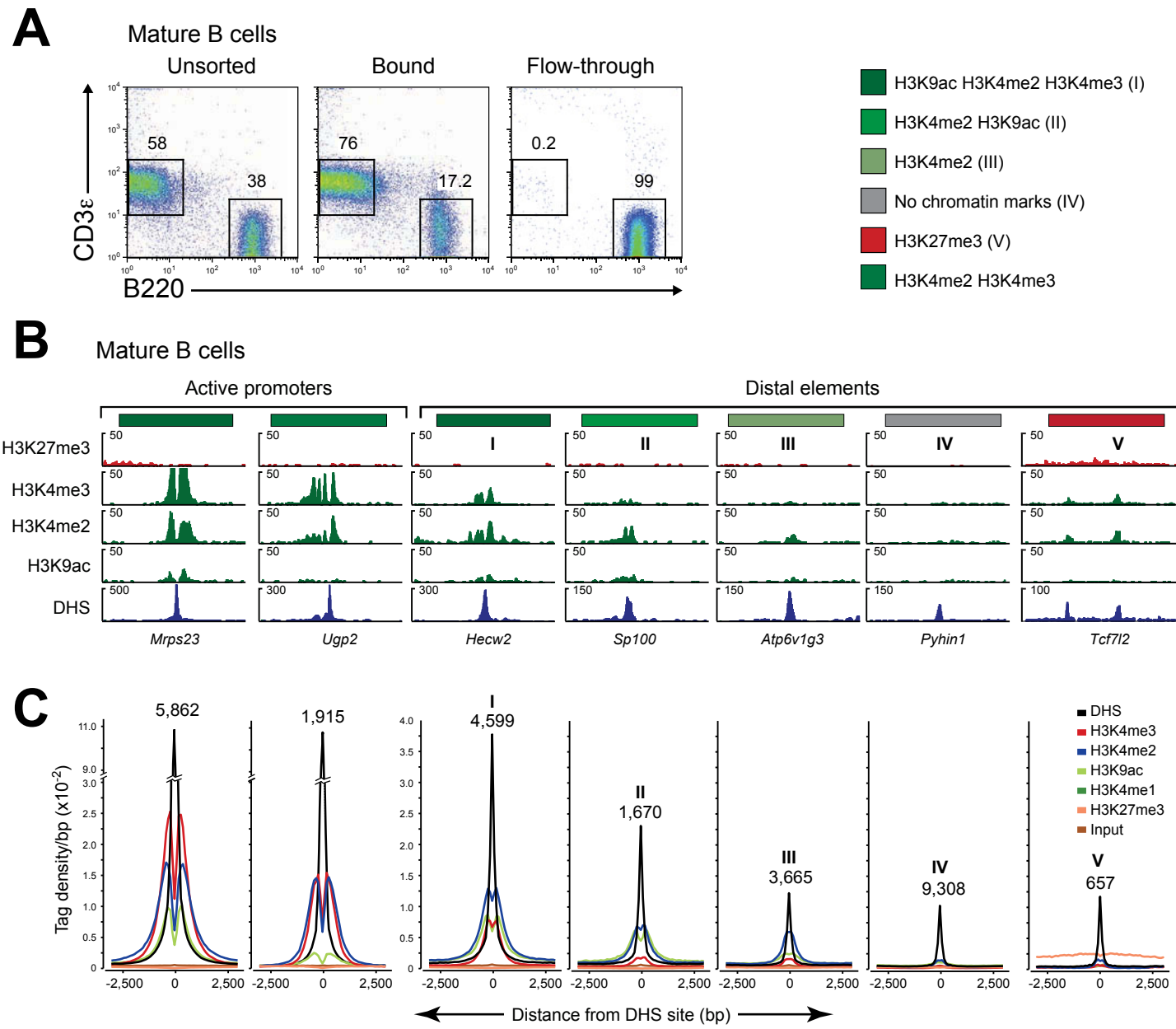
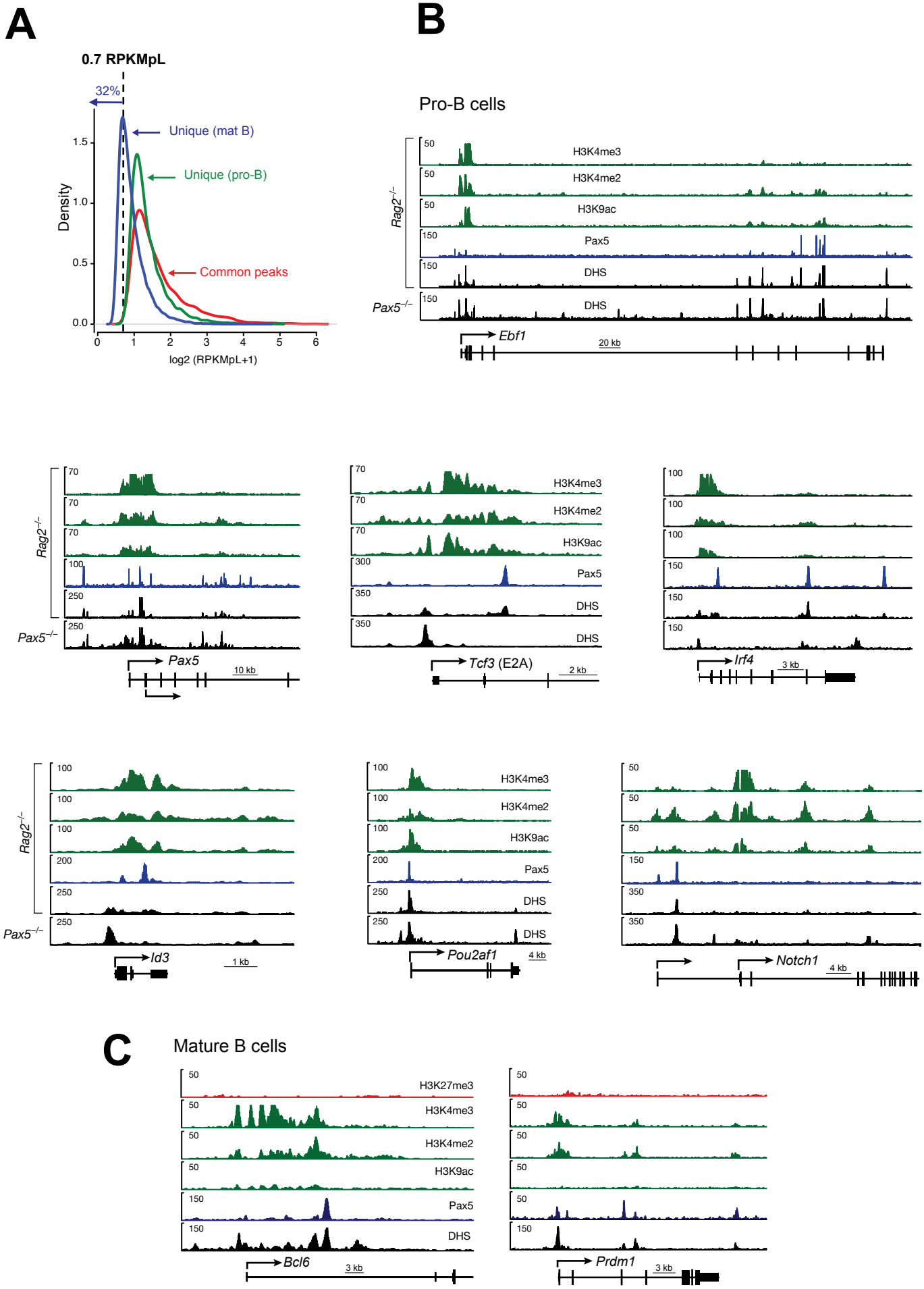

