# An intact gut microbiome protects genetically predisposed mice against leukemia

*Vicente-Dueñas*, Janssen*, Oldenburg* et al.*

## SUPPLEMENTAL INFORMATION

## CONTENTS:

## METHODS

## SUPPLEMENTAL DATA TABLES:

**Supplemental Table 1.** Gene list of differentially expressed genes in bone marrow leukemic cells from *Pax5*$^{+/-}$ mice treated with antibiotics compared to wild-type bone marrow proB and preB cells. (7032 genes-probe sets with FDR = 0.01). Provided as an Excel file.

**Supplemental Table 2.** Whole exome sequencing results from leukemic *Pax5*$^{+/-}$ mice treated with antibiotics.

**Supplemental Table 3.** Gene list of differentially expressed genes in bone marrow leukemic cells from *Pax5*$^{+/-}$ mice treated with antibiotics compared to tumoral bone marrow cells from untreated *Pax5*$^{+/-}$ mice. (2119 genes-probe sets with FDR = 0.01). Provided as an Excel file.

**Supplemental Table 4.** Cell marker schemes used for flow cytometric identification of cell populations.

**SUPPLEMENTAL FIGURES:**

**Supplemental Figure 1.** Source tracking of microbial compositions of genotype-mixed cages.

**Supplemental Figure 2.** Taxonomic composition of specific pathogen free (SPF) and conventional (CF) housing facilities across time.

**Supplemental Figure 3.** B cells in Peyer's patches are reduced in *Pax5*$^{+/-}$ mice.

**Supplemental Figure 4.** Experimental flow chart of the *in vivo* study design.

**Supplemental Figure 5.** Experimental design of the "abx" cohort.

**Supplemental Figure 6.** ASVs differentially abundant between WT and *Pax5*$^{+/-}$ DNA stool samples.

**Supplemental Figure 7.** Antibiotics alter the gut microbiota composition**.**

**Supplemental Figure 8.** Impact of removing mitochondrial and chloroplast raw reads.

**Supplemental Figure 9.** Peripheral blood FACS analysis of *Pax5*$^{+/-}$ mice after antibiotic treatment.

**Supplemental Figure 10.** Bone marrow FACS analysis of *Pax5*$^{+/-}$ mice after antibiotic treatment.

**Supplemental Figure 11.** Spleen FACS analysis of *Pax5*$^{+/-}$ mice after antibiotic treatment.

**Supplemental Figure 12.** Lymph node FACS analysis of *Pax5*$^{+/-}$ mice after antibiotic treatment.

**Supplemental Figure 13.** Thymus flow cytometric analysis of *Pax5*$^{+/-}$ mice after

antibiotic treatment.

**Supplemental Figure 14.** Reduction of peripheral blood B cells in *Pax5*$^{+/-}$ mice treated with antibiotics.

**Supplemental Figure 15.** pB-ALL development in *Pax5*$^{+/-}$ mice treated with antibiotics and exposed to infectious conditions.

**Supplemental Figure 16.** pB-ALL in *Pax5*$^{+/-}$ mice after depletion of gut commensal bacteria.

**Supplemental Figure 17.** Analysis of BCR clonality of leukemias arising in *Pax5*$^{+/-}$ mice treated with antibiotics.

**Supplemental Figure 18.** Gene expression analysis of leukemic *Pax5*$^{+/-}$ pB-ALL with altered gut microbiome.

**Supplemental Figure 19.** Number of called somatic mutations in *Pax5*$^{+/-}$ pB-ALL of mice treated with antibiotics.

**Supplemental Figure 20.** Antibiotics treatment alters the gut microbiota composition in control WT mice.

**Supplemental Figure 21.** WT mice treated with antibiotics develop B cell neoplasias.

**Supplemental Figure 22.** B cell neoplasia in WT mice following depletion of gut commensal bacteria.

**Supplemental Figure 23.** Histological images of B cell neoplasia in WT mice following depletion of gut commensal bacteria.

**Supplemental Figure 24.** BCR clonality in antibiotic-treated WT mice treated with antibiotics that developed with B cell neoplasias

**Supplemental Figure 25.** Numbers of T cells in Peyer's patches are decreased following depletion of gut commensal bacteria by antibiotic treatment.

**Supplemental Figure 26.** pB-ALL development in *Pax5*$^{+/-}$;*nu/nu* mice.

**Supplemental Figure 27.** FACS analysis of pB-ALL in *Pax5*$^{+/-}$*;nu/nu* mice.

**Supplemental Figure 28.** Treatment with antibiotics alters the gut microbiota composition in *Pax5*$^{+/-}$ mice housed in SPF.

**Supplemental Figure 29.** Bacterial depletion of the gut promotes pB-ALL development in genetically predisposed mice in the absence of natural infectious stimuli.

**Supplemental Figure 30.** Gating strategy used in FACS analyses.

## METHODS

### Generation of mouse strains

Heterozygous *Pax5*$^{+/-}$ mice as well as *Sca1-ETV6-RUNX1* mice used in the study have been described previously[1] [2]. *Pax5*$^{+/-}$ and control *Pax5*$^{+/+}$ wild-type (WT) mice were born and kept in the SPF facility until exposure to an infectious environment as previously described.[3] The health status of all mice included in the study was monitored from November-2015 to October-2019 by serological analysis to determine the microorganisms they were exposed to. All animal work was conducted according to relevant national and international guidelines and has been approved by the Bioethics Committee of University of Salamanca and by the Bioethics Subcommittee of Consejo Superior de Investigaciones Científicas (CSIC). In order to avoid gender-biases in microbiome results, only female WT, *Pax5*$^{+/-}$ and *Sca1-ETV6-RUNX1* mice of a mixed C57BL/6 x CBA background were included in the study. We used WT and *Pax5*$^{+/-}$ littermates of the same breeding. In order to generate *Pax5*$^{+/-}$*;nu/nu* and *Pax5*$^{+/-}$*;nu/*+ mice, we crossed *Pax5*$^{+/-}$ mice with thymus-deficient *nu/nu* mice (Crl:NU-Foxn1$^{nu}$ ; Charles River). *Pax5*$^{+/-}$*;nu/nu (n=33), Pax5*$^{+/-}$*;nu/*+ (n=30), *nu/nu* (n=31) and control WT (n=20) mice were born and kept in the SPF facility until exposed to an infectious environment. Upon signs of disease, all mice in the study were sacrificed and subjected to standard necropsy procedures. All major organs were examined under the dissecting microscope. Tissue samples were taken from homogenous portions of the resected organ and fixed immediately after excision. Differences in Kaplan-Meier survival plots of transgenic and WT mice were analysed using the log-rank (Mantel-Cox) test. Statistical analyses were performed by using GraphPad Prism v5.01 (GraphPad Software).

