Appendix

Table of contents

1. Appendix Supplementary Methods

Mice

The following mice were maintained on the C57BL/6 genetic background: $Pax5^{+/-}$ (Urbánek et al, 1994), $Cdkn2ab^{+/-}$ (Krimpenfort et al, 2007), $Rosa26^{BirA/BirA}$ (Driegen et al, 2005), $Meox2^{Cre/+}$ (Tallquist & Soriano, 2000), $Trp53^{fl/fl}$ (Jonkers et al, 2001), $Nras^{LSL-G12D/+}$ (Haigis et al, 2008) the transgenic FLPe (Rodriguez et al, 2000) and transgenic EµSRα-$Stat5a$(S711F) mouse strain (referred to as ca-$Stat5a$ in this paper) (Joliot et al, 2006). All animal experiments were carried out according to valid project licenses, which were approved and regularly controlled by the Austrian Veterinary Authorities. Mice were monitored for weight loss on a weekly basis and other signs of disease. Mice were sacrificed, when they lost more than 10% of their body weight or showed signs of terminal illness.

Generation of $Pax5^{Etv6/+}$, $Pax5^{Foxp1/+}$ and $Pax5^{Prd/+}$ mice

All three alleles were generated by cloning of targeting vectors as shown in Appendix Fig. S2A-C and described in detail in the corresponding figure legend. To convert the $Pax5$-$Etv6$ targeting vector to the $Pax5$-$Foxp1$ targeting vector, we replaced the $PAX5$-$ETV6$ cDNA sequence with the corresponding $PAX5$-$FOXP1$ cDNA sequence by taken advantage of a unique BspEI site in $PAX5$ exon 4 and a KpnI site upstream of the C-terminal tag sequence. To generate the $Pax5$-Prd targeting vector, we replaced the $PAX5$-$ETV6$ cDNA sequence with mouse exon 4 sequences fused in frame to the same C-terminal tag sequence and linked to a downstream IRES-luciferase gene. Targeting was performed in the ES cell line A9. PCR-positive clones were verified by Southern blot analysis (Appendix Fig. S2D) before injection into C57BL/6 blastocysts and the generation of $Pax5^{Stop-Etv6-Neo/+}$, $Pax5^{Stop-Foxp1-Neo/+}$ and $Pax5^{Prd-neo/+}$ mice. The $Pax5^{Stop-}$

$Pax5^{Etv6/+}$ and $Pax5^{Stop-Foxp1/+}$ and were obtained by crossing the $Pax5^{Stop-Etv6-Neo/+}$, $Pax5^{Stop-Foxp1-Neo/+}$ with the transgenic FLPe line to remove the neomycin resistance gene (Appendix Fig. S2A,B). The $Pax5^{Etv6/+}$ and $Pax5^{Foxp1/+}$ mice were generated by subsequent crossing with the $Meox2^{Cre/+}$ mouse. The $Pax5^{Prd/+}$ mouse was obtained by crossing the $Pax5^{Prd-neo/+}$ mice with the transgenic FLPe mouse (Appendix Fig. S2C). The following primers were used for PCR genotyping of $Pax5^{Etv6/+}$, $Pax5^{Foxp1/+}$ and $Pax5^{Prd/+}$ mice: 5'-GGTTCTGAGGGTCCAGGAAT-3' (a), 5'-AGGTTCAGCCCTTGGAGAAT-3' (b) and 5'-CGAGAACTTGTTTATTGCAGCTT-3' (c). The $Pax5^{Etv6/+}$, $Pax5^{Foxp1/+}$ and $Pax5^{Prd/+}$ alleles were identified as a 320-bp PCR fragment by using the primer pair b/c and the wild-type *Pax5* allele as 744-bp PCR fragment by using the primer pair a/b.

Antibodies

The following monoclonal antibodies were used for flow cytometric analysis of the mouse spleen, lymph node and bone marrow: B220/CD45R (RA3-6B2), CD2 (RM2-5), CD11b/Mac1 (M1/70), CD19 (1D3/6D5), CD21/CD35 (7G6), CD23 (B3B4), CD25/IL-2Rα (PC61), CD43 (S7), CD49b/DX5 (DX5), CD90.2/Thy1.2 (53-2.1), CD93/AA4.1 (AA4.1), CD115/M-CSFR (AFS98), CD117/c-Kit (2B8), CD127/IL-7Rα (A7R34), CD135/Flt3 (A2F10), CD161/NK1.1 (PK136), Gr1 (RB6-8C5), IgD (11.26c), Igκ (187.1), IgM (II/41 or eB121-15F9), pre-BCR (SL156), TCRβ (H57-597).

The following antibodies (produced in-house) were used; anti-Pax5 (rabbit polyclonal antibody recognizing the N-terminal Prd domain of Pax5 [codon 17-145]) (Adams et al, 1992) for immunoblot analysis and anti-Pax5 (rabbit polyclonal antibody recognizing the central and C-terminal sequences of Pax5 [codon 189-391]) (Adams et al, 1992) for ChIP-seq analysis. The anti-B220/CD45R (RA3-6B2) was used for immunohistochemical analysis.

Antibodies recognizing the following histone tail modifications were used for ChIP analysis: anti-H3K4me2 (rabbit polyclonal Ab 07-030; Merck Millipore), anti-H3K4me3 (rabbit polyclonal Ab ab8580-100; Abcam), anti-H3K9ac (rabbit polyclonal Ab 07-352; Merck Millipore), H3K27me3 (rabbit polyclonal Ab; from T. Jenuwein, Max Planck Institute of Immunobiology and Epigenetics, Freiburg).

Flow cytometric sorting and definition of mouse hematopoietic cell types

The different hematopoietic cell types were identified by flow cytometry or sorted with a FACS Aria machine (Becton Dickinson) as follows: pro-B ($CD19^+B220^+$c-Kit$^+CD2^-$[or $CD25^-$]IgM$^-$ IgD$^-$), large pre-B cells ($CD19^+B220^+CD2^+$c-Kit$^-$IgM$^-$IgD$^-$FSC^{hi}), small pre-B ($CD19^+B220^+CD2^+$[or $CD25^+$]c-Kit$^-$IgM$^-$IgD$^-$FSC^{lo}), immature B ($CD19^+B220^+$IgM^{hi}IgD^{lo}), transitional B ($CD19^+B220^+$IgM^{hi}IgD^{hi}), mature B ($CD19^+B220^+$IgM^{lo} IgD^{hi}), recirculating B ($CD19^+B220^+$IgM^{lo} IgD^{hi}), total B cells ($CD19^+B220^+$), granulocytes (Gr1^{hi}Mac1^{hi}), macrophages ($Mac1^+$M-CSFR$^+$), NK cells ($NK1.1^+DX5^+TCR\beta^-$), DN ($CD4^-CD8^-$Thy1.2$^+$), DP ($CD4^+CD8^+$), CD4 SP ($CD4^+CD8^-$) and CD8 SP ($CD4^-CD8^+$) thymocytes.

Competitive bone marrow transplantation

Bone marrow cells from *Pax5*$^{+/-}$ mice (Ly5.2) were mixed at a 1:1 ratio with bone marrow cells from WT mice (C57BL/6-Ly5.1) followed by intravenous injection (3×10^6 donor cells per mouse) into C57BL/ 6-Ly5.1 mice, which were irradiated with a single dose of 10 Gy 24 h before injection. Bone marrow cells were analyzed by flow cytometry 12 weeks after bone marrow transplantation.

Transwell migration

Short-term cultured pro-B cells were resuspended at a density of 5×10^6 cells per ml in IMDM supplemented with 0.25% heat-inactivated FCS, 1 mM glutamine, 50 μM β-mercaptoethanol and 1% conditioned supernatant of J558L mouse plasmacytoma cells that produce recombinant IL-7. The pro-B cell suspension (100 μl) was placed in the upper compartment and pro-B cell medium (IMDM containing IL-7) (600 μl) containing 400 ng/ml CXCL12 (SDF-1α) in the lower compartment of a Transwell chamber (pore size, 5 μm; Corning). Pro-B cells that migrated into the lower chamber were measured after 2 h with a Casy cell counter, and their frequency relative to the total cells per well was calculated.

Transplantations

Lymph node cells from tumor mice ($0.1\text{-}1\times10^6$ donor cells per mouse) were injected into sublethally irradiated C57BL/6 mice (4.5-6 Gy). Tumor development was monitored, and mice were sacrificed when loosing more than 10% of body weight or showing signs of terminal illness.

Nuclear extract preparation and streptavidin pulldown

Nuclear extracts were prepared from short-term cultured $Rosa26^{BirA/+}$ $Pax5^{Etv6/+}$, $Rosa26^{BirA/+}$ $Pax5^{Foxp1/+}$ and $Rosa26^{BirA/+}$ $Pax5^{Prd/+}$ mice pro-B cells and streptavidin pulldown experiments were performed with Dynabeads (M280; Invitrogen) as described in detail (Minnich et al, 2016). The precipitated proteins were resuspended in 2x SDS sample buffer, eluted from the beads by boiling and separated by SDS-PAGE followed by immunoblot analysis with an anti-Pax5 antibody detecting the paired (Prd) domain.