Common
1,699
Unique pro-B
3,086
Unique mature B
1,193
Pro-B
Mat B
80
80
400
400
75
75
Siglecg
2 kb
Vpreb1
Ccl5
2 kb
Common + unique pro-B
1,995
Common + unique mature B
1,391
Pro-B
Mat B
200
200
400
400
6 kb
Msh6
Cxcr5
5 kb
Complex
2,312
Pro-B
Mat B
80
80
20 kb
Bcl2

Figure S2. Genes with different assortments of distal elements in pro-B and mature B cells.
A representative gene is shown together with the total gene number present in each of the 6 different gene categories defined by the presence of common and unique distal elements in pro-B and mature B cells. Unique distal elements are indicated in red, and the scale bars are in kilobases (kb).



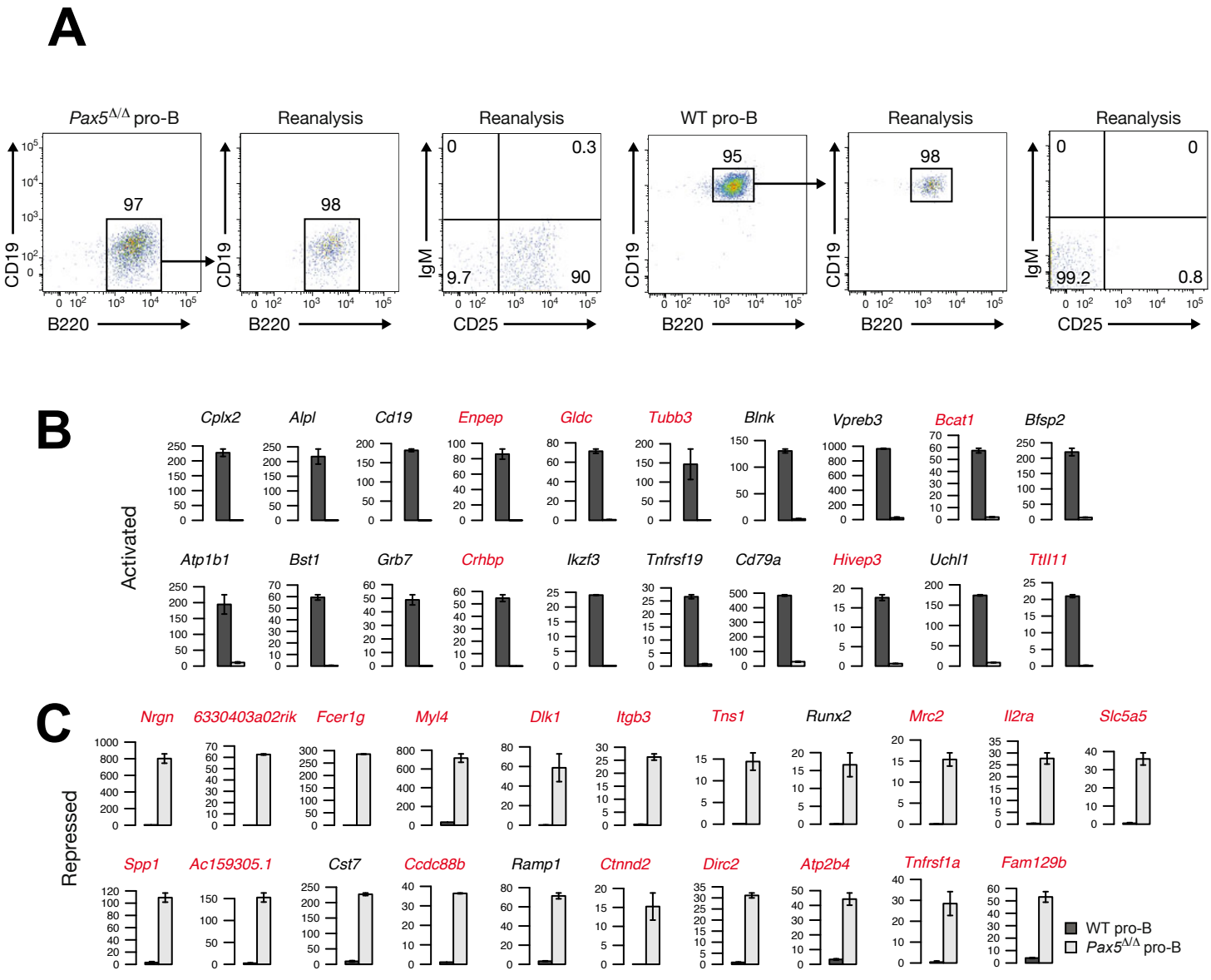
Supplementary Figure S3 (Revilla et al.)

Figure S3. Chromatin profiling of active enhancers or inactive distal elements in mature B cells. (**A**) Purification of mature B cells from lymph nodes of wild-type mice by MACS depletion of non-B cells with anti-PE beads after staining with PE-labeled TCRβ, CD4, CD8a, Mac1, Gr1, DX5 and Ly6C antibodies. (**B**) Chromatin marks at elements representative of different combinations of active (H3K4me2, H3K4me3, H3K9ac) and repressive (H3K27me3) histone modifications in mature B cells isolated from lymph nodes. The corresponding DHS sites are additionally shown. The names of the genes associated with the different elements are indicated. (**C**) Average sequence tag density profiles indicating the extent and intensity of the different histone modifications at regulatory elements in mature pro-B cells. The profiles were aligned at the center of the DHS sites, and the number of elements in each class is indicated. The genomic coordinates of the different distal elements are provided in Supplementary Table S5. (**D**) Changes of chromatin signatures at distal elements during the development of pro-B cells to mature B cells. The number of distal elements within each class (I to V) in pro-B cells was set to 100% and the presence (%) and chromatin signature (color code) of the distal elements at the same position in mature B cells are shown. (**E**) Interaction of CTCF and cohesin with distal elements in *Rag2*$^{-/-}$ pro-B cells. Genome-wide binding of the transcription factor CTCF and cohesin subunit Rad21 was recently determined by ChIP-sequencing (Ebert et al, 2011). The binding of CTCF and Rad21 in *Rag2*$^{-/-}$ pro-B cells is shown a relative percentage of all DHS sites within each class of distal elements. A similar distribution of CTCF-binding sites was observed at distal elements in mature B cells (data not shown).