**Antibiotics treatment**

Mice were treated with an antibiotic regimen as previously described.[4] Groups of mice were given a cocktail of antibiotics (ampicillin, 1 g/L, Ratiopharm; vancomycin, 500 mg/L, Cell Pharm; ciprofloxacin, 200 mg/L, Bayer Vital; imipenem, 250 mg/L, MSD; metronidazole, 1 g/L, Fresenius) into their drinking water *ad libitum* for a period of eight weeks.

**Bone marrow transplantation (BMT) and stool sample collection.**

Recipient female C57BL/6J mice (8–12 weeks old) were irradiated with two split doses of 600 cGy 2 hr apart. This dose is sufficient to eliminate endogenous hematopoiesis completely. *Pax5*$^{+/-}$ and WT bone marrow (BM) cells were injected into the tail vein of the irradiated mice at 2–4 × $10^6$ cells per mouse for hematopoietic reconstitution. All recipients were maintained in microisolator cages on sterilized food and acidified sterile water. Stool samples were collected at defined time points after BMT for microbiome analysis.

**FACS analysis**

Nucleated cells were obtained from total mouse bone marrow (flushing from the long bones), peripheral blood, thymus, or spleen. Contaminating red blood cells were lysed with RCLB lysis buffer and the remaining cells were washed in PBS with 1% FCS. After staining, all cells were washed once in PBS and were resuspended in PBS with 1% FCS containing 10 µg/mL propidium iodide (PI) to exclude dead cells from both analyses and sorting procedures. The samples and the data were acquired in an AccuriC6 Flow Cytometer and analysed using Flowjo software. Specific fluorescence of FITC, PE, PI and APC excited at 488

nm (0.4 W) and 633 nm (30 mW), respectively, as well as known forward and orthogonal light scattering properties of mouse cells were used to establish gates. Nonspecific antibody binding was suppressed by preincubation of cells with CD16/CD32 Fc-block solution (cat. 553142) from BD Biosciences. For each analysis, at least 50,000 viable (PI-) cells were assessed.

The antibodies used to characterize the hematopoietic compartments as well as to detect leukemia development are listed in **Supplemental Table 4**. All antibodies were purchased from BioLegend and used at a 1:100 dilutions unless otherwise indicated. The gating strategy is shown in **Supplemental Figure 30**.

**Preparation of serum samples for metabolome analyses**

Serum samples from preleukemic *Pax5*$^{+/-}$ (n=10), leukemic *Pax5*$^{+/-}$ (n=10) and healthy *WT* mice (n=10) were prepared as follows: whole blood was collected and allowed to clot at room temperature (15-30 min) followed by centrifugation at 1,000–2,000 x g for 10 minutes. The samples were stored at –70°C.

**Metabolite extraction**

Metabolite extraction was conducted according to Barupal et al. 2019[5] with minor modifications. Briefly, extraction solution (acetonitrile (ACN) / isopropanol (IPA) / water (H2O) (3:3:2, v/v/v)) was prepared and cooled to −20 °C. 1 ml of extraction solution was mixed with 30 µl of 25 µM internal standard (ISTD) solution (ribitol and N,N-dimethylphenylalanine) per sample. Serum samples were then thawed on ice and 20 µl of sample was added to the extraction/ISTD solution. Samples were vortexed for 10 seconds, shaken for 5 min and

centrifuged for 2 min at 14,000 rcf at 4°C. Next, two 450 µL supernatant aliquots were transferred to new tubes and 500 µl of an ice-cold ACN / water (50:50, v/v) solution was added to remove any excess protein. After additional centrifugation for 2 min at 14,000 rcf, the supernatant was transferred to a pre-cooled tube and dried via vacuum centrifugation.

**Gas chromatography - Mass Spectrometry (GC-MS) analysis**

Dried samples were reconstituted in 150 µl of the extraction solution, transferred into a glass vial and dried again via vacuum centrifugation. Each sample was derivatized with methoxyamine hydrochloride and N-methyl-N-(trimethylsilyl)trifluoroacetamide as described in Gu et al., 2012[6]. After incubation for two hours at room temperature, 1 mL was injected into a GC-MS system (7890A GC and a 5977B MSD both Agilent Technologies) and chromatography was performed as described in Shim et al., 2019[7]. Identification of metabolites was performed by two different methods. A quality control (QC) sample containing a mixture of target compounds was used as a reference to identify target compounds in the sample based on mass spectra similarity and retention time (annotation level: reference). The AMDIS software ((http://chemdata.nist.gov/mass-spc/amdis/) v. 2.72, 2014) was used for deconvolution of mass spectra of target peaks before comparing spectra to the NIST14 Mass Spectral Library (https://www.nist.gov/srd/nist-standard-reference-database-1a-v14). Matches with more than 80% mass spectra similarity were assigned accordingly (annotation level: NIST match). Peaks were integrated using the software MassHunter Quantitative (v b08.00, Agilent

Technologies). For relative quantification, metabolite peak areas were normalized to the peak area of the internal standard ribitol.

**Histology**

Tissue samples were formalin-fixed and embedded in paraffin. Pathology assessment was performed on hematoxylin-eosin stained sections under the supervision of Dr. Oscar Blanco, an expert pathologist at the Salamanca University Hospital.