Control B-ALL tumor

Murine control B-ALLs driven by Myc, Nras^{G12D} and loss of Trp53 were generated by retroviral transduction of vectors expressing c-Myc and Cre-ER^{T2} (pMSCV-c-Myc-IRES-rtTA3-PGK-CreERT), into MACS-selected B220$^+$ bone marrow progenitors isolated from $Nras^{LSL\text{-}G12D/+}$ $Trp53^{fl/fl}$ mice. Following transduction, cells were treated with 0.2 μM 4-OHT and cultured for 10 days in pro-B cell conditions (on OP9 feeder cells in RPMI1640 supplemented with 10% FCS, 10 mM HEPES pH7.5, 50 μM β-mercaptoethanol, 20 mM glutamine and 1 ng/μl IL-7). The cultured cells were transplanted into syngeneic recipient mice, where they promoted the development of aggressive B-ALL (B220$^+$CD19$^+$CD43$^+$, predominantly IgM$^-$) with a median survival of 34 days (data not shown). RNA-seq analysis was performed with primary B-ALL cells that were isolated from the bone marrow of terminally diseased mice and cultured for 2 passages in RPMI1640 supplemented with 10% FCS, 10 mM HEPES pH7.5 and 10 mM glutamine.

Patient samples and human cells

Normal human pre–B cells were sorted as CD34$^-$CD19$^+$CD10$^+$ cells from the bone marrow of healthy donors. Informed consent for all patient samples was obtained from all subjects.

Histopathology

Sections (2-μm) of fixed and paraffin-embedded tissues were stained with hematoxylin and eosin for histological analysis or with a B220-specific antibody (RA3-6B2) for immunohistochemical analysis. Peripheral blood smears were analyzed with May-Grünwald-Giemsa staining. Specimens were scanned using a Mirax slide scanner.

Bio-ChIP-seq analysis

Pro-B cells isolated from the bone marrow of $Rosa26^{BirA/+} Pax5^{Etv6/+}$ or $Rosa26^{BirA/+} Pax5^{Prd/+}$ mice were cultured for 4-5 days on OP9 cells in the presence of IL-7. Chromatin from ~$5x10^7$ pro-B cells or $Cdkn2ab^{+/-} Pax5^{Etv6/+} Rosa26^{BirA/+}$ B-ALL cells (Tu-43) was prepared using a lysis buffer containing 0.25% SDS prior to chromatin precipitation by streptavidin pulldown (Bio-ChIP), as described (Ebert et al, 2011; Revilla-i-Domingo et al, 2012). The precipitated genomic DNA was quantified by real-time PCR, and about 1-5 ng of ChIP-precipitated DNA was used for library preparation.

cDNA preparation for RNA-sequencing

RNA from *ex vivo* sorted B cells was isolated with the RNeasy Plus Mini Kit (Qiagen). mRNA was obtained by two rounds of poly(A) selection using the Dynabeads mRNA purification kit (Invitrogen) followed by fragmentation to a size of 100-600 nucleotides. The fragmented mRNA was used as template for first-strand cDNA synthesis with random hexamers using the Superscript III First-Strand Synthesis System (Invitrogen). The second-strand cDNA was synthesized with 100 mM dATP, dCTP, dGTP and dUTP in the presence of RNase H, E. coli DNA polymerase I and DNA ligase (Invitrogen). The incorporation of dUTP allowed elimination of the second strand during library preparation (see below), thereby preserving strand specificity (Parkhomchuk et al, 2009).

Library preparation and Illumina deep sequencing

About 1-5 ng of cDNA or ChIP-precipitated DNA were used as starting material for the generation of sequencing libraries with the NEBNext Quick Ligation Module, NEBNext Endrepair Module and NEBNext dA-Tailing module or NEBNext Ultra Ligation Module and NEBNext End Repair/dA-Tailing module. DNA fragments of the following sizes were selected

for the different experiments: 200–500 bp for ChIP-seq and 150–700 bp for RNA-seq with AMPure XP beads (Beckman Coulter). For strand-specific RNA-sequencing, the uridines present in one cDNA strand were digested with uracil-N-glycosylase (New England Biolabs) as described (Parkhomchuk et al, 2009) followed by PCR amplification with the KAPA Real Time Amplification kit (KAPA Biosystems). Completed libraries were quantified with 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 HiSeq 2000/2500 system with a read length of 50 nucleotides (single-read) or 100/125 nucleotides (paired-end) according to the manufacturer's guidelines. Table EV6 provides further information about all sequencing experiments of this study.

Sequence alignment

Sequence reads that passed the Illumina quality filtering were considered for alignment. For ChIP-seq, the reads were aligned to the mouse genome assembly version of July 2007 (NCBI37/mm9), using the Bowtie program version 12.5. For comparing Pax5-Etv6, Pax5 and Prd binding, we truncated all reads from 50 to 36 nucleotides before sequence alignment.

In case of mouse RNA-seq experiments, reads corresponding to mouse ribosomal RNAs (BK000964.1 and NR046144.1) were removed, and the remaining reads were cut down to a read length of 44 nucleotides. As the mouse RNA-seq data were either sequenced as single-read or paired-end, only the first read (in the case of paired-end data) was considered for alignment to the mouse transcriptome (genome assembly version of July 2007 NCBI37/mm9) using TopHat version 1.4.1 (Trapnell et al, 2009).

For human RNA-seq data, reads corresponding to human ribosomal RNAs (U13369.1 and NR_023371.1) were removed. The remaining reads were cut down to a read length of 94 nucleotides and aligned to the human transcriptome (genome assembly version of February 2009 NCBI37/hg19) using TopHat version 1.4.1 (Trapnell et al, 2009) in the paired-end mode.

Database of RefSeq-annotated genes

Peak-to-gene assignment and calculation of RNA expression values were all based on the RefSeq database, which was downloaded from UCSC on January 10th, 2014. The annotation of immunoglobulin and T cell receptor genes was incorporated from Ensembl release 67 and Ensembl release 75 for mouse and human, respectively (Cunningham et al, 2015). Genes with overlapping exons were flagged and double entries (i.e. exactly the same gene at two different

genomic locations) were renamed. Identical genes with more than one assigned gene symbol were flagged. Genes with several transcripts were merged to consensus-genes consisting of a union of all underlying exons using the fuge software (I. Tamir, unpublished). This resulted in 24,726 mouse and 25,714 human gene models.

Orthologous genes between mouse and human were defined by the HomoloGene database release 68 (http://www.ncbi.nlm.nih.gov/pubmed/26615191), which was downloaded from NCBI on September 22, 2014. Only genes with a single ortholog match in the other genome were considered for gene set enrichment analysis, which resulted in 17,081 and 16,657 matched genes for mouse and human, respectively.

Peak calling ChIP-seq data and target gene assignment

Peaks were determined by using the MACS program version 1.3.6.1 (Zhang et al, 2008) with default parameters, a cutoff of a *P* value of $< 10^{-10}$ and a genome size of 2,654,911,517 bp (mm9). The identified peaks were then assigned to target genes as described (Revilla-i-Domingo et al, 2012). For comparisons of ChIP-seq data, we down-sampled all reads before peak calling to the ChIP-seq experiment with the lowest number of aligned reads.

Read density heat maps

Read densities were calculated using the JNOMICS program (I. Tamir, unpublished). Associated heat map visualizations were implemented using R (http://www.R-project.org) and were wrapped with customized bash scripts for command line usage.

De novo motif discovery

For *de novo* motif discovery, we used the MEME-ChIP suite (version 4.9.1) (Machanick & Bailey, 2011) to predict the most significant motifs present in the 300 base pairs centered at the peak summit of the top 300 sequences, as sorted by the fold enrichment score of the MACS program. For motif scanning, the 300 base pair fragments of all peaks were subsequently scanned for the Pax5 and Ets motif using FIMO with a sensitive *P* value threshold of 0.1 (Grant et al, 2011). The best scoring motif per peak was used for further analysis.

Analysis of RNA-seq data

For analysis of differential gene expression, the number of reads per gene was counted using HTseq version 0.5.3 (Anders et al, 2015) with the overlap resolution mode set to 'union'. The

datasets were grouped according to cell type and organism and analyzed using the R package DESeq2 version 1.4.1 (Love et al, 2014). Sample normalizations and dispersion estimations were conducted using the default DESeq2 settings. In detail, the following DESeq2 analyses were performed: all mouse pro B cell samples were analyzed together considering the genotype effects (model design formula: "~ genotype"). Mouse tumor cells and mouse wild-type large pre-B cells were also grouped and analyzed in the same manner, as were human PAX5-ETV6$^+$ B-ALL cells and normal pre-B cells from healthy donors.

Principal component analysis

For principal component analysis, read counts were rlog normalized with the blind option set to true, using the R package DESeq2 version 1.4.1. Only the top 500 most varying genes of wild-type pro-B, large pre-B and small pre-B cell samples were considered for calculating the principal components. Tumor samples were then classified using the previously identified components.

Fusion gene detection and quantification

The RNA-seq reads of tumor samples were cut down to 94 bp and aligned to the human genome assembly version of February 2009 (NCBI37/hg19) with the TopHat2 version 2.0.10 (Kim et al, 2013) in the paired-end and fusion-search mode. Predicted *PAX5-ETV6* fusion breakpoints were then used to create a custom human genome reference by adding the corresponding *PAX5-ETV6* fusion sequences. The tumor sample reads were again aligned to the custom genome with TopHat version 1.4.1 (Trapnell et al, 2009) in the paired-end mode. For the data shown in Appendix Fig. S7A, potential optical or PCR duplicates were detected and filtered by using MarkDuplicates provided by the Picard toolkit version 1.82 (http://broadinstitute.github.io/picard). Reads per million were calculated based on uniquely matching reads only. Reads covering the exonic *PAX5-ETV6* fusion junction by at least one base pair were used as proof for the fusion detection.

For measuring the relative abundance of full-length *PAX5* and *PAX5-ETV6* fusion transcripts (shown in Appendix Fig. S7B,C), the reads covering the *PAX5* exon 4 and *PAX5* exon 5 (detecting the full-length *PAX5* transcript) or covering the predicted *PAX5-ETV6* breakpoints (detecting the fusion *PAX5-ETV6* transcript) were counted and the relative ratio of the two transcripts was calculated.