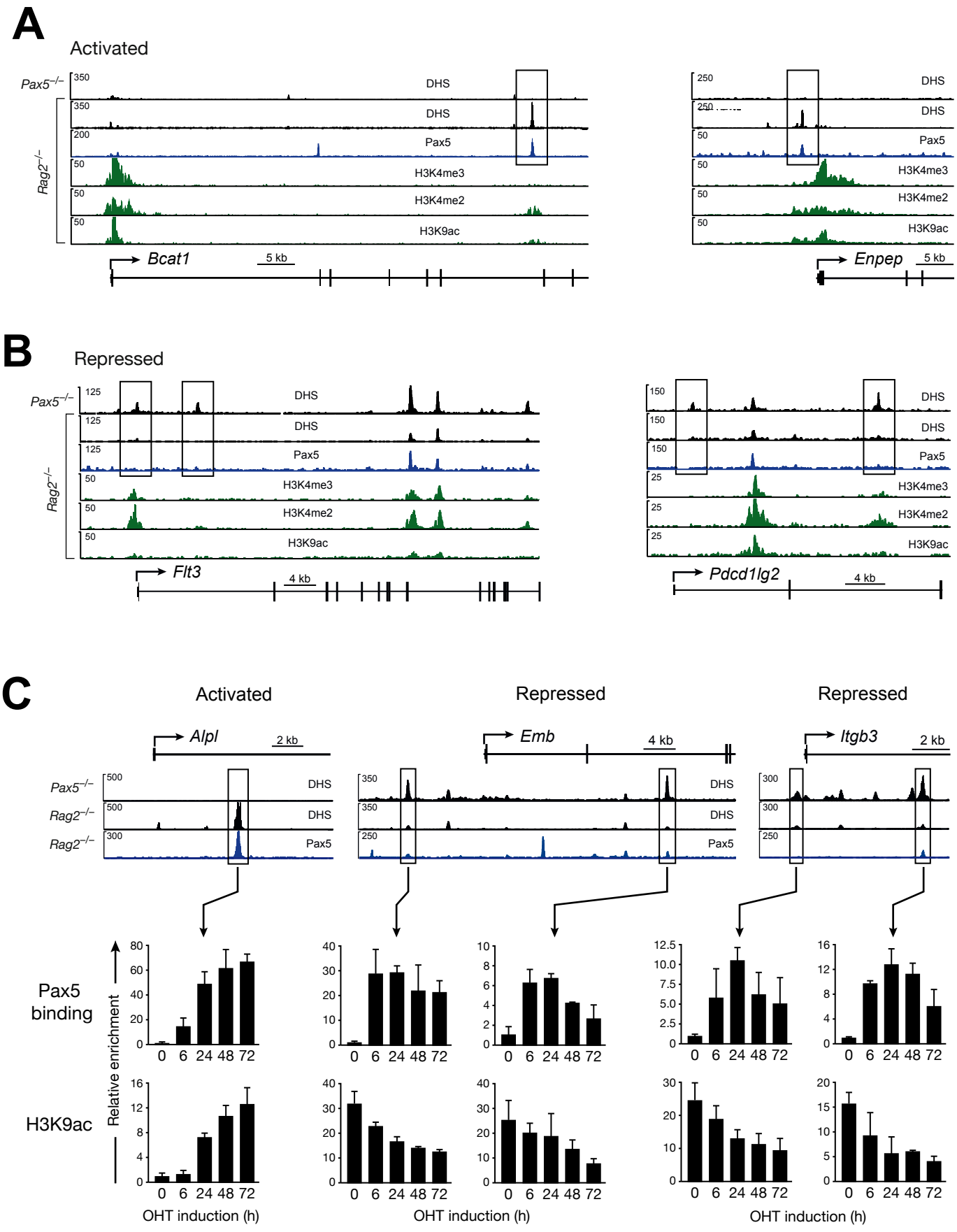
Supplementary Figure S4 (Revilla et al.)

Figure S4. Pax5-binding and epigenetic profile of important regulatory genes.
(**A**) Normalization of ChIP-seq experiments by eliminating small peaks. Pax5 peaks were identified by the MACS program with a p-value of $< 10^{-10}$, and their normalized peak height was calculated as RPKMpL (*r*eads *p*er 1 *k*b of peak width per *m*illion reads in peaks *p*er total *l*ength of all peaks; see Supplementary Materials and Methods). The RPKMpL values (x-axis) of common Pax5 peaks (blue) as well as unique peaks in pro-B (green) and mature (red) B cells were plotted together with their Gaussian kernel density estimation (y-axis). The cutoff was set at an RPKMpL value of > 0.7, which removed only 0.5% (112) of all Pax5 peaks (20,725) identified in pro-B cell. In contrast, the cutoff of 0.7 RPKMpL eliminated 7,226 (32%) very small and unique peaks of 22,694 Pax5 peaks in mature B cells. (**B**, **C**) Pax5 binding and epigenetic profile at key transcription factor genes. Pax5 binding (blue), active (geen) and inactive (red) histone modifications as well as DHS sites (black) are shown for selected genes in *Rag2*$^{-/-}$ pro-B cells (B) or wild-type mature B cells (C). The DHS sites in *Pax5*$^{-/-}$ pro-B cells are also shown.



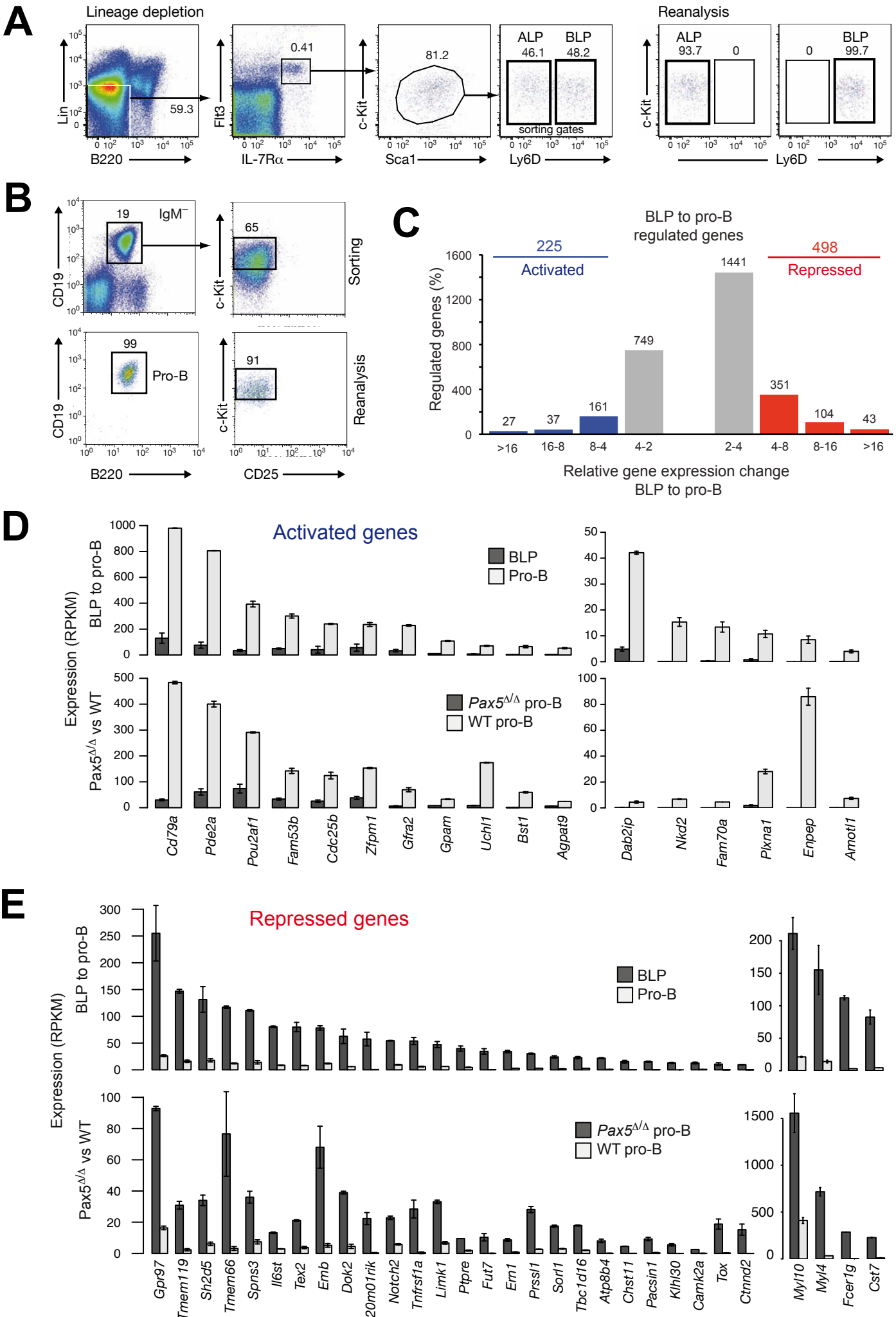
Supplementary Figure S5 (Revilla et al.)