**V(D)J recombination**

Immunoglobulin rearrangements were amplified by PCR using the primers shown below. Cycling conditions consisted of an initial heat-activation at 95ºC followed by 31-37 cycles of denaturation for 1 min at 95ºC, annealing for 1 min at 65ºC, and elongation for 1 min 45 s at 72ºC. This was followed by a final elongation for 10 min at 72ºC. To determine the DNA sequences of individual V(D)J rearrangements, the PCR fragments were isolated from the agarose gel and cloned into the pGEM-Teasy vector (Promega); the DNA inserts of at least ten clones corresponding to the same PCR fragment were then sequenced. The following primer pairs were used:

| | | |
|---|---|---|
| $V_H$J558 | forward | CGAGCTCTCCARCACAGCCTWCATGCARCTCARC |
| | reverse | GTCTAGATTCTCACAAGAGTCCGATAGACCCTGG |
| $V_H$7183 | forward | CGGTACCAAGAASAMCCTGTWCCTGCAAATGASC |
| | reverse | GTCTAGATTCTCACAAGAGTCCGATAGACCCTGG |
| $V_H$Q52 | forward | CGGTACCAGACTGARCATCASCAAGGACAAYTCC |
| | reverse | GTCTAGATTCTCACAAGAGTCCGATAGACCCTGG |
| DH | forward | TTCAAAGCACAATGCCTGGCT |
| | reverse | GTCTAGATTCTCACAAGAGTCCGATAGACCCTGG |
| Cμ | forward | TGGCCATGGGCTGCCTAGCCCGGGACTT |
| | reverse | GCCTGACTGAGCTCACACAAGGAGGA |

**Microarray Data Analysis**

Total RNA was isolated in two steps using TRIzol (Life Technologies) followed by RNeasy Mini-Kit (Qiagen) purification according to the manufacturer's RNA Clean-up Protocol including the optional On-Column DNase treatment. The integrity and the quality of the RNA were verified by electrophoresis, and its concentration was measured. Samples were analysed using Affymetrix Mouse Gene 1.0 ST arrays.

Briefly, the robust microarray analysis algorithm was used for background correction, intra- and inter-microarray normalization, and expression signal calculation.[8-10] Once the absolute expression signal for each gene (i.e., the signal value for each probe set) was calculated in each microarray, a method called significance analysis of microarray (SAM)[11] was applied to calculate significant differential expression and find the genes that characterized the tumor-bearing BM from *Pax5*$^{+/-}$ mice compared with BM-derived B220$^{+}$ cells from WT mice. The method uses permutations to provide a robust statistical inference of the most significant genes and provides *P* values adjusted to multiple testing using FDR.[12] A cutoff of FDR < 0.05 was used for the differential expression calculations. We applied all these methods using R[13] and Bioconductor.[14]Gene Set Enrichment Analyses were performed using GSEAv2.2.2[15,16] and hallmark collection of gene sets.[17] Gene expression signatures that were specifically upregulated or downregulated in human B-ALL[18-20] were also tested for enrichment within tumor specimens using GSEA.

**Mouse exome library preparation and next generation sequencing**

The AllPrep DNA/RNA Mini Kit (Qiagen) was used to purify DNA according to the manufacturer's instructions. Exome library preparation was performed using the Agilent SureSelectXT Mouse All Exon v1 kit with modifications. Targeted capture by hybridization to an RNA library was performed according to the manufacturer's protocol. Purification and enrichment of the captured library was achieved by binding to MyOne Streptavidin T1 Dynabeads (LifeTechnolgies) and off-bead PCR amplification in the linear range. 2x100 bp sequencing with an 8 bp index read was performed using the TruSeq SBS Kit v3 on the HiSeq 2500 (Illumina).

For data analysis, Fastq files were generated by using bcl2fastq 2.19.0.316 (Illumina). BWA 0.7.12. was used to align sequence data to the mouse reference genome (GRCm38.71). Conversion steps were carried out using Samtools 1.3.2 followed by removal of duplicate reads Picard tools 2.0.1. Local realignment around indels, SNP-calling, annotation and recalibration was facilitated by GATK 3.5.0. Mouse dbSNP138 and dbSNP for the used mouse strains were used as training datasets for recalibration. Resulting variation calls were annotated with Variant Effect Predictor using the Ensembl database (v70) and imported into an in-house SQL database to facilitate automatic and manual annotation, reconciliation and data analysis via complex database queries. Loss of function prediction scores for PolyPhen2 and SIFT were extracted from this Ensemble release.

Somatic calls were produced using MuTect 1.1.7 and VarScan2 2.4.3. As suggested by the author of VarScan2, false-positive filtering was applied. To further clear the resulting file of suspicious results, only entries with at least 9%

difference in allele frequency between tumor and normal were kept for further analysis. Cancer-related genes were determined by translating the cancer gene consensus from COSMIC using ENSEMBL's BioMart.

**16S rRNA sequencing and analysis**

Fresh fecal samples were collected from each mouse (WT or *Pax5*$^{+/-}$ mice) at time points indicated in **Supplemental Figures 5 and 6.** Fecal DNA was isolated using QIAamp DNA Stool Mini Kit (Qiagen) using the stool pathogen detection protocol. DNA was extracted following the manufacturer's instructions with a modification: heating step in ASL buffer at 95°C and eluting the DNA in 100µl AE buffer.

*Targeted 16S V4 region sequencing:* Following the Earth Microbiome Project protocols[21,22] we targeted and amplified the V4 region of the 16S rRNA gene by PCR using barcoded primers.[23] V4 paired-end sequencing[23] was performed using an Illumina MiSeq (La Jolla, CA) according to manufacturer's protocols.

*Full-length 16S amplicon sequencing:* In addition to targeted V4 sequencing, we performed long-read full-length 16S rRNA gene sequencing on the same samples. To this end, one round of PCR was performed with 16S-specific primers concatenated with sample-specific barcode sequences (Pacific Biosciences). Both primers for each sample were tailed with the same barcode sequence. 2.5 ng of fecal DNA were used for PCR employing the KAPA HiFi HotStart ReadyMix PCR Kit (KAPA Biosystems, 3 µM each primer, cycling program: 95°C for 3 min followed by 23 cycles of amplification (95°C for 20 s, 57°C for 30 s and 72°C for 30 s), followed by final extension at 72°C for 1 min). PCR products were purified with 0.6X AMPure PB beads (Pacific Biosciences)

on a SciClone G3 (Perkin Elmer) and eluted in 38 µl 1x Tris-HCl (pH 8.5). PCR products were diluted 1:10 and used for post-amplification quality control performed on a Fragment Analyser (Advanced Analytical) using the HS NGS Fragment kit (Agilent). PCR products were pooled in 10 equimolar samples with ~100 ng per sample. Each pool included positive (mock community, ATCC MSA-1003) and negative control (no DNA). DNA pools were concentrated with 1X AMPure PB beads to 48 µl and DNA concentrations (1:10 dilution) were measured with Qubit dsDNA HS Kit (Thermo Fischer). Library preparation was performed using the Express Template Prep Kit 2.0 (Pacific Biosciences) according to the manufacturer's instructions. Sequencing primers v4 (Pacific Biosciences) were annealed to 20 ng of each library at a molar ratio of 20:1 (primer:library template) at 20°C for 1 h, and polymerase 3.0 was bound to the libraries at a molar ratio of 30:1 (polymerase:library template) at 30°C for 1 h. Polymerase-bound templates were purified with 1X AMPure PB beads and DNA concentrations were measured with Qubit dsDNA HS Kit (Thermo Fischer). Sequencing was performed on a PacBio Sequel System with the sequencing chemistry 3.0. Each pool (7 pM) was loaded on one SMRT cell and sequenced with 60 min pre-extension and 600 min collection time. Circular consensus reads were generated and de-multiplexing was performed with SMRT Link 7.0.