Gene set enrichment analysis

Gene Set Enrichment Analysis was performed using the GSEA software from the Broad Institute (Subramanian et al, 2005). Genes were ranked based on $\log_2$ fold change from the DESeq2 package and compared to our own defined gene sets (Figs. 6E,F and 7A). For interspecies comparisons, gene sets were converted to their corresponding orthologous genes before analysis (Fig. 7A). For the analysis shown in Fig. S7F, the RNA-seq data of 9 human PAX5-JAK2[+] B-ALLs (3 samples sequenced in-house and 6 samples sequenced by the St. Jude-Washington Pediatric Genome Project) were compared to the RNA-seq data of the 9 human PAX5-ETV6[+] B-ALLs as described above (Analysis of RNA-seq data). The ranked genes were then compared to the following gene sets: "Upregulated in pre-BCR[+] B-ALLs" containing 19 genes that are upregulated in pre-BCR[+] B-ALLs compared to pre-BCR[−] B-ALLs (Fig. 2A-C [middle panel] of Geng et al, 2015) and "Downregulated in pre-BCR[+] B-ALLs" containing 21 genes that are downregulated in pre-BCR[+] B-ALLs compared to pre-BCR[−] B-ALLs (Fig. 2A-C [bottom panel] of Geng et al, 2015).
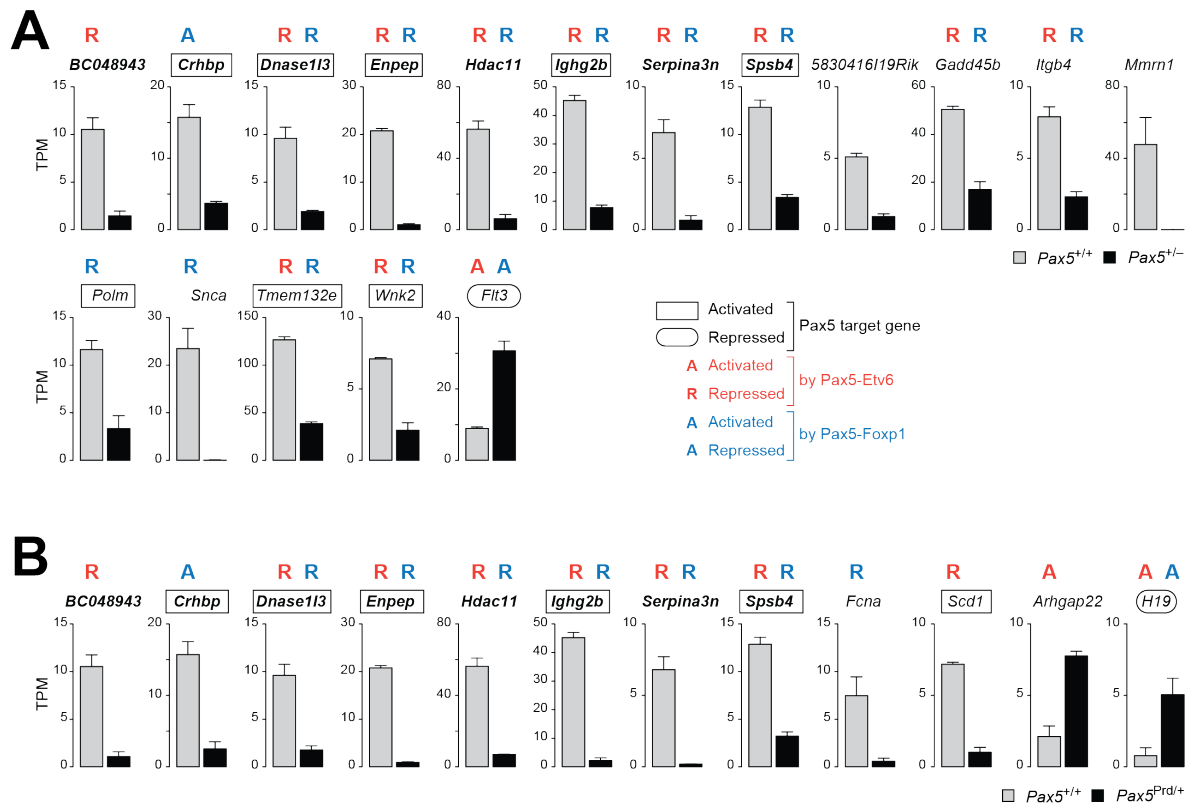
2. Appendix Supplementary References

Adams B, Dörfler P, Aguzzi A, Kozmik Z, Urbánek P, Maurer-Fogy I, Busslinger M (1992) *Pax-5* encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev* 6: 1589-1607

Anders S, Pyl PT, Huber W (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166-169

Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L, Giron CG, Gordon L, Hourlier T, Hunt SE, Janacek SH, Johnson N, Juettemann T, Kahari AK, Keenan S, et al (2015) Ensembl 2015. *Nucleic Acids Res* 43: D662-D669

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

Geng H, Hurtz C, Lenz KB, Chen Z, Baumjohann D, Thompson S, Goloviznina NA, Chen WY, Huan J, LaTocha D, Ballabio E, Xiao G, Lee JW, Deucher A, Qi Z, Park E, Huang C, Nahar R, Kweon SM, Shojaee S, et al (2015) Self-enforcing feedback activation between BCL6 and pre-B cell receptor signaling defines a distinct subtype of acute lymphoblastic leukemia. *Cancer Cell* **27:** 409-425

Grant CE, Bailey TL, Noble WS (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27: 1017-1018

Haigis KM, Kendall KR, Wang Y, Cheung A, Haigis MC, Glickman JN, Niwa-Kawakita M, Sweet-Cordero A, Sebolt-Leopold J, Shannon KM, Settleman J, Giovannini M, Jacks T (2008) Differential effects of oncogenic K-Ras and N-Ras on proliferation, differentiation and tumor progression in the colon. *Nat Genet* 40: 600-608

Joliot V, Cormier F, Medyouf H, Alcalde H, Ghysdael J (2006) Constitutive STAT5 activation specifically cooperates with the loss of p53 function in B-cell lymphomagenesis. *Oncogene* 25: 4573-4584

Jonkers J, Meuwissen R, van der Gulden H, Peterse H, van der Valk M, Berns A (2001) Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. *Nat Genet* 29: 418-425

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

Krimpenfort P, Ijpenberg A, Song J-Y, van der Valk M, Nawijn M, Zevenhoven J, Berns A (2007) p15Ink4b is a critical tumour suppressor in the absence of p16Ink4a. *Nature* 448: 943-946

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

Machanick P, Bailey TL (2011) MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 27: 1696-1697

Minnich M, Tagoh H, Bönelt P, Axelsson E, Fischer M, Cebolla B, Tarakhovsky A, Nutt SL, Jaritz M, Busslinger M (2016) Multifunctional role of the transcription factor Blimp-1 in coordinating plasma cell differentiation. *Nat Immunol* 17: 331-343

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

Revilla-i-Domingo R, Bilic I, Vilagos B, Tagoh H, Ebert A, Tamir IM, Smeenk L, Trupke J, Sommer A, Jaritz M, Busslinger M (2012) The B-cell identity factor Pax5 regulates distinct transcriptional programmes in early and late B lymphopoiesis. *EMBO J* 31: 3130-3146

Rodriguez CI, Buchholz F, Galloway J, Sequerra R, Kasper J, Ayala R, Stewart AF, Dymecki SM (2000) High-efficiency deleter mice show that FLPe is an alternative to Cre-*loxP*. *Nat Genet* 25: 139-140

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102: 15545-15550

Tallquist MD, Soriano P (2000) Epiblast-restricted Cre expression in MORE mice: a tool to distinguish embryonic vs. extra-embryonic gene function. *Genesis* 26: 113-115

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25: 1105-1111

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

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
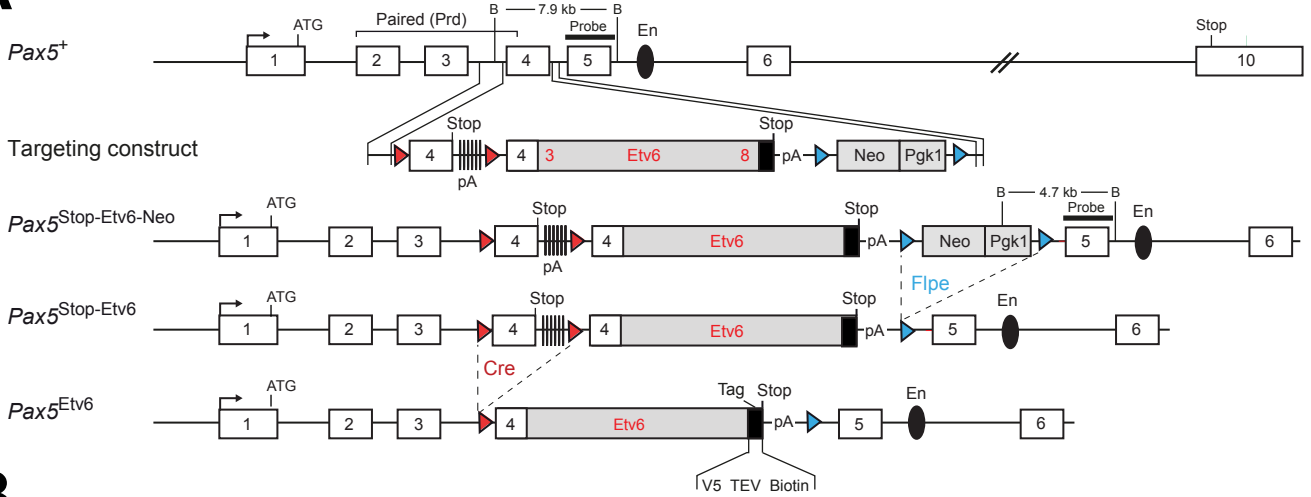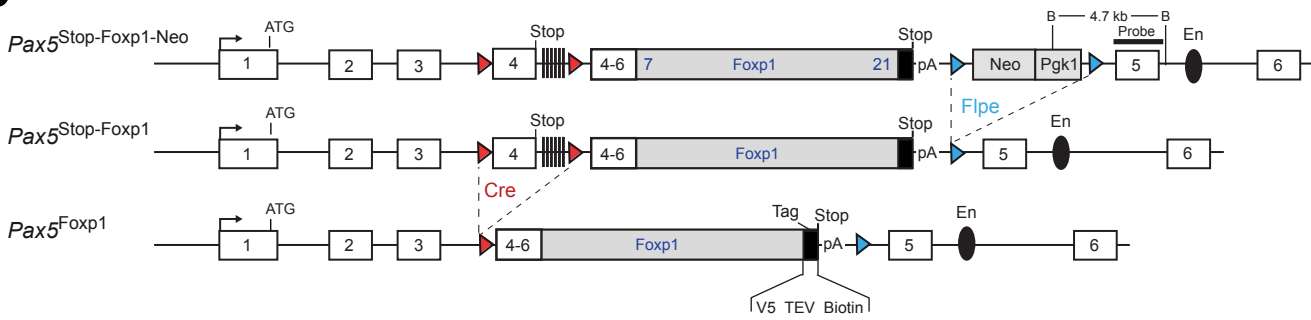
3. Appendix Figures