Figure S5. Sorting of pro-B cells and expression of regulated Pax5 target genes.
(**A**) Purification of wild-type and *Pax5* mutant pro-B cells for RNA-sequencing. Short-term cultured $Pax5^{fl/fl}$ (WT) pro-B and *Vav*-Cre $Pax5^{fl/fl}$ ($Pax5^{\Delta/\Delta}$) pro-B cells were FACS-sorted as $CD19^{+}B220^{+}$ and $CD19^{-}B220^{+}$ cells, respectively. As shown by FACS reanalysis, the $Pax5^{\Delta/\Delta}$ pro-B cells did not express CD19 and up-regulated CD25 like $Pax5^{-/-}$ pro-B cells (Nutt et al, 1997), consistent with full deletion of the floxed *Pax5* allele. (**B**, **C**) Expression of the most highly (>16 fold) activated (B) and repressed (C) Pax5 target genes identified in Figure 4D. The expression of each gene in short-term cultured $Pax5^{\Delta/\Delta}$ (grey bar) and control $Rag2^{-/-}$ (black bar) pro-B cells is shown as average expression value (RPKM) with standard deviation, which was determined by two independent RNA-seq experiments of both cell types. Genes indicated in red refer to newly identified Pax5 target genes, whereas genes shown in black correspond to Pax5 target genes recently identified by ChIP-chip analysis (McManus et al, 2011).



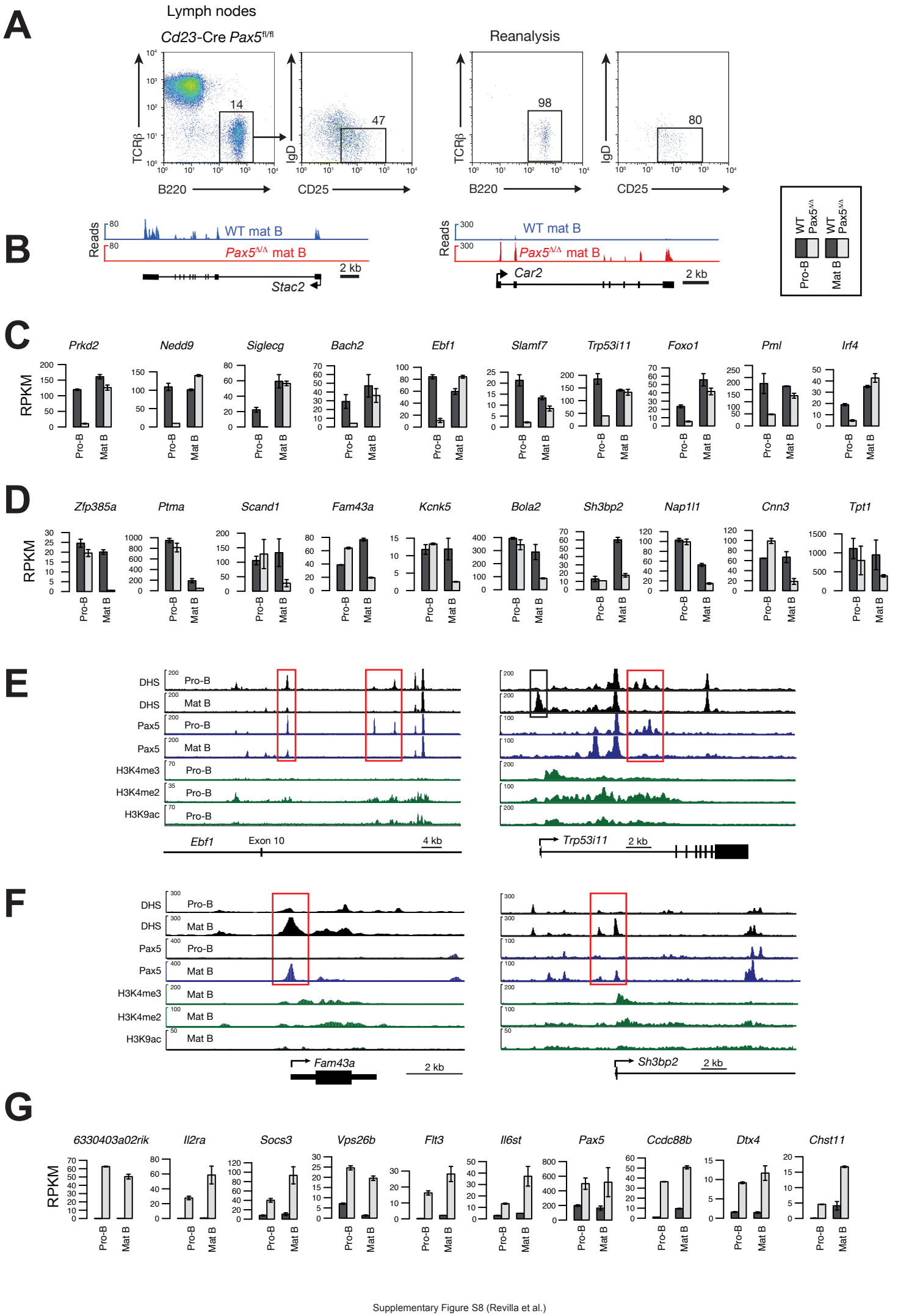
Supplementary Figure S6 (Revilla et al.)