*Sequencing data processing and quality control*: Raw V4 sequence reads were demultiplexed using Illumina's bcl2fastq 2.19.0.316 software. Primers were trimmed via cutadapt 1.18,[24] uploaded to Qiita[25] and quality controlled using the defaults. Forward reads were trimmed to the first 150 nucleotides. The primary feature table was generated using Deblur 1.1.0[26] and can be found in Qiita (qiita.ucsd.edu) as studies 11758, 11953, 12981 and 13189 with artefacts

67817, 75878, 86044 and 90669 respectively. Sequences can additionally be found in EBI under accession numbers ERP117561, ERP117668, ERP122650 and ERP122653, respectively. Taxonomy assignment was done via Qiime2's[27] "feature-classifier" plugin version 2020.2 against Greengenes 13_8 99% OTUs from 515F/806R region of sequences. Bacterial features with assigned taxonomy containing the labels c__Chloroplast or f__mitochondria were considered of host or plant origin and removed from the feature table prior to rarefaction. Differential abundance analysis was performed via discrete FDR[28] as implemented in calour.[29] PacBio data were transformed into a feature table via DADA2[30] as described.[31] Processing was performed analogously to the V4 sequences with following exceptions: Taxonomy assignment was done against Greengenes 13_8 99% OTUs full-length sequences. Rarefaction depth was 12,000. Phylogeny was created by aligning sequences via Mafft 7.310[32] and de-novo tree construction via Fasttree 2.1.10.[33]

*Statistical analysis:* In order to compute phylogenetic diversity distances, unique V4 sequence fragments were inserted into the reference phylogenetic tree of Greengenes 13.8 99%.[34] The resulting tree was used to compute weighted and unweighted UniFrac[35] beta distances and Faith's PD alpha diversity. Rarefaction curves for all samples were computed via Faith's PD and an optimal rarefaction depth of 35,000 reads per sample was chosen manually, resulting in 657 samples with suitable read counts. Bacterial features with less than 10 reads across all 657 samples were omitted, resulting in 3983 bacterial features for downstream analyses. Differences in community composition for groups of samples (i.e. beta diversity) were assessed via pairwise Permanova[36] tests with 999 permutations as implemented in scikit-bio 0.5.5.

scikit-learn 0.22.1[37] was used for Random Forest analyses. Comparison of alpha diversity values and changes in beta diversity was done using Mann-Whitney tests as implemented in scipy 1.4.1.[38] Jupyter notebooks for data analysis are available as supplemental material.

**Data Availability**

Authors can confirm that all relevant data are included in the paper and/or its supplemental information files. The gene expression data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (GEO)[39] and are accessible through GEO Series accession numbers: GSE62529 (for pB-ALL from *Pax5*$^{+/-}$ non-treated mice) and GSE139547 (for pB-ALL from *Pax5*$^{+/-}$ antibiotics-treated mice).

All sequence data related to this study are available upon request from the European Nucleotide Archive (ENA) under the EBI accession numbers: ERP117561, ERP117668, ERP122650 and ERP122653.

## SUPPLEMENTAL DATA TABLES:

**Supplemental Table 1.** Gene list of differentially expressed genes in bone marrow leukemic cells from *Pax5*$^{+/-}$ mice treated with antibiotics compared to wild-type bone marrow proB and preB-cells. (7032 genes-probe sets with FDR = 0.01). Provided as an Excel file.

**Supplemental Table 2.** Whole exome sequencing results from leukemic *Pax5*$^{+/-}$ mice treated with antibiotics.

| Mice cohort | Mouse number | Mouse age at disease (months) | SEX | Antibiotic treatment | Housing conditions | Mutations identified by WES | | | |
|---|---|---|---|---|---|---|---|---|---|
| Pax5$^{+/-}$ (abx-CF) | L220 | 8,77 | Female | YES | CF | Ptpn11 (p:Glu69Lys) | | | |
| | L222 | 9,30 | Female | YES | CF | Kit (p.Asp822Gly) | | | |
| | L203 | 10,03 | Female | YES | CF | | | | |
| | L643 | 8,30 | Female | YES | CF | Ptpn11 (p.Ser506Leu) | | | |
| | L452 | 12,30 | Female | YES | CF | Cbl (p.Cys402Phe) | | | |
| | L712 | 10,10 | Female | YES | CF | Ptpn11 (p.Gly60Arg) | Jak1 (p:Ser645Phe) | Jak3 (p.Ala568Val) | |
| | L713 | 10,07 | Female | YES | CF | Kit (p.Val562Asp) | Cbl (p.Gln365Pro) | | |
| | L980 | 9,40 | Female | YES | CF | Flt3 (p.Leu840Pro) | Flt3 (p.Asp838Val) | Flt3 (p.Asp842insGly) | |
| | I338 | 7,53 | Female | YES | CF | | | | |
| | L832 | 11,13 | Female | YES | CF | Ptpn11 (p.Ser506Pro) | Flt3 (p.Asp838Glu) | | |
| | L245 | 15,50 | Female | YES | CF | | | | |
| | L380 | 14,73 | Female | YES | CF | Ptpn11 (p:Glu69Lys) | Jak3 (p.Arg653His) | | |
| | L381 | 14,97 | Female | YES | CF | Flt3 (p.Asp838His) | | | |
| | L181 | 20,00 | Female | YES | CF | Ptpn11 (p.Ser506Leu) | Jak3 (p.Arg653His) | | |
| | I337 | 12,67 | Female | YES | CF | | | | |
| | I331 | 19,50 | Female | YES | CF | | | | |
| | L979 | 21,77 | Female | YES | CF | Jak3 (p.Arg653His) | | | |
| Pax5$^{+/-}$ (abx-SPF) | A955 | 11,47 | Female | YES | SPF | Trp53 (p.Arg270Pro) | Ptpn11 (p.Ser506Leu) | Cbl (p.Gly849Arg) | |
| | A428 | 14,73 | Female | YES | SPF | Pax5 (p.Pro80Arg) | Ptpn11 (p.Ser506Leu) | Wnt1 (p.Ala253Ser) | |
| | A707 | 13,23 | Female | YES | SPF | Pax5 (p.Pro80Arg) | Ptpn11 (p.Ser506Leu) | | |
| | A956 | 12,47 | Female | YES | SPF | Ptpn11 (p.Glu76Lys) | | | |
| | A733 | 14,33 | Female | YES | SPF | Jak3 (p.Arg653His) | Jak1 (p.Leu652Phe) | Jak1 (p.Phe733Leu) | Jak1 (p.Leu909Pro) |
| | A734 | 14,70 | Female | YES | SPF | Ptpn11 (p.Glu69Lys) | Nras (p.Gln61Arg) | | |