Appendix Figure S1

Appendix Figure S1. Differentially expressed genes in *Pax5*$^{+/-}$ and *Pax5*$^{Prd/+}$ pro-B cells.
(A,B) Differential gene expression between *ex vivo* sorted *Pax5*$^{+/+}$ (gray) and *Pax5*$^{+/-}$ (black) pro-B cells (A) as well as *Pax5*$^{+/+}$ (gray) and *Pax5*$^{Prd/+}$ (black) pro-B cells (B). Eight genes, which were similarly regulated in *Pax5*$^{+/-}$ and *Pax5*$^{Prd/+}$ pro-B cells, are indicated in bold face. The expression of each gene is shown as normalized TPM value with SEM, based on at least two RNA-seq experiments for each cell type. Differentially expressed genes were identified in Figures 1D (*Pax5*$^{+/-}$) and 2C (*Pax5*$^{Prd/+}$) by an expression difference of > 3-fold, an adjusted *P* value of < 0.05 and a TPM value of > 5 in one of the two cell types. TPM, Transcripts per million. Genes, which were previously identified as activated (boxed) or repressed (oval) Pax5 target genes in pro-B cells (Revilla-i-Domingo et al, *EMBO J* 31: 3130-3146), are indicated together with those genes that were activated (A) or repressed (R) by Pax5-Etv6 (red) or Pax5-Foxp1 (blue), as shown in Fig. 3B,C. Notably, most Pax5-activated genes identified by decreased expression in *Pax5*$^{+/-}$ or *Pax5*$^{Prd/+}$ pro-B cells appeared to be 'repressed' by Pax5-Etv6 and Pax5-Foxp1 (Fig. 3B,C), although this effect is likely caused by the loss of one wild-type *Pax5* allele rather than the expression of the fusion protein in *Pax5*$^{Etv6/+}$ and *Pax5*$^{Foxp1/+}$ pro-B cells.
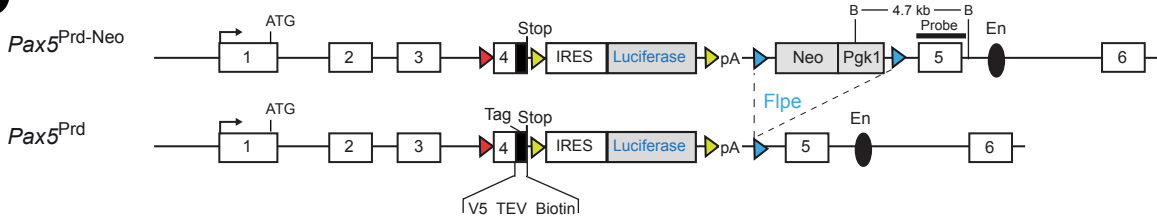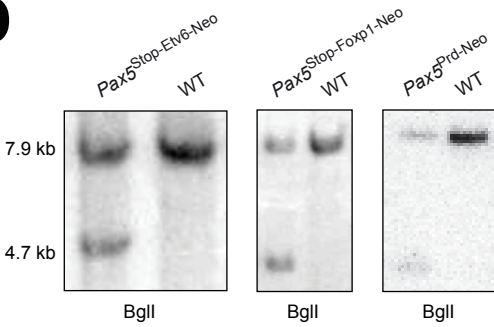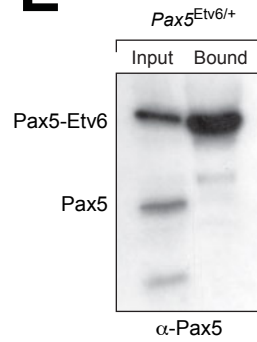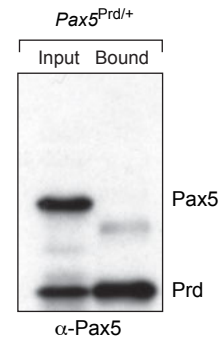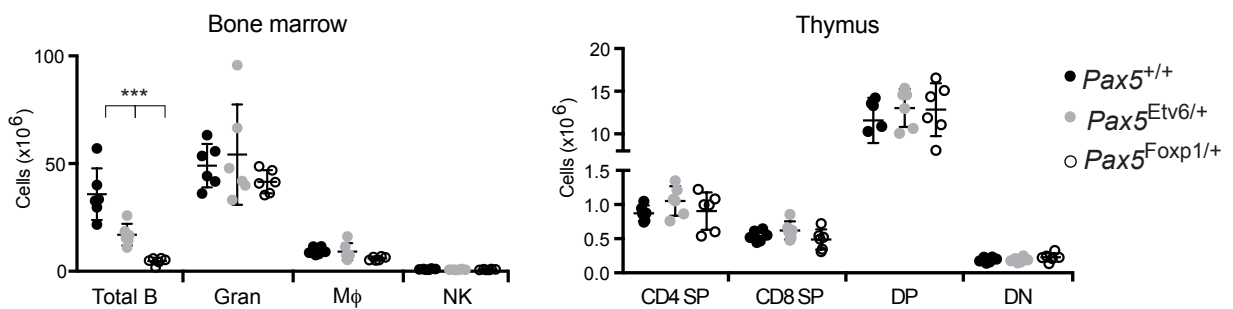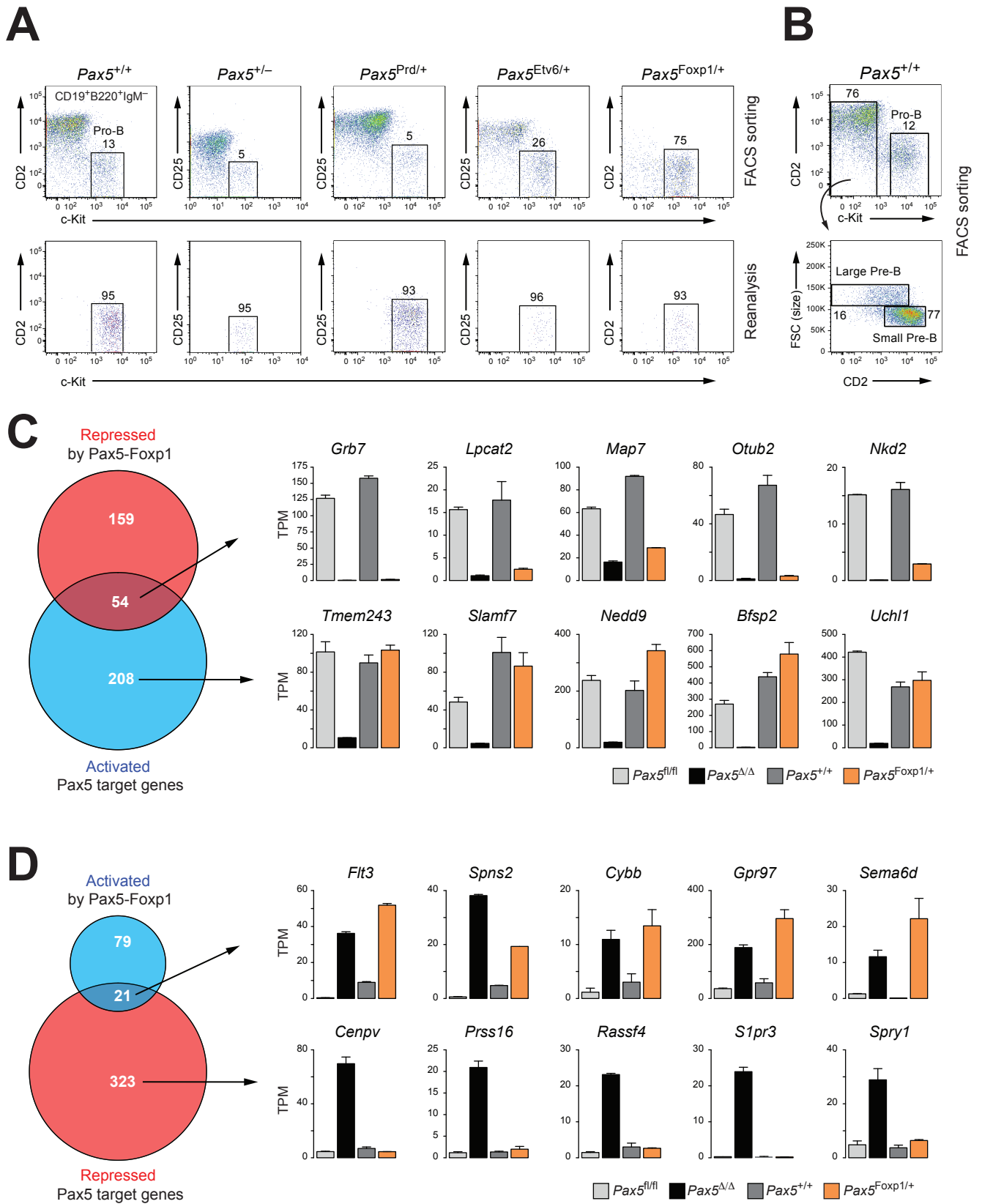
Appendix Figure S2