Figure S6. Pax5-dependent gain and loss of enhancers at regulated target genes in pro-B cells. (**A**, **B**) Induction of enhancers at the activated Pax5 target genes *Bcat* and *Enpep* (A) and loss of DHS sites at the repressed Pax5 target genes *Flt3* and *Pdcd1lg2* (B) in $Rag2^{-/-}$ pro-B cells. DHS sites (black) are shown for $Rag2^{-/-}$ and $Pax5^{-/-}$ pro-B cells together with the active chromatin profile (green) in $Rag2^{-/-}$ pro-B cells and the Pax5-binding pattern (blue) analyzed in $Pax5^{Bio/Bio}$ $Rag2^{-/-}$ pro-B cells. Regions with differences in DHS sites are boxed. (**C**) Transient binding of Pax5 and loss of H3K9ac at the repressed target genes *Emb* and *Itgb3* in contrast to the activated target gene *Alpl*. The Pax5-binding regions analyzed by ChIP-qPCR assay are boxed in the upper part of panel C indicating Pax5-binding and DHS sites at the three genes in $Pax5^{-/-}$ and $Rag2^{-/-}$ pro-B cells (Figure 4E,G). $Pax5^{-/-}$ pro-B cells expressing the Pax5-oestrogen receptor (ER) fusion protein (KO-Pax5-ER pro-B cells; Nutt et al, 1998) were treated for the indicated time (h) with the oestrogen analogue 4-hydroxytamoxifen (OHT, 1 μM) prior to ChIP analysis with Pax5 or H3K9ac antibodies as described in detail (McManus et al, 2011). Input and precipitated DNA were quantified by real-time PCR with primer pairs amplifying the indicated Pax5-binding regions and an inactive region downstream of the *Cd19* gene (Table S9). The relative enrichment of H3K9ac or Pax5 binding was determined by dividing the percentage of precipitated DNA at the Pax5-binding site (ChIP/input) by the percentage of precipitated DNA at the downstream region of the *Cd19* gene (ChIP/input). The relative enrichment of Pax5 binding at time point 0 was set to 1. The average values and standard deviations of two independent experiments are shown.



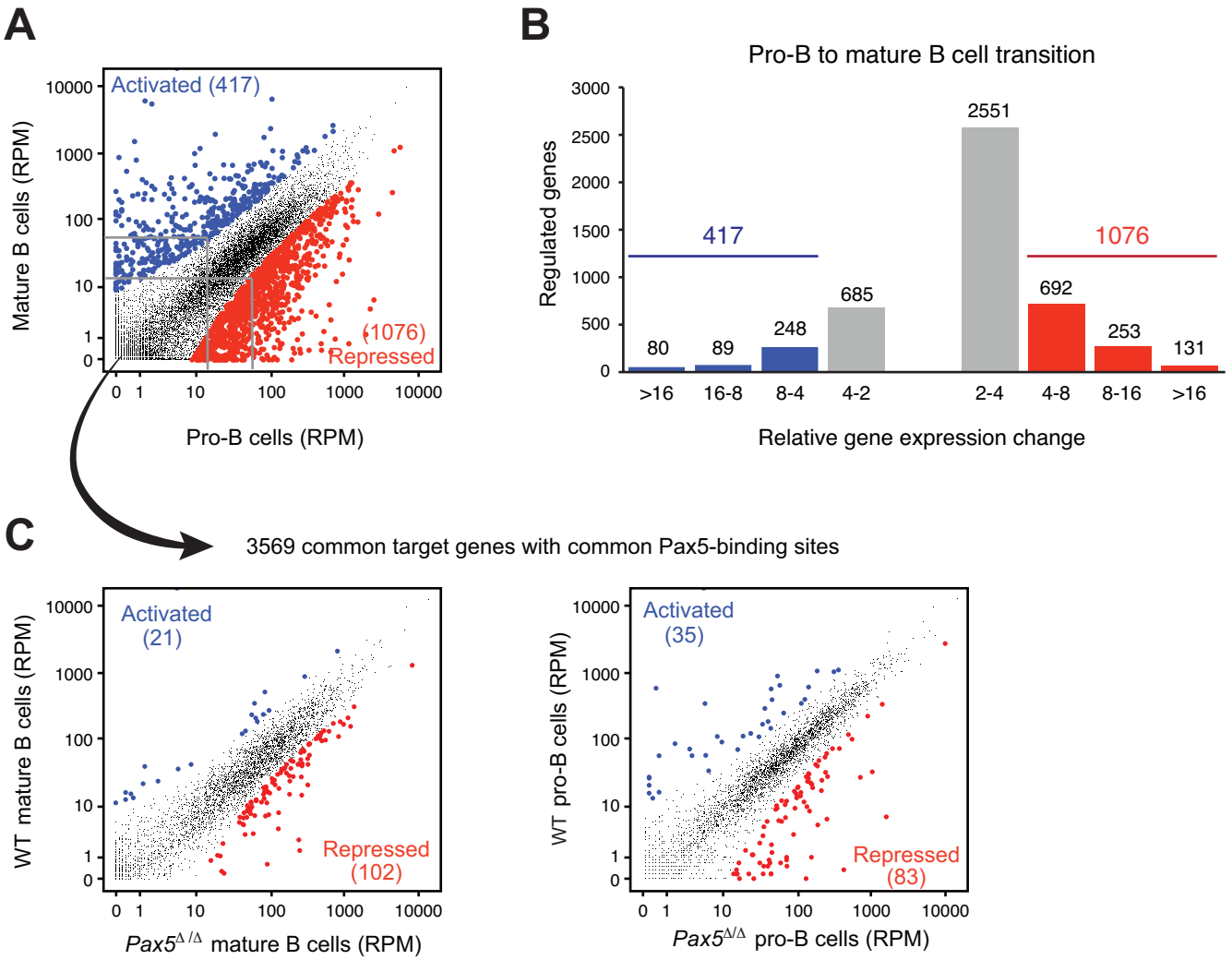
Supplementary Figure S7 (Revilla et al.)

Figure S7. Gene expression changes during the BLP to pro-B cell transition.
(**A**, **B**) FACS isolation of wild-type ALPs and BLPs (A) as well as *Rag2*$^{-/-}$ pro-B cells (B) using the indicated sorting gates. The purity of the sorted cell population was determined by flow cytometric reanalysis. (**C**) Activated and repressed genes during the developmental transition from BLPs to pro-B cells. The number of differentially expressed genes in *Rag2*$^{-/-}$ pro-B cells compared to BLPs are shown for the indicated expression changes referring to a gene with a transcript level of 15 RPM in the lower expressing cell type. The RNA-seq data of the genes regulated during this developmental transition are shown in Supplementary Table S8. (**D**) Expression of genes that are 8-16-fold activated during the BLP to pro-B cell transition and correspond to activated Pax5 target genes (see Figure 5D). Average expression values (RPKM) with standard deviations are shown for ex vivo sorted BLPs (black) and *Rag2*$^{-/-}$ (grey) pro-B cells (top row) and short-term cultured *Pax5*$^{\Delta/\Delta}$ (black) and wild-type (grey) pro-B cells (bottom row). (**E**) Expression of genes that are 8-16-fold repressed during the BLP to pro-B cell transition and qualify as repressed Pax5 target genes in pro-B cells (see Figure 5D).



Supplementary Figure S8 (Revilla et al.)