Note: CF (Conventional Facility); SPF (Specific Pathogen Free).

**Supplemental Table 3.** Gene list of differentially expressed genes in bone marrow leukemic cells from *Pax5*$^{+/-}$ mice treated with antibiotics compared to tumoral bone marrow cells from untreated *Pax5*$^{+/-}$ mice. (2119 genes-probe sets with FDR = 0.01). Provided as an Excel file.

**Supplemental Table 4.** Cell marker schemes used for flow flow cytometry identification of cell populations.

| ***Peripheral blood cells*** | ***Markers*** |
|---|---|
| $CD4^+$ T cells | $CD4^+$ |
| $CD8^+$ T cells | $CD8^+$ |
| B cells | $B220^+$ $IgM^{+/-}$ |
| Mature B cells | $CD19^+B220^+$ |
| Transitional T1 B cells | T1 ($IgD^{low}$ $IgM^+$) |
| Transitional T2 B cells | T2 ($IgD^+$ $IgM^{+)}$ |
| Mature B cells | M ($IgD^+$ $IgM^{-)}$ |
| Myeloid cells | $CD11b^+$ $Ly6G^+$ |
| Erythroid cells | $TER119^+$ |

| ***Bone marrow cells*** | ***Markers*** |
|---|---|
| $CD4^+$ T cells | $CD4^+$ |
| $CD8^+$ T cells | $CD8^+$ |
| B cells | $B220^+$ $IgM^{+/-}$ |
| Pre and ProB cells | $IgM^-$ $B220^{low}$ |
| Immature B cells | $IgM^+$ $B220^{low}$ |
| Recirculating B cells | $IgM^+$ $B220^{high}$ |
| pro-B cells | $CD19^+$ $cKit^+$ |
| pre-B cells | $CD25^+$ $B220^{low}$ |
| Mature B cells | $CD19^+B220^+$ |
| Transitional T1 B cells | T1 ($IgD^{low}$ $IgM^+$) |
| Transitional T2 B cells | T2 ($IgD^+$ $IgM^+$) |
| Mature B cells | M ($IgD^+$ $IgM^-$) |
| Follicular B cells | $CD23^+$ $CD21^+$ |
| Myeloid cells | $CD11b^+$ $Ly6G^+$ |
| Erythroid cells | $TER119^+$ |
| LSK | Lin- $Sca1^+$ $cKit^+$ |

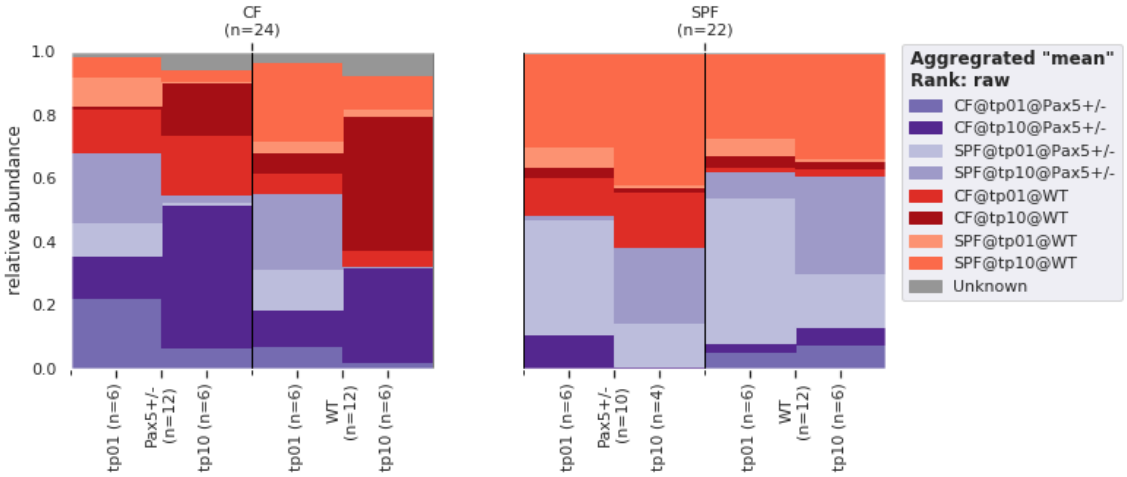
| ***Thymus cells*** | ***Markers*** |
|---|---|
| $CD4^+$ T cells | $CD4^+$ |
| $CD8^+$ T cells | $CD8^+$ |
| DP | $CD4^+$ $CD8^+$ |
| DN | $CD4^-$ $CD8^-$ |

| ***Spleen cells*** | ***Markers*** |
|---|---|
| $CD4^+$ T cells | $CD4^+$ |
| $CD8^+$ T cells | $CD8^+$ |
| B cells | $B220^+$ $IgM^{+/-}$ |
| Mature B cells | $CD19^+B220^+$ |
| Transitional T1 B cells | T1 ($IgD^{low}$ $IgM^+$) |
| Transitional T2 B cells | T2 ($IgD^+$ $IgM^+$) |
| Mature B cells | M ($IgD^+$ $IgM^-$) |
| Follicular B cells | $CD23^+$ $CD21^+$ |
| MZ B cells | CD23- CD21+ |
| Myeloid cells | $CD11b^+$ $Ly6G^+$ |
| Erythroid cells | $TER119^+$ |