Appendix Figure S2. Generation and characterization of $Pax5^{Etv6}$, $Pax5^{Foxp1}$ and $Pax5^{Prd}$ alleles.
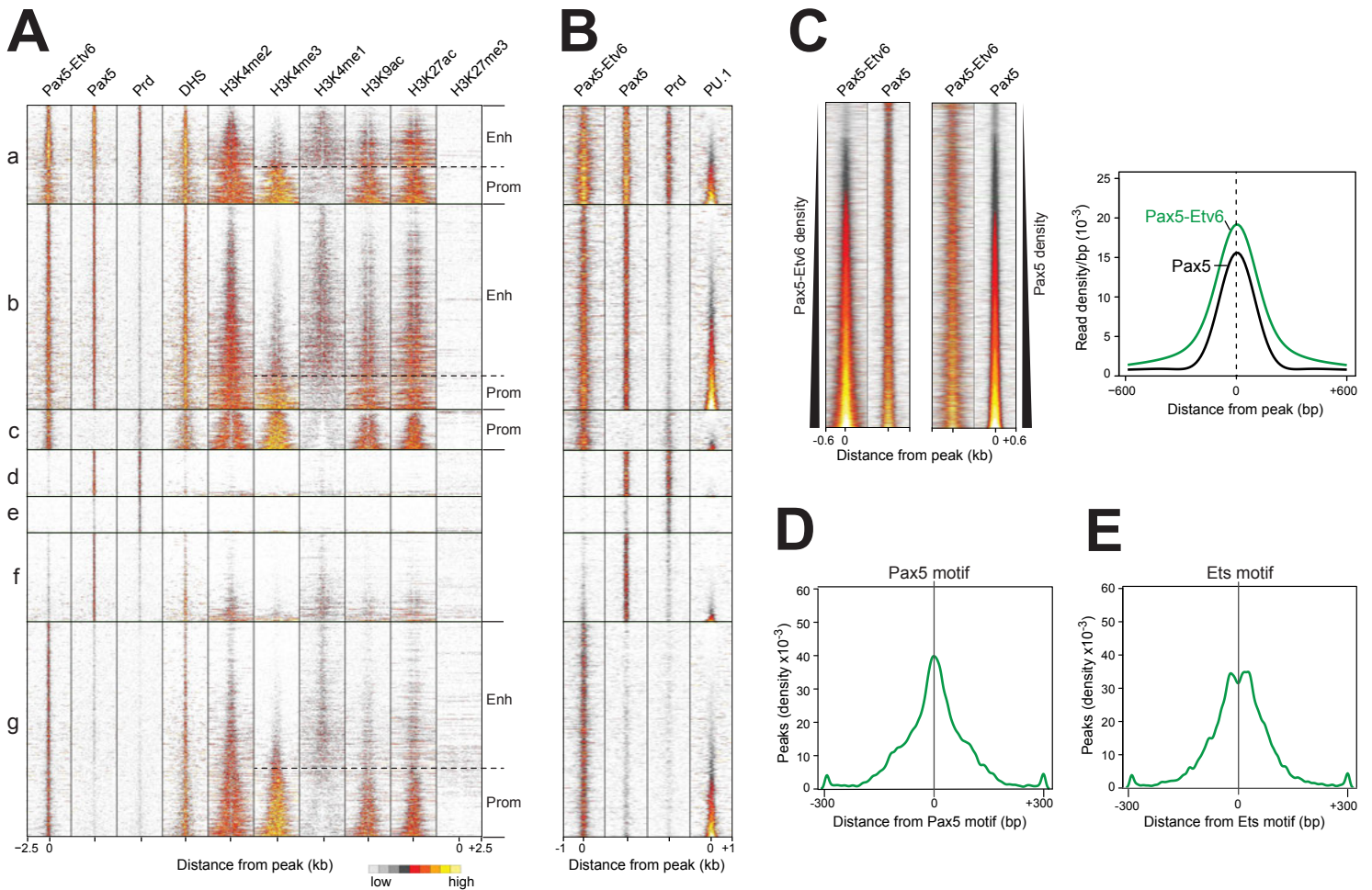
(A-C) Generation of the $Pax5^{Etv6}$ (A), $Pax5^{Foxp1}$ (B) and $Pax5^{Prd}$ (C) alleles. The $Pax5^{Etv6}$ allele was generated by replacing exon 4 of *Pax5* with the following sequences in the 5' to 3' direction: (i) a *loxP*-flanked 1.6-kb DNA fragment containing exon 4 linked to a stop codon and six copies of the SV40 polyadenylation region, (ii) a 1.63-kb DNA fragment containing the 3' splice site and start of *Pax5* exon 4 linked to human cDNA sequences starting in *PAX5* exon 4 and encoding the remaining PAX5-ETV6 protein coupled to a C-terminal tag sequence and a SV40 polyadenylation region and (iii) *frt*-flanked 2-kb DNA fragment containing the mouse phosphoglycerate kinase (*Pgk1*) promoter linked to the neomycin (neo^r) resistance gene. Brackets indicate the two homology regions mediating recombination in ES cells. The $Pax5^{Foxp1}$ allele (B) was generated by replacing the *ETV6* cDNA sequence with the corresponding *FOXP1* cDNA sequence in the targeting construct. The $Pax5^{Prd}$ allele (C) was generated by directly adding the C-terminal tag sequence in frame in *Pax5* exon 4 followed by a *rox*-flanked IRES-Luciferase expression cassette and the *frt*-flanked DNA sequence containing the mouse phosphoglycerate kinase (*Pgk1*) promoter linked to the neomycin (neo^r) resistance gene. *LoxP*, *frt* and *rox* sites are indicated by red, blue and yellow arrowheads, respectively. The C-terminal tag sequences added at the last codon of human *ETV6* and *FOXP1* as well as in *Pax5* exon 4 (Prd) contained an epitope for V5 antibodies, two cleavage sites for the TEV protease and a biotin acceptor sequence (Biotin) for biotinylation by the *E. coli* biotin ligase BirA (de Boer *et al.*, 2003, Proc. Natl. Acad. Sci. USA *100*, 7480-7485). The mouse *Pax5* exons are numbered and shown as open boxes. Only the number of the exons coding for the start and end of the human *ETV6* and *FOXP1* cDNA sequences are indicated. A black oval denotes the B cell-specific enhancer (En) in *Pax5* intron 5. The BglI (B) fragments of the wild-type ($Pax5^+$) and all three targeted alleles, which were used for allele identification by Southern blot analysis with the indicated probe, are shown together with their length (in kilobases, kb). The $Pax5^{Etv6}$ (A), $Pax5^{Foxp1}$ (B) and $Pax5^{Prd}$ (C) alleles were generated by sequential Cre- and Flpe-mediated deletion of the *loxP*-flanked *Pax5* exon 4-6xpA sequences and the *frt*-flanked *Pgk1-neo* cassette, respectively. pA, polyadenylation site. (D) Southern blot analysis of wild-type (WT) ES cells and correctly targeted ES cell clones of the indicated genotypes by hybridization of BglI-digested DNA with the DNA probe shown in (A-C). (E, F) Efficient precipitation of *in vivo* biotinylated Pax5-Etv6 and Prd proteins by streptavidin pulldown. Bone marrow pro-B cells from $Pax5^{Etv6/+}$ $Rosa26^{BirA/+}$ (E) and $Pax5^{Prd/+}$ $Rosa26^{BirA/+}$ (F) mice were short-term cultured *in vitro* prior to nuclear extract preparation and precipitation of the biotinylated Pax5-Etv6 and Prd proteins with streptavidin beads. The input (1/10) and streptavidin-bound precipitate were analyzed by immunoblotting with a Pax5 antibody recognizing the N-terminal Prd domain. (G) Hematopoiesis in $Pax5^{Etv6/+}$ and $Pax5^{Foxp1/+}$ mice. Absolute cell numbers of the indicated cell types were determined by flow cytometric analysis of the bone marrow and thymus from $Pax5^{+/+}$, $Pax5^{Etv6/+}$ and $Pax5^{Foxp1/+}$ mice (n = 6 per genotype). Data are presented as average cell number with SEM. Gran, granulocytes; Mϕ, macrophage; NK, natural killer cell; DP, double-positive; DN, double-negative; CD4 and CD8 single-positive (SP) thymocytes. Horizontal bars indicate the mean value, and error bars represent SD. *** $P < 0.001$, as determined by two-way analysis of variance (ANOVA).

Appendix Figure S3. Regulation of Pax5 target genes by Pax5-Foxp1 in pro-B cells.