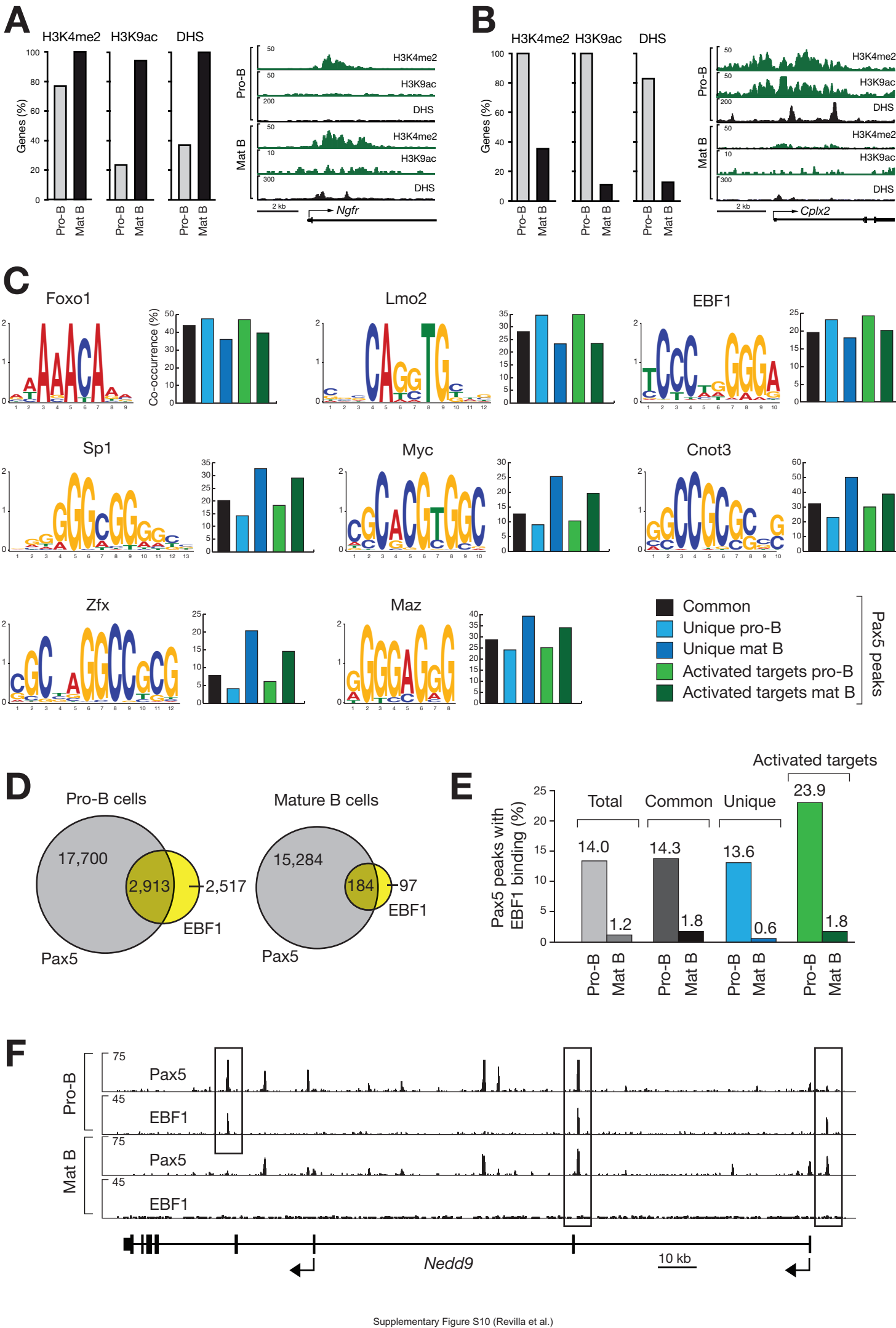
Figure S8. Regulation of Pax5 target genes in mature B cells.
(**A**) Isolation of Pax5-deficient mature B cells. *Pax5*$^{\Delta/\Delta}$ mature B cells were FACS-sorted as B220^{+}IgD^{lo}CD25^{+}TCRβ^{-} cells (Horcher et al, 2001) from the lymph nodes of *Cd23*-Cre *Pax5*$^{fl/fl}$ mice by using the indicated sorting gates. Wild-type mature B cells were purified from lymph nodes as shown in Supplementary Figure S3A. (**B**) Identification of *Stac2* (SH3 and cysteine-rich domain 2) and *Car2* (carbonic anhydrase 2) as Pax5-activated and Pax5-repressed genes in mature B cells by RNA-sequencing. (**C**, **D**) Target genes, which are activated by Pax5 only in pro-B (C) or mature B cells (D). The expression of the different genes is shown as normalized expression value (RPKM) with standard deviation as determined by two RNA-seq experiments. Black bars indicate the expression in wild-type (WT) pro-B and mature B cells and grey bars in *Pax5*$^{\Delta/\Delta}$ pro-B and mature B cells. (**E**, **F**) Loss of regulation by Pax5 coincides with the selective absence of DHS sites and Pax5 binding. The *Ebf1* and *Trp53i11* genes (E), which are no longer under Pax5 control in mature B cells (C), have lost DHS sites together with Pax5 binding (red boxes) at "pro-B cell-specific" enhancers in mature B cells. The *Fam43a* and *Sh3bp2* genes (F) lack DHS sites and Pax5 binding (red boxes) in pro-B cells where they are not regulated by Pax5 (D). Visual inspection of all 30 genes, that qualified as activated Pax5 target genes in pro-B cells in the absence of Pax5 regulation in mature B cells (C), and all 23 genes, that were activated Pax5 target genes in mature B cells in the absence of regulation in pro-B cells (D), demonstrated that 60% of the genes in both categories revealed the loss of a Pax5 peak together with a loss of the corresponding DHS site in the cell type where the respective gene was not regulated by Pax5, as shown in (E,F). The appearance of DHS sites without Pax5 binding in the "non-responsive" cell type (as shown by a black box for Trp53i11 in mature B cells; panel E) was a less frequent event. (**G**) Commonly repressed Pax5 target genes. Ten of 18 genes, which were repressed at least 3-fold in pro-B and mature B cells, are shown. The remaining 8 genes are *Mcoln2*, *Lax1*, *Impact*, *Gimap9*, *Leprotl1*, *Rapgef3*, *Ptpre* and *Ac159305.1*. These commonly repressed genes correspond to 7.4% and 9.7% of all repressed Pax5 target genes in pro-B and mature B cells, respectively.



E

GO catergory (Biological processes)	Genes (Total genes)	$Log_{10}(p)$
Cellular metabolic process	1644 (6999)	-50.4
Cellular macromolecule metabolic process	1262 (5179)	-42.4
Metabolic process	1840 (8281)	-40.7
Primary metabolic process	1614 (7058)	-40.5
Macromolecule metabolic process	1339 (5825)	-31.5
Nucleobase nucleoside nucleotide and nucleic acid metabolic process	935 (3842)	-28.4
Cellular protein metabolic process	681 (2649)	-26.3
Cellular nitrogen compound metabolic process	987 (4166)	-25.7
Nitrogen compound metabolic process	1000 (4256)	-24.7
Nucleic acid metabolic process	765 (3136)	-22.6
Gene expression	788 (3279)	-21.4
RNA metabolic process	683 (2771)	-21.0
Macromolecule modification	540 (2100)	-20.3

Figure S9. Massive gene expression changes during the pro-B to mature B cell transition.
(**A**) Gene expression changes during the developmental transitions of pro-B cells to mature B cells. Normalized expression values (RPM), which were determined by RNA sequencing of sorted *Rag2*$^{-/-}$ pro-B cells and mature B cells, were analyzed by the edgeR program (Robinson et al, 2010). Genes with a difference corresponding to a significance value of > 20 (blue) or < –20 (red) are highlighted in the scatter plot displaying the expression pattern at the two B cell types. The cutoff value of 20 corresponds to a 4-fold expression change for a gene expressed at 15 RPM in the lower expressing cell type (grey lines). (**B**) Activated and repressed genes during the development of pro-B cells to mature B cells. The number of differentially expressed genes in mature B cells compared to *Rag2*$^{-/-}$ pro-B cells are shown for the indicated expression changes referring to a transcript level of 15 RPM in the lower expressing cell type. The RNA-seq data of the genes regulated during the transition from pro-B to mature B cells are shown in Supplementary Table S8. (**C**, **D**) Pax5-dependent regulation of common Pax5 target genes, which were defined by their similar expression (black color in [A]) and the presence of common Pax5-binding sites (D) in pro-B and mature B cells. (**E**) Gene Ontology analysis of the common Pax5 target genes. The 13 most significant classes are shown together with the number of common Pax5 target genes and total genes (in brackets) belonging to each class, as defined by the analysis of biological processes using the GoMiner program (Zeeberg et al, 2003).