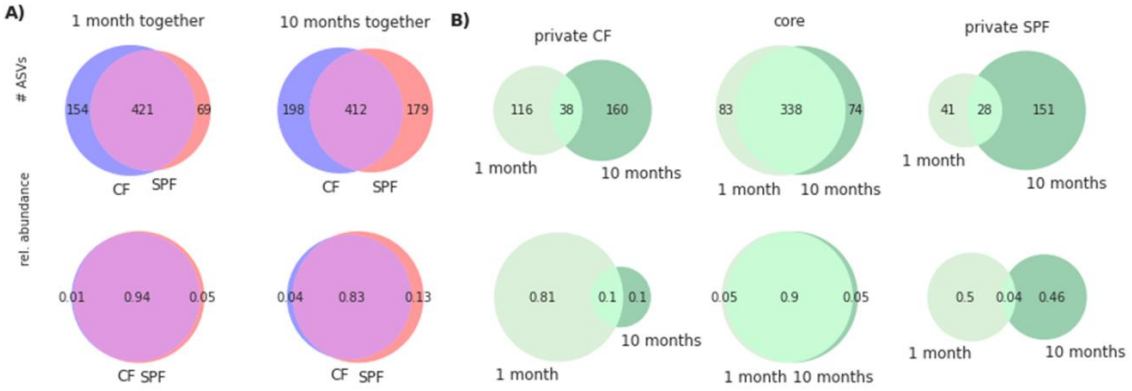
| ***Lymph nodes cells*** | ***Markers*** |
|---|---|
| $CD4^+$ T cells | $CD4^+$ |
| $CD8^+$ T cells | $CD8^+$ |
| B cells | $B220^+$ $IgM^{+/-}$ |
| Mature B cells | $CD19^+B220^+$ |
| Recirculating B cells | $IgD^+$ $B220^+$ |
| Follicular B cells | $CD23^+$ $CD21^+$ |
| Myeloid cells | $CD11b^+$ $Ly6G^+$ |
| Erythroid cells | $TER119^+$ |

| ***Peyer's Patches cells*** | ***Markers*** |
|---|---|
| $CD3^+$ T cells | $CD3^+$ |
| $CD4^+$ T cells | $CD4^+$ |
| $CD8^+$ T cells | $CD8^+$ |
| B cells | $B220^+$ $IgM^{+/-}$ |
| Mature B cells | $CD19^+B220^+$ |
| Myeloid cells | $CD11b^+$ $Ly6G^+$ |