(A) Flow cytometric sorting of pro-B cells from the bone marrow of $Pax5^{+/+}$, $Pax5^{+/-}$, $Pax5^{Prd/+}$, $Pax5^{Etv6/+}$ and $Pax5^{Foxp1/+}$ mice. The purity of the sorted cells was determined by flow cytometric reanalysis. The flow cytometric data of the different cell types cannot be directly compared, as they were generated on different days. Note that the expression of the cell surface markers CD2 and CD25 are known to be similarly induced during the pro-B-to-pre-B cell transition in adult bone marrow and are therefore interchangeably used for the definition of pro-B as $CD19^{+}$c-$Kit^{+}CD2^{-}$ or $CD19^{+}$c-$Kit^{+}CD25^{-}$ cells. (B) Flow cytometric sorting of pro-B cells ($CD19^{+}B220^{+}$c-$Kit^{+}CD2^{-}IgM^{-}IgD^{-}$), large pre-B cells ($CD19^{+}CD2^{+}$c-$Kit^{-}IgM^{-}IgD^{-}$, large size) and small pre-B cells ($CD19^{+}CD2^{+}$c-$Kit^{-}IgM^{-}IgD^{-}$, small size) from the bone marrow of $Pax5^{+/+}$ mice. (C) Venn diagram indicating the overlap between activated Pax5 target genes and Pax5-Foxp1-repressed genes in pro-B cells. The expression of genes indicated to the right was determined by RNA-sequencing of *in vitro* cultured $Pax5^{fl/fl}$ (light grey) and *Vav*-Cre $Pax5^{fl/fl}$ ($Pax5^{\Delta/\Delta}$, black) pro-B cells as well as *ex vivo* sorted $Pax5^{+/+}$ (dark grey) and $Pax5^{Foxp1/+}$ (orange) pro-B cells. Gene expression is shown as normalized expression value (TPM) with SEM, based on two independent RNA-seq experiments for each cell type. Notably, 6 activated Pax5 target genes, which were apparently 'repressed' by Pax5-Foxp1, showed decreased expression already in response to a 2-fold reduction of full-length *Pax5* expression in $Pax5^{+/-}$ pro-B cells (Appendix Fig. S1A). (D) Venn diagram indicating the overlap between repressed Pax5 target genes and Pax5-Foxp1-activated genes in pro-B cells, as described in (C).

A

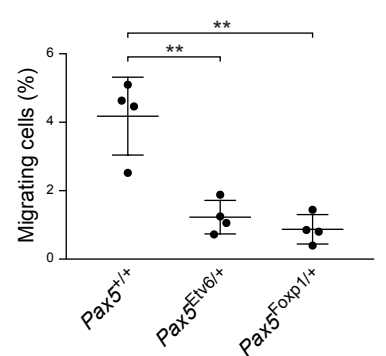Pax5-Etv6 | Pax5 | Prd | DHS | H3K4me2 | H3K4me3 | H3K4me1 | H3K9ac | H3K27ac | H3K27me3

a — Enh / Prom
b — Enh / Prom
c — Prom
d
e
f
g — Enh / Prom

Distance from peak (kb) −2.5 0 / 0 +2.5

low — high

B

Pax5-Etv6 | Pax5 | Prd | PU.1

Distance from peak (kb) −1 0 / 0 +1

C

Pax5-Etv6 | Pax5 (Pax5-Etv6 density)
Pax5-Etv6 | Pax5 (Pax5 density)

Distance from peak (kb) −0.6 0 / 0 +0.6

Read density/bp (10⁻³)
Pax5-Etv6
Pax5
Distance from peak (bp) −600 0 +600

D

Pax5 motif
Peaks (density ×10⁻³)
Distance from Pax5 motif (bp) −300 0 +300

E

Ets motif
Peaks (density ×10⁻³)
Distance from Ets motif (bp) −300 0 +300

F

Activated Pax5-Etv6 target genes

Signal transducer
Gng4
Rgs10
Efhd1
Rgs18
Dtx4
Dlg2
Cd247
Rassf4
Gimap3 *
Gimap4 *
Spry1
Arhgap22
Pde8a
Plcd3
Dusp5 *

Surface receptor
S1pr3
Gpr110
Trem1 *
Cmklr1
Tgfbr3
Ifitm1
Flt3 *
Pcdh9 *
Sell *
Sema7a
Il2ra *
Tmem119

Secreted protein
Ccl9
Tnfsf11 *
Emcn
Sema3g *
Pgf
Ctla2a *

Transcription regulator
Klf4
Id2
Ets2
Sox15

Repressed Pax5-Etv6 target genes

Signal transducer
Pde9a
Nlrp1a *
Grb7 *
Bcar3
Parvb *
Ccm2l
Axin2
Wnk2 *
Mapk11
Dab2ip
Gadd45b *
Gna15
Ppap2b *

Surface receptor
Itgb4 *
Enpep *
Ildr1
Fzd9
Robo3
Faim3 *
Lgr5
Cd22 *

Secreted protein
Slit1 *

Transcription regulator
Nr1d1
Zfp94
Tle6
Ikzf3
Bcl6b

H

Migrating cells (%)
** **
Pax5^{+/+} | Pax5^{Etv6/+} | Pax5^{Foxp1/+}

G

S1pr3 | Ccl9 | Gpr110 | Cmklr1 | Emcn | Sema3g | Arhgap22 | Tgfbr3 | Pcdh9 | Plcd3 | Sell

Sema7a | Cnn1 | Myl10 | Bfsp2 | Itgb4 | Grb7 | Bcar3 | Parvb | Map7 | Dab2ip | Myo1e

TPM

Pax5^{+/+}
Pax5^{Etv6/+}

Appendix Figure S4. Genome-wide Pax5-Etv6 binding in the context of the chromatin landscape of pro-B cells.