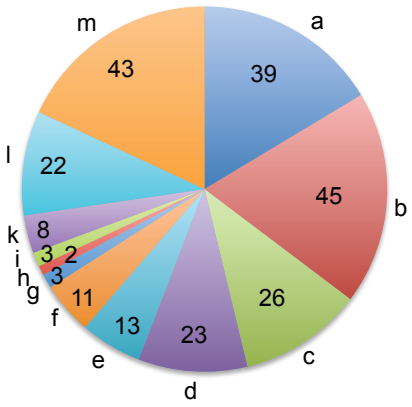
Supplementary Figure S10 (Revilla et al.)

Figure S10. EBF1 binding and enrichment of transcription factor-binding motifs at Pax5 peaks in pro-B and mature B cells.

(**A**, **B**) Epigenetic silencing of Pax5 target genes in the B cell type where they are not or only lowly expressed. The presence or absence of DHS sites and active histone marks (H3K4me2 and H3K9ac) was evaluated by visual inspection of the genes shown in Figure 6F and 6G. (A) Active histone marks and DHS sites were present at activated Pax5 target genes in mature B cells (Figure 6F), but H3K9ac and DHS sites were absent or strongly reduced at most of these genes in pro-B cells, whereas H3K4me2 (indicative of poised chromatin) was present at 80% of these genes as shown for *Ngfr*. (B) As exemplified for *Cplx2*, active histone marks and DHS sites were present at activated Pax5 target genes in pro-B cells, but were absent or strongly reduced at most of these genes in mature B cells, where they are no longer or lowly expressed (Figure 6G). (**C**) Enrichment of transcription factor-binding motifs at Pax5 peaks. Sequences from -100 bp to +100 bp relative to the center of the Pax5 peak were analyzed for the presence of transcription factor-binding motifs that are enriched at common and unique Pax5 peaks (Figure 3B) or at Pax5 peaks of activated Pax5 target genes in pro-B cells (Figure 4D) or mature (mat) B cells (Figure 6C), as described in Supplementary Materials and Methods. The motif with its information content (bits) is shown together with the frequency of its co-occurrence relative to all Pax5 peaks of the respective category. The motifs were taken from the TRANSFAC database (Foxo1, Lmo1, Maz, Sp1, Cnot3), Chen et al (2008: Myc, Zfx), and Vilagos et al (2012; EBF1). (**D**) Overlap of Pax5 peaks (this study) with EBF1 peaks (Vilagos et al, 2012) in pro-B and mature B cells, as determined by ChIP-sequencing. Pax5 and EBF1 peaks were equally called with a stringent p-value of $< 10^{-10}$. (**E**) Frequency of co-occurrence of EBF1 and Pax5 peaks relative to all common and unique Pax5 peaks as well as to Pax5 peaks at activated target genes in pro-B and mature B cells. (**F**) Absence of EBF1 binding at the *Nedd9* locus in mature B cells, where this gene is no longer regulated by Pax5 (Supplementary Figure S8C).

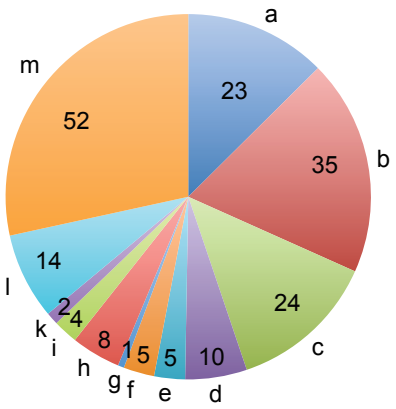
A

Repressed targets
Pro-B cells



B

Repressed targets
Mature B cells



a: receptor
b: transducer
c: transcription
d: metabolism
e: cytoskeleton
f: transporter
g: cell cycle
h: turnover
i: trafficking
k: secreted
l: other
m: unknown

Signal
Transducer

Tns1
Ccdc88b
Fgr (K)
Dusp5 (P)
Grap2 (A)
Gng12
Chn2
Frmd4b (A)
Osbpl3 (A)
Trio
Tbc1d16
Ppm1h (P)
Cass4 (A)
Lax1 (A)
Sik1 (K)
Rab27a
Ern1
Ahnak
Rinl (A)
Socs3
Sla2 (A)
Ppp1r3b (P)
Tyrobp (A)
Dok2 (A)
Limk1 (K)
Pik3r1 (K)
Rapgef3
Sh2d2a (A)
Daam1
Adap1
Camk2a (K)
Ptpre (P)
Dtx4
Arhgap8
Tnfaip8l2
Anxa5
Rasgrf2
Prkca (K)
Spry1
Pik3r6 (K)
Zap70 (K)
Irs1 (A)
Dock9
Pde4a
Rgl1

Surface
Receptor

Fcer1g
Dlk1
Itgb3
Mrc2
Il2ra
Ramp1
Tnfrsf1a
S1pr3
Cd28
Flt3
Nfam1
Lrp1
Tmem66
Pdcd1
Emb
Il18rap
Itgb7
Sema4a
Gpr97
Notch1
Emp1
Tspan2
Sema6d
A430110n23rik
Mrc1
Cxcr5
Cd9
Cdh1
Notch2
Csf1r
Sv2a
Il6st
Gpr146
Il4ra
Leprotl1
Gfra1
Il1r2
Emp3
Thy1

Transcription
Regulator

Nr4a1
Runx2
Fos
Tox
Hnf1b
Mn1
Gfi1b
Nfe2
Creb5
Ctbp2
Id2
Eya2
Lsr
Zfhx3
Tcf7
Ets2
Med13l
Maml3
Mxd4
Bcl9l
Bhlhe40
Pttg1ip
Ell3
Elk3
Mta3
Pax5

Signal
Transducer

Socs3
Gbp2 (A)
Ahnak
Irgm2
Rgs2
Trp53inp1
Arhgap18
Map3k5 (K)
Ccdc88b
Lax1 (A)
Dtx4
Tgtp1
Hck (K)
Pde4b
Abr
Nod2
Ptprj (P)
Cdc42se2
Cd300lf
Rhobtb1
Rgs10
Rock2 (K)
Tns3
Gimap9
Akap9
Pde7a
Ptprs (P)
Magi3
Akt3 (K)
Peli1
Itpr1
Dyrk2 (K)
Dock10
Braf
Irgm1