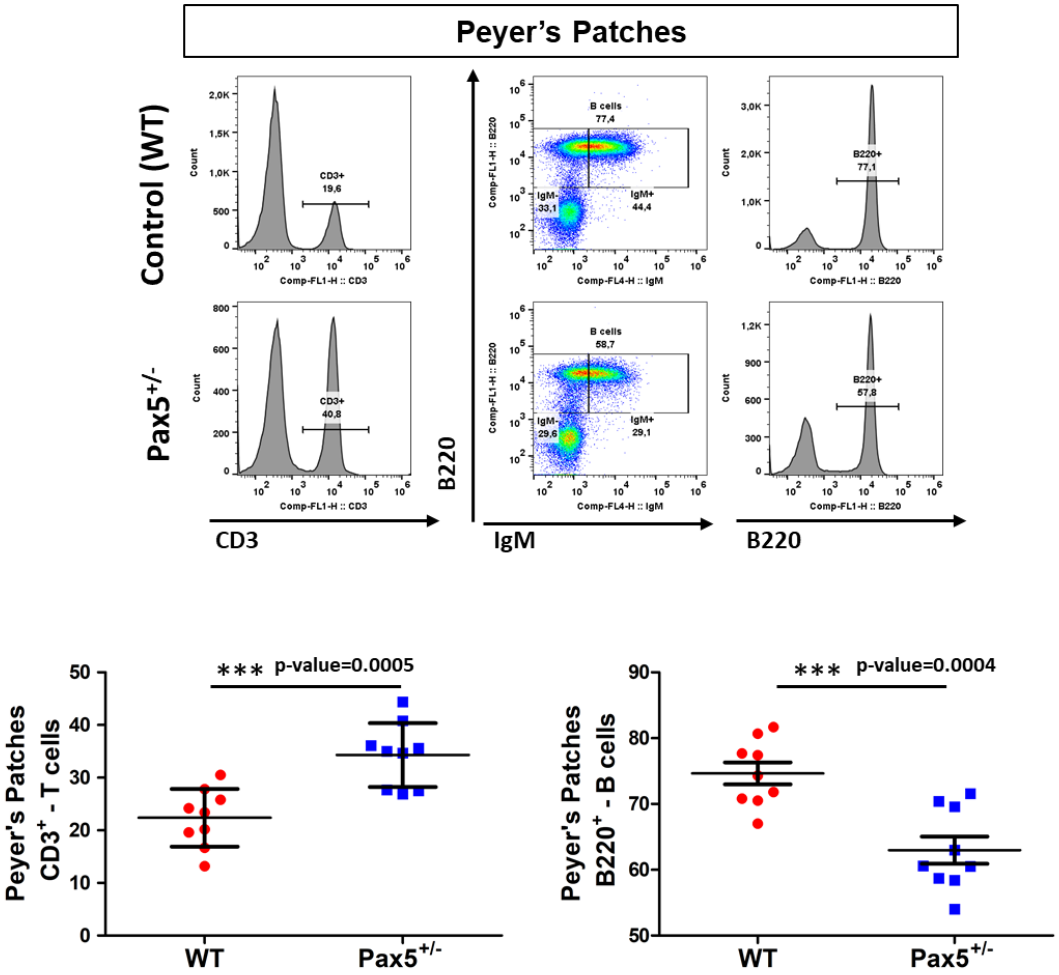
**SUPPLEMENTAL FIGURES:**



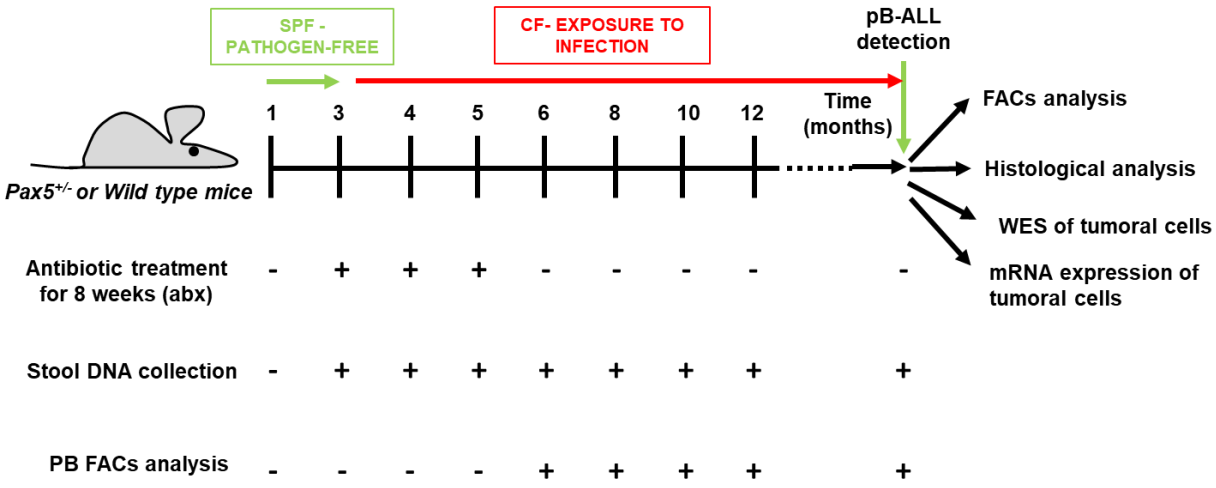

**Supplemental Figure 1.** Source tracking of microbial compositions of genotype-mixed cages**.** Defining all genotype homogeneous samples of the co-housing cohort, stratified by genotype (*Pax5*$^{+/-}$ purple, *WT* red), facility (CF darker shades, SPF lighter shades) and time point (tp01 and tp10), as eight "sources" and the remaining samples as "sinks", we performed a SourceTracker2 analysis to investigate mixing of the microbiota of mice in genotype-mixed heterozygous cages due to their coprophagous behavior. Contributions from individual samples were merged (by computing their mean). Six samples from the mixed-genotype housed *Pax5*$^{+/-}$ mice in CF at the first time point (very first column) show an almost equal mixing of both genotype specific microbiomes, with slight dominance ~60% (compare all reddish to all purplish colors) of *Pax5*$^{+/-}$ "source" microbiomes. This ratio dropped when mice lived 10 months together (second column) to ~50%. Similar trends were observed for WT mice, with a slightly larger dominance of the WT "source" microbiome (third and fourth column). Observations are similar for SPF (right part of the figure), but samples were more dominated by "source" microbiota of the first time point, which is in line with the observation that the change across time for SPF is rather small. CF (Conventional Facility); SPF (Specific Pathogen Free).