(A) Interaction of Pax5-Etv6, Pax5 and Prd with the chromatin landscape in pro-B cells, as revealed by density heat maps. The densities of DHS sites, the 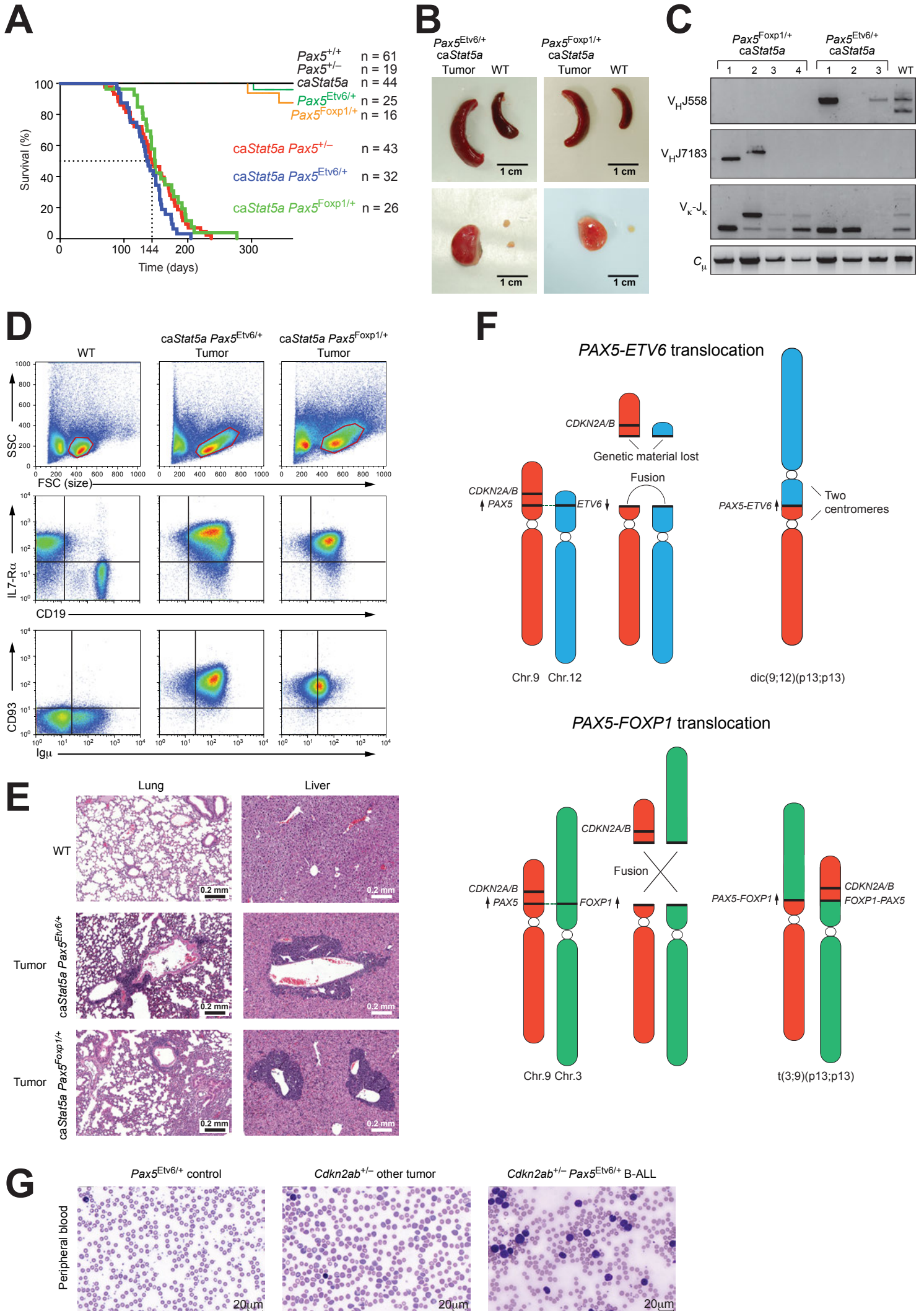repressive chromatin mark H3K27me3 and five active histone modifications (H3K4me1, H3K4me2, H3K4me3, H3K9ac and H3K27ac) are displayed for all Pax5-Etv6, Pax5 and Prd peaks in a region extending from -2.5 kb to +2.5 kb relative to the Pax5 or Pax5-Etv6 peak summit. The heat maps were sorted according to the increasing density of H3K4me3. The pattern of Pax5-Etv6, Pax5 and Prd peaks was determined by multiple overlap analysis as described in Figure 4A. A dashed line indicates the approximate boundary between active promoters (Prom; H3K4me3^{high}, H3K4me1^{low}) and enhancers (Enh; H3K4me3^{low}, H3K4me1^{high}). A density scale from low (grey) to high (yellow) is shown. (B) Binding of PU.1 at Pax5-Etv6 peaks. The heat maps of Pax5-Etv6, Pax5, Prd and PU.1 were sorted according to the increasing density of PU.1 binding and are shown for a region extending from -1 kb to +1 kb relative to the peak summit. (C) Comparison of the width of Pax5 and Pax5-Etv6 peaks in pro-B cells. A density heat map of the common Pax5 and Pax5-Etv6 peaks (sectors a and b in Figure 4A,B) is shown for a region extending from -0.6 kb to +0.6 kb relative to the peak summit and was sorted according to the increasing binding density of Pax5-Etv6 (left) and Pax5 (right). The average binding densities of Pax5-Etv6 (green) and Pax5 (black) is shown for the same region (to the left). (D) Presence of the Pax5 motif at the center of maximal Pax5-Etv6 binding density. The region from -300 bp to +300 bp relative to the Pax5-Etv6 peak summit (sectors a and b in Figure 4A,B) was divided into hundred 6-bp bins, and the position of the bin with the highest Pax5-Etv6 binding density was plotted for each peak on the horizontal axis relative to the centered Pax5 motif (D), which was identified in each peak by scanning with the Pax5 motif (Figure 4A). The vertical axis displays the distribution of numbers of peaks calculated by a density function. (E) Peaks of maximal Pax5-Etv6 binding density on both sides of the centered Ets motif. The highest Pax5-Etv6 binding density of each peak was plotted relative to the centered Ets motif that was found to be closest to the Pax5 motif by scanning with the Ets motif (Figure 4A). (F) Regulated Pax5-Etv6 target genes coding for secreted proteins, cell surface receptors, signal transducers and transcriptional regulators. The indicated target genes were ranked according to their fold expression changes observed between $Pax5^{Etv6/+}$ and $Pax5^{+/+}$ pro-B cells (Figure 3B). The color code refers to >10-fold (green), 5-10-fold (blue), 4-5-fold (red) and 3-4-fold (black). Pax5-Etv6 target genes, which were also activated or repressed by Pax5-Foxp1 (Figure 3B), are indicated by asterisks. (G) Regulated Pax5-Etv6 target genes with functions in cell adhesion and migration. The names of genes that were also activated or repressed by Pax5-Foxp1 are boxed. The expression of the indicated genes is shown as normalized expression value (TPM) with SEM, based on at least two independent RNA-seq experiments of $Pax5^{Etv6/+}$ (green) and $Pax5^{+/+}$ (dark grey) pro-B cells, respectively. (H) Migration of $Pax5^{Etv6/+}$, $Pax5^{Foxp1/+}$ and $Pax5^{+/+}$ pro-B cells (in IL-7-containing IMDM) from the upper compartment of a Transwell through a 5 μm filter into the lower compartment with the same medium additionally containing CXCL12 (400 ng/ml). The migrated cells were counted after 2 h and presented as frequency of migrating cells relative to the total cells per well. Two mice were analyzed per genotype, and short-term cultured pro-B cells of each mouse were separately analyzed in two wells. The data of one of two independent experiments are shown as average frequency with SEM and were analyzed by the Student's t-test; **$P < 0.01$. Each symbol represents the analysis of one well.
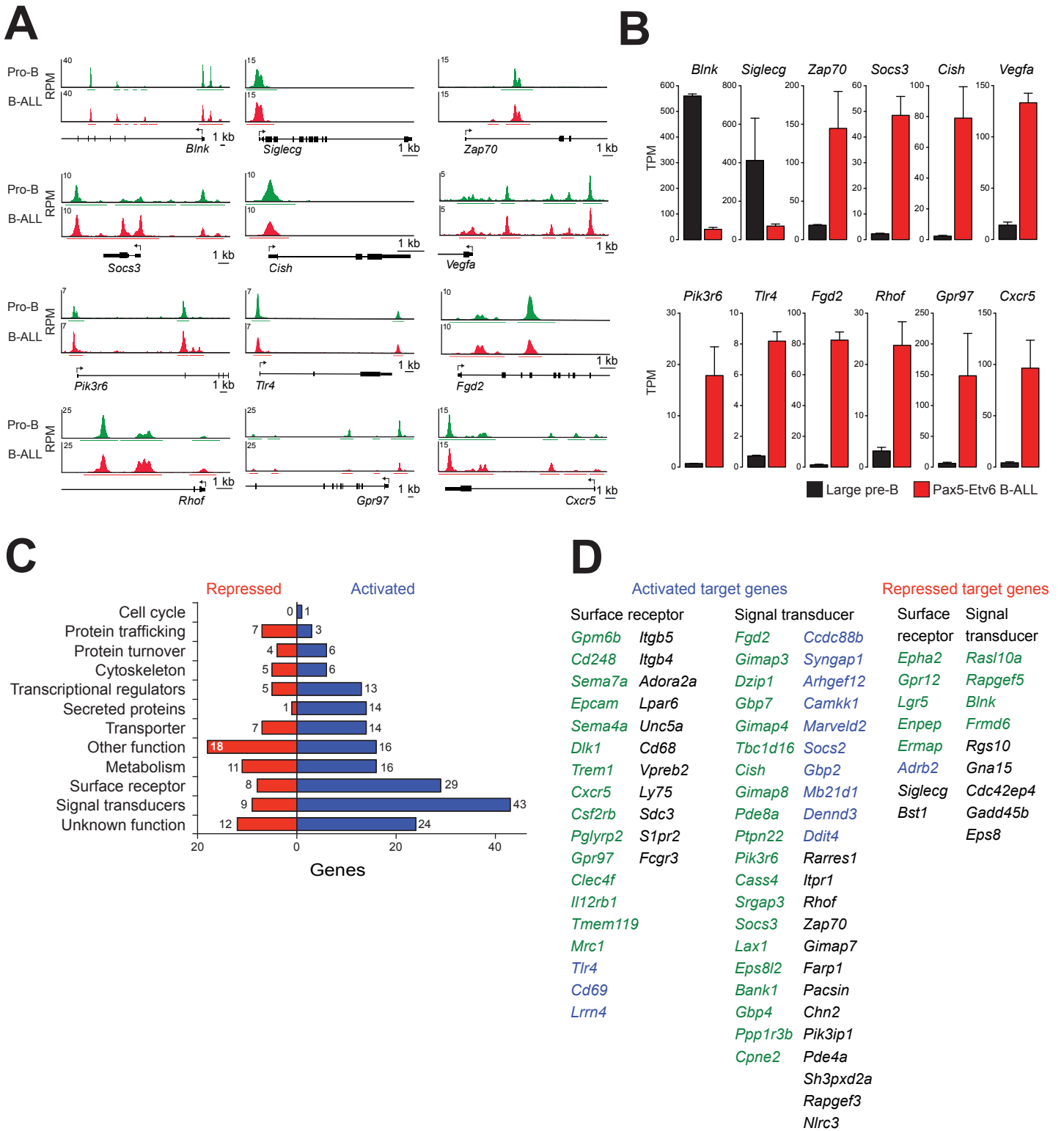
Appendix Figure S5

A Survival (%) vs Time (days)

Pax5$^{+/+}$ n = 61
Pax5$^{+/-}$ n = 19
ca*Stat5a* n = 44
Pax5$^{Etv6/+}$ n = 25
Pax5$^{Foxp1/+}$ n = 16

ca*Stat5a Pax5*$^{+/-}$ n = 43
ca*Stat5a Pax5*$^{Etv6/+}$ n = 32
ca*Stat5a Pax5*$^{Foxp1/+}$ n = 26

B *Pax5*$^{Etv6/+}$ ca*Stat5a* — Tumor / WT; *Pax5*$^{Foxp1/+}$ ca*Stat5a* — Tumor / WT; scale bars 1 cm

C *Pax5*$^{Foxp1/+}$ ca*Stat5a* (1 2 3 4); *Pax5*$^{Etv6/+}$ ca*Stat5a* (1 2 3); WT
V_HJ558; V_HJ7183; V_κ-J$_\kappa$; C_μ

D WT; ca*Stat5a Pax5*$^{Etv6/+}$ Tumor; ca*Stat5a Pax5*$^{Foxp1/+}$ Tumor
SSC vs FSC (size); IL7-Rα vs CD19; CD93 vs Igμ

E Lung / Liver — WT; Tumor ca*Stat5a Pax5*$^{Etv6/+}$; Tumor ca*Stat5a Pax5*$^{Foxp1/+}$; scale bars 0.2 mm

F *PAX5-ETV6* translocation
CDKN2A/B — Genetic material lost
CDKN2A/B ↑ PAX5; ETV6 ↓; Fusion
Chr.9 Chr.12; PAX5-ETV6 ↑ — Two centromeres; dic(9;12)(p13;p13)

PAX5-FOXP1 translocation
CDKN2A/B; CDKN2A/B ↑ PAX5; FOXP1 ↑; Fusion
Chr.9 Chr.3; PAX5-FOXP1 ↑; CDKN2A/B FOXP1-PAX5; t(3;9)(p13;p13)

G Peripheral blood — *Pax5*$^{Etv6/+}$ control; *Cdkn2ab*$^{+/-}$ other tumor; *Cdkn2ab*$^{+/-}$ *Pax5*$^{Etv6/+}$ B-ALL; scale bars 20 µm

Appendix Figure S5. Inactivation of one *Pax5* allele, but not expression of the Pax5 fusion protein, is required for leukemogenesis in ca*Stat5a Pax5*$^{Etv6/+}$ and ca*Stat5a Pax5*$^{Foxp1/+}$ mice.