Surface
Receptor

Il2ra
Cd93
Lair1
Flt3
Il6st
P2ry13
Tlr7
Itga4
Cd274
Hrh2
Ly6a
Gpr183
Cd84
Crim1
Insr
Thbd
Gpr34
Plxnc1
Antxr2
Pdcd1lg2
Il6ra
Tspan13
Lphn2

Transcription
Regulator

Trps1
Elk4
Pax5
Mll1
Zbtb20
Nrip1
Sfmbt1
Stat4
Maml2
Mll3
Prdm1
Crebbp
Tead2
Stat1
Cecr2
Tbl1xr1
Ankrd12
Tcf20
Irf1
Mll2
Med13
Ssbp2
Rb1
Tef

Figure S11. Function of repressed Pax5 target genes in pro-B and mature B cells.
(**A**, **B**) Pie diagram indicating the different functional classes of repressed Pax5 target genes in pro-B cells (A) and mature B cells (B). The repressed Pax5 target genes shown below are ranked according to the significance of their differential expression as determined by RNA-sequencing of *Rag2*$^{-/-}$ and Pax5-deficient pro-B cells (Figure 4D) or wild-type and Pax5-deficient mature B cells (Figure 6C). Genes with a significance value of > 83 (> 16 fold) or between 83 and 47 (16-8 fold) are indicated by red and green boxes, respectively, whereas the remaining genes are characterized by a significance value ranging from 47 to 20 (8-4 fold). Genes that are commonly repressed in both cell types are indicated in bold. Signaling adaptors (A), kinases (K) and phosphatases (P) are indicated. The different functional categories shown in the pie diagram include: (a) cell surface receptors and transmembrane proteins; (b) signal transducers; (c) transcriptional regulators; (d) enzymes involved in metabolism; (e) cytoskeletal proteins and cytoskeleton-associated proteins; (f) transporters and channels; (g) cell cycle regulators; (h) regulators of protein turnover and processing; (i) regulators of protein trafficking; (k) secreted protein; (l) proteins with other functions; (m) genes with unknown functions.

Supplementary References

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37:** W202-W208

Bouvier G, Watrin F, Naspetti M, Verthuy C, Naquet P, Ferrier P (1996) Deletion of the mouse T-cell receptor β gene enhancer blocks αβ T-cell development. *Proc Natl Acad Sci USA* **93:** 7877-7881

Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A (2008) JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res* **36:** D102-D106

Cappabianca L, Thomassin H, Pictet R, Grange T (1999) Genomic footprinting using nucleases. *Methods Mol Biol* **119:** 427-442

Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Arner P, Chesi A, Gustincich S, Persichetti F, Suzuki H, Grimmond SM, Wells CA, Orlando V, Wahlestedt C, Liu ET, Harbers M, Kawai J, Bajic VB, Hume DA, Hayashizaki Y (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38:** 626-635

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, Loh YH, Yeo HC, Yeo ZX, Narang V, Govindarajan KR, Leong B, Shahab A, Ruan Y, Bourque G, Sung WK, Clarke ND, Wei CL, Ng HH (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133:** 1106-1117

de Boer J, Williams A, Skavdis G, Harker N, Coles M, Tolaini M, Norton T, Williams K, Roderick K, Potocnik AJ, Kioussis D (2003) Transgenic mice with hematopoietic and lymphoid specific expression of Cre. *Eur J Immunol* **33:** 314-325

Decker T, Pasca di Magliano M, McManus S, Sun Q, Bonifer C, Tagoh H, Busslinger M (2009) Stepwise activation of enhancer and promoter regions of the B cell commitment gene *Pax5* in early lymphopoiesis. *Immunity* **30:** 508-520

Driegen S, Ferreira R, van Zon A, Strouboulis J, Jaegle M, Grosveld F, Philipsen S, Meijer D (2005) A generic tool for biotinylation of tagged proteins in transgenic mice. *Transgenic Res* **14:** 477-482

Ebert A, McManus S, Tagoh H, Medvedovic J, Salvagiotto G, Novatchkova M, Tamir I, Sommer A, Jaritz M, Busslinger M (2011) The distal V_H gene cluster of the *Igh* locus contains distinct regulatory elements with Pax5 transcription factor-dependent activity in pro-B cells. *Immunity* **34:** 175-187

Heng TS, Painter MW (2008) The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol* **9:** 1091-1094

Horcher M, Souabni A, Busslinger M (2001) Pax5/BSAP maintains the identity of B cells in late B lymphopoiesis. *Immunity* **14:** 779-790

Jothi R, Cuddapah S, Barski A, Cui K, Zhao K (2008) Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res* **36:** 5221-5231

Kwon K, Hutter C, Sun Q, Bilic I, Cobaleda C, Malin S, Busslinger M (2008) Instructive role of the transcription factor E2A in early B lymphopoiesis and germinal center B cell development. *Immunity* **28:** 751-762

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25

Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* **34:** D108-D110

McManus S, Ebert A, Salvagiotto G, Medvedovic J, Sun Q, Tamir I, Jaritz M, Tagoh H, Busslinger M (2011) The transcription factor Pax5 regulates its target genes by recruiting chromatin-modifying proteins in committed B cells. *EMBO J* **30:** 2388-2404

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5:** 621-628

Nutt SL, Urbánek P, Rolink A, Busslinger M (1997) Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced *V*-to-*DJ* recombination at the *IgH* locus. *Genes Dev* **11:** 476-491

Nutt SL, Morrison AM, Dörfler P, Rolink A, Busslinger M (1998) Identification of BSAP (Pax-5) target genes in early B-cell development by loss- and gain-of-function experiments. *EMBO J* **17:** 2319-2333

Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, Lehrach H, Soldatov A (2009) Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res* **37:** e123

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:** 139-140

Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, Cao H, Yu M, Rosenzweig E, Goldy J, Haydock A, Weaver M, Shafer A, Lee K, Neri F, Humbert R, Singer MA, Richmond TA, Dorschner MO, McArthur M, Hawrylycz M, Green RD, Navas PA, Noble WS, Stamatoyannopoulos JA (2006) Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods* **3:** 511-518

Schebesta A, McManus S, Salvagiotto G, Delogu A, Busslinger GA, Busslinger M (2007) Transcription factor Pax5 activates the chromatin of key genes involved in B cell signaling, adhesion, migration and immune function. *Immunity* **27:** 49-63

Shinkai Y, Rathbun G, Lam K-P, Oltz EM, Stewart V, Mendelsohn M, Charron J, Datta M, Young F, Stall AM, Alt FW (1992) RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* **68:** 855-867

Thijs G, Marchal K, Lescot M, Rombauts S, De Moor B, Rouze P, Moreau Y (2002) A Gibbs sampling method to detect overrepresented motifs in the upstream regions of coexpressed genes. *J Computational Biol* **9:** 447-464

Urbánek P, Wang Z-Q, Fetka I, Wagner EF, Busslinger M (1994) Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell* **79:** 901-912

Vilagos B, Hoffmann M, Souabni A, Sun Q, Werner B, Medvedovic J, Bilic I, Minnich M, Axelsson E, Jaritz M, Busslinger M (2012) Essential role of EBF1 in the generation and function of distinct mature B cell types. *J Exp Med* **209:** 775-792

Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, Riss J, Barrett JC, Weinstein JN (2003) GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol* **4:** R28

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