(A) Kaplan-Meier survival analysis of mice of the indicated genotypes. The number (n) of mice analyzed is shown. (B) Enlarged spleen and lymph nodes in ca*Stat5a Pax5*$^{Etv6/+}$ and ca*Stat5a Pax5*$^{Foxp1/+}$ tumor mice compare to wild-type (WT) mice. (C) PCR analysis of V_HJ558-DJ$_H$, V_H7183-DJ$_H$ and V_κ-J$_\kappa$ rearrangements in lymph node tumors of the indicated genotypes. V(D)J rearrangements to all J$_H$ and J$_\kappa$ segments were detected in WT splenic B cells. (D) Flow cytometric analysis of lymph node cells from a WT mouse as well as ca*Stat5a Pax5*$^{Etv6/+}$ and ca*Stat5a Pax5*$^{Foxp1/+}$ tumor mice. (E) Eosin-hematoxylin-stained sections of the lung and liver of a ca*Stat5a Pax5*$^{Etv6/+}$ and ca*Stat5a Pax5*$^{Foxp1/+}$ tumor mouse, revealing infiltration of these organs by leukemic cells in contrast to a WT mouse. (F) Schematic diagram describing the generation of the human *PAX5-ETV6* and *PAX5-FOXP1* translocations. Arrows indicated the direction of transcription at the *PAX5*, *ETV6* and *FOXP1*. (G) May-Grünwald-Giemsa staining of peripheral blood smears from a control *Pax5*$^{Etv6/+}$ mouse, a *Cdkn2ab*$^{+/-}$ mouse with a non-lymphoid tumor and a *Cdkn2ab*$^{+/-}$ *Pax5*$^{Etv6/+}$ mouse with a B-ALL tumor.
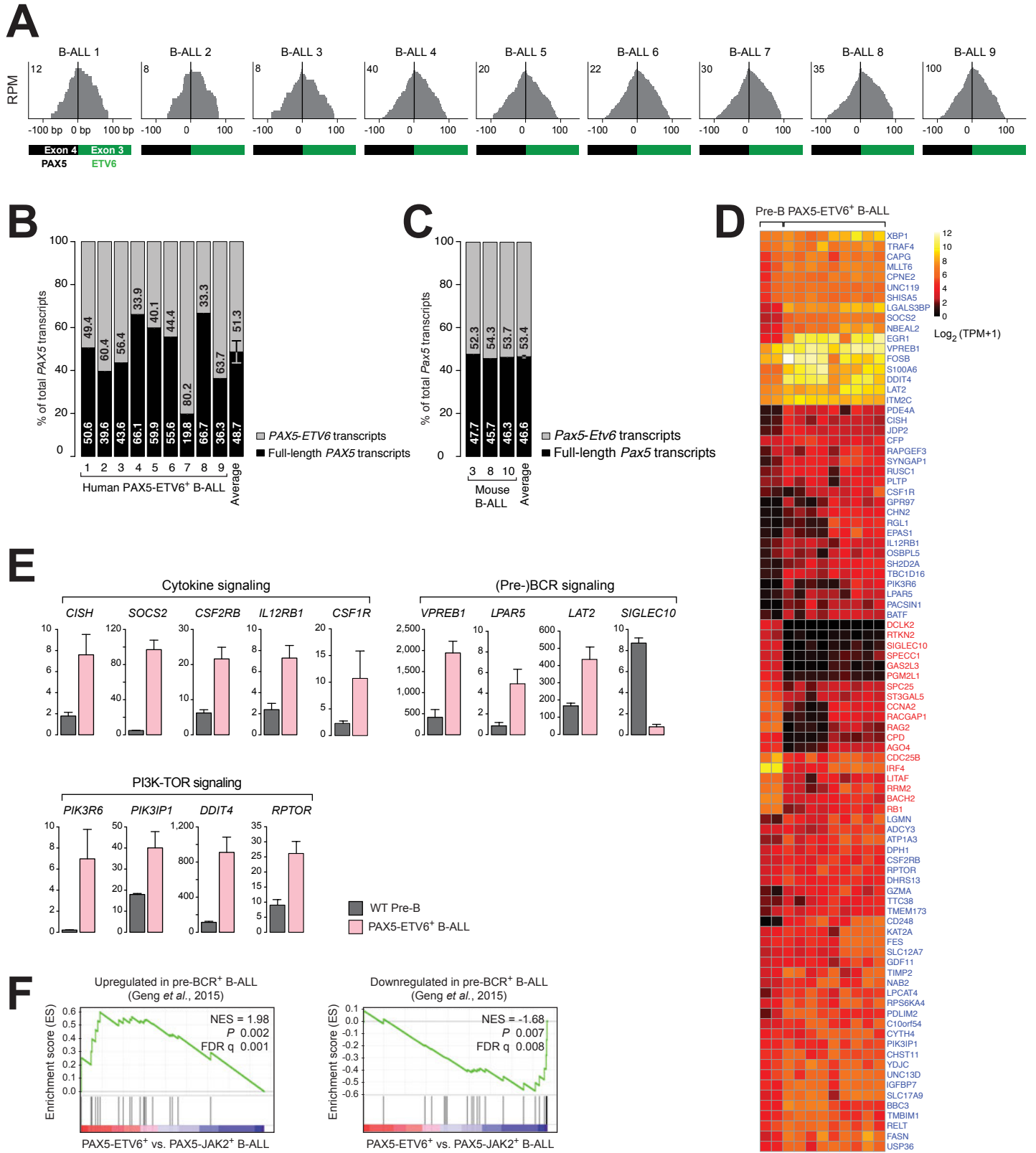
Appendix Figure S6. Molecular function of the Pax5-Etv6 oncoprotein in *Cdkn2ab$^{+/-}$ Pax5$^{Etv6/+}$* **B-ALLs.**

(A) Binding of Pax5-Etv6 at the indicated genes in *Pax5$^{Etv6/+}$ Rosa26$^{BirA/+}$* pro-B cells and *Cdkn2ab$^{+/-}$ Pax5$^{Etv6/+}$ Rosa26$^{BirA/+}$* B-ALL cells (Tu-43), as determined by Bio-ChIP-seq. Horizontal bars below the ChIP-seq tracks indicate binding regions identified by MACS peak calling. The exon-intron structure and a scale bar in kilobases (kb) are shown. (B) Expression of the genes, shown in (A), in wild-type large pre-B cells (black) and *Cdkn2ab$^{+/-}$ Pax5$^{Etv6/+}$* B-ALL cells (red). Gene expression is shown as normalized expression value (TPM) with SEM, based on two (large pre-B) and three (B-ALL) independent RNA-seq experiments. (C) Functional classification and quantification (numbers) of the proteins encoded by > 5-fold activated and repressed Pax5-Etv6 target genes in *Cdkn2ab$^{+/-}$ Pax5$^{Etv6/+}$* B-ALLs. (D) Regulated Pax5-Etv6 target genes coding for cell surface receptors and signal transducers. The indicated target genes were ranked according to their fold expression changes observed between *Cdkn2ab$^{+/-}$ Pax5$^{Etv6/+}$* B-ALL cells and wild-type large pre-B cells (Figure 6I). The color code refers to > 10-fold (green), 7-10-fold (blue) and 5-7-fold (black).

Appendix Figure S7. Molecular characterization of human PAX5-ETV6$^+$ B-ALLs.

(A) Detection of the *PAX5-ETV6* translocation in all 9 B-ALL samples. The frequency of RNA-seq reads, which spanned the junction between PAX5 exon 4 and ETV6 exon 3 in the different PAX5-ETV6$^+$ B-ALL samples, were plotted as normalized RPM values from -100 bp to + 100 bp relative to exon junction. (B,C) Relative abundance of the full-length *PAX5* and *PAX5-ETV6* fusion transcripts in 9 human PAX5-ETV6$^+$ B-ALLs (B) and 3 mouse *Cdkn2ab$^{+/-}$ Pax5$^{Etv6/+}$* ALLs (C). The RNA-seq data of these B-ALLs were used to calculate the percentage of each transcript relative to the total *PAX5* transcripts as described in the Appendix Supplementary Methods. (D) Heat map displaying differentially expressed genes in normal human pre-B cells (n = 2) and the 9 human PAX5-ETV6$^+$ B-ALLs, which were analyzed by hierarchical clustering. The 71 commonly activated (blue) and 19 repressed (red) Pax5-Etv6 target genes in mouse *Cdkn2ab$^{+/-}$ Pax5$^{Etv6/+}$* ALLs and human PAX5-ETV6$^+$ B-ALLs (Fig. 7D) are shown. The log$_2$ [TPM+1] expression value of each gene is visualized according to the indicated scale. The pathway annotation of individual genes is shown in Fig. 7E. (E) Expression of the indicated pathway genes in human PAX5-ETV6$^+$ B-ALLs (rosa) and pre-B cells of healthy donors (gray). Gene expression is shown as normalized expression value (TPM) with SEM, based on two (pre-B) and nine (B-ALL) independent RNA-seq experiments. (F) GSEA analysis of the 19 upregulated (left) and 21 downregulated (right) signature genes in pre-BCR$^+$ B-ALLs compared to pre-BCR$^-$ B-ALLs (Geng et al., *Cancer Cell* **27:** 409-425). These genes were compared to the ranked log$_2$-fold gene expression changes in human PAX5-ETV6$^+$ B-ALLs (n = 9) versus human PAX5-JAK2$^+$ B-ALLs (n = 9). NES, normalized enrichment score; FDR, false discovery rate. For further description, see Appendix Supplementary Methods.