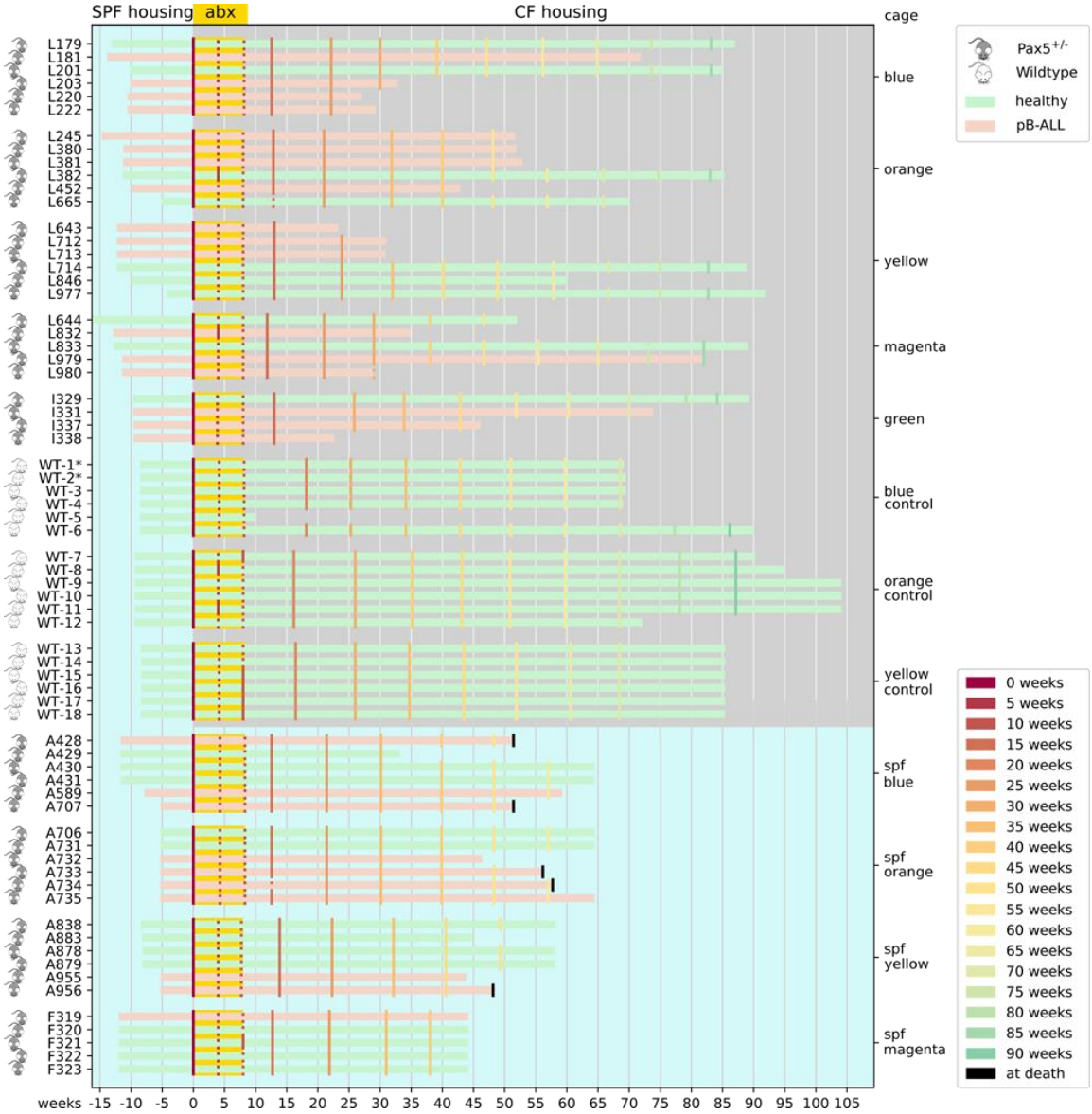

**Supplemental Figure 2. Taxonomic composition of specific pathogen free (SPF) and conventional (CF) housing facilities across time. A)** 421 distinct amplicon sequence variants (ASVs) were shared between CF and SPF at the first time point (top, leftmost Venn diagram). Together, they made up 94% of total relative abundance (bottom, leftmost Venn diagram). We only considered ASVs that were present, (more than one read in the rarefied feature table, in at least 5 samples in each facility). In addition, there were 154 ASVs only found in CF and 69 that are private for SPF. Nine months later, the core set decreased by 14 ASVs and 11% relative abundance (second column) indicating a growing difference according to facility. **B)** Of the 154 and 198 ASVs private to the CF, only 38 were shared across time. They contributed 0.1% or 0.9% of total relative abundance at first and second time points, respectively (first column). Similarly, only 28 ASVs persisted over time as private SPF (last column), with an even lower contribution to the total relative abundance. Majority of the abundance (89.3% and 78.9%) was stable between the first and second time point, respectively (center column) consisting of a stable set of 338 ASVs.



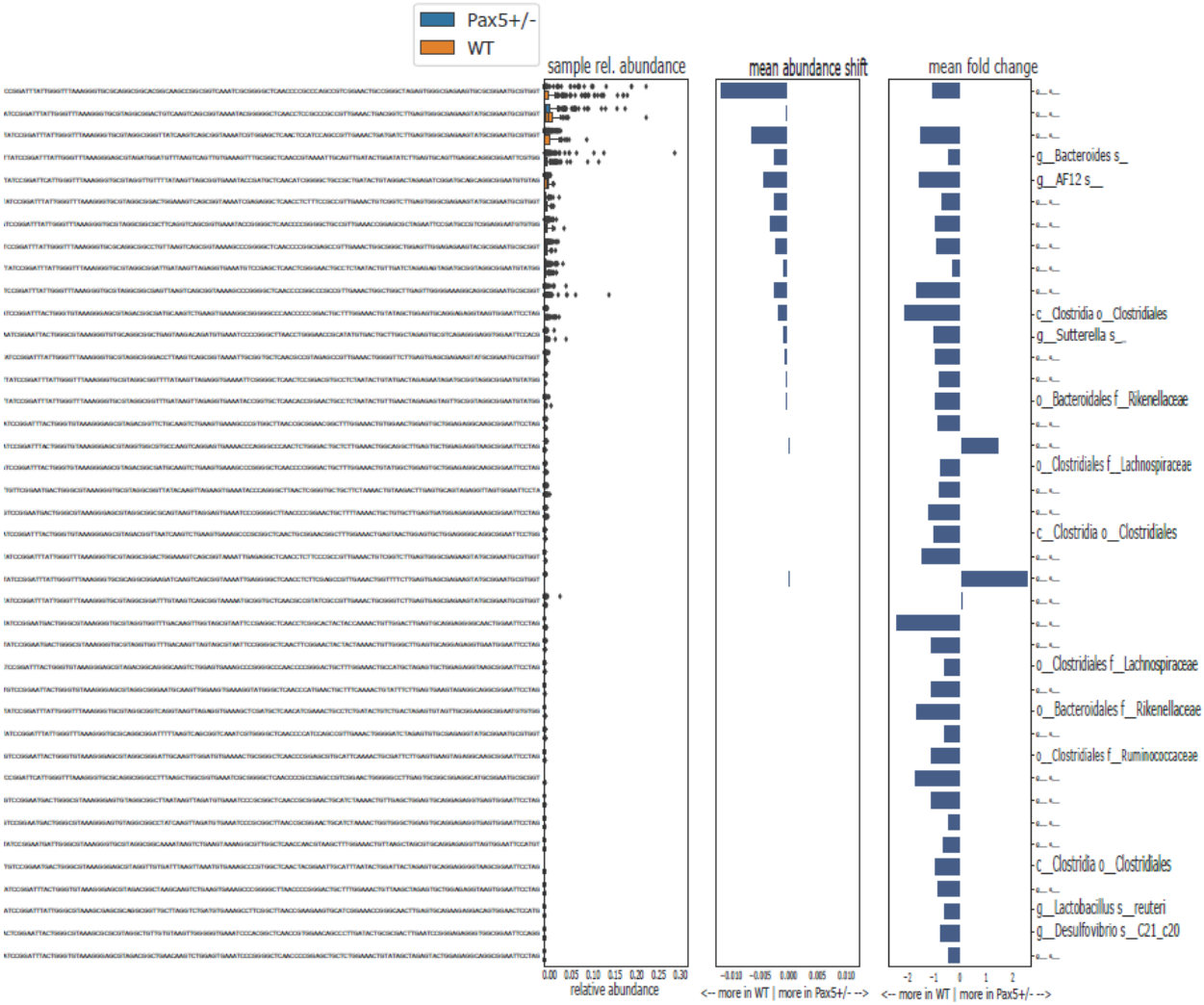

**Supplemental Figure 3. B cells in Peyer's patches are reduced in *Pax5*$^{+/-}$ mice.** Percentage of B cells (B220$^{+}$IgM$^{+/-}$) and T cells (CD3$^{+}$) in Peyer's patches of *Pax5*$^{+/-}$ mice (n = *9*) compared with Peyer's patches of age-matched (3 months-old) WT mice (n = *9*) were analysed by FACS. A significant decrease in B cells was observed in the Peyer's patches of *Pax5*$^{+/-}$ mice, while T cells are increased. Error bars represent the mean and SD. Unpaired *t*-test p-value is indicated in each case.




**Supplemental Figure 4. Experimental flow chart of the *in vivo* study design.** *Pax5*$^{+/-}$ and WT mice were born in the SPF and at the age of 3 months they were moved to the CF. In the CF, mice were exposed to common infections and concomitantly treated with a cocktail of antibiotics for 8 weeks. Throughout the study, stool DNA collection and peripheral blood (PB) FACS analysis was carried out as indicated in the flow chart. At the end of the study, every single mouse was analysed by FACS and all tissues were histologically analysed. Leukemic cells were genetically analysed by WES and by gene expression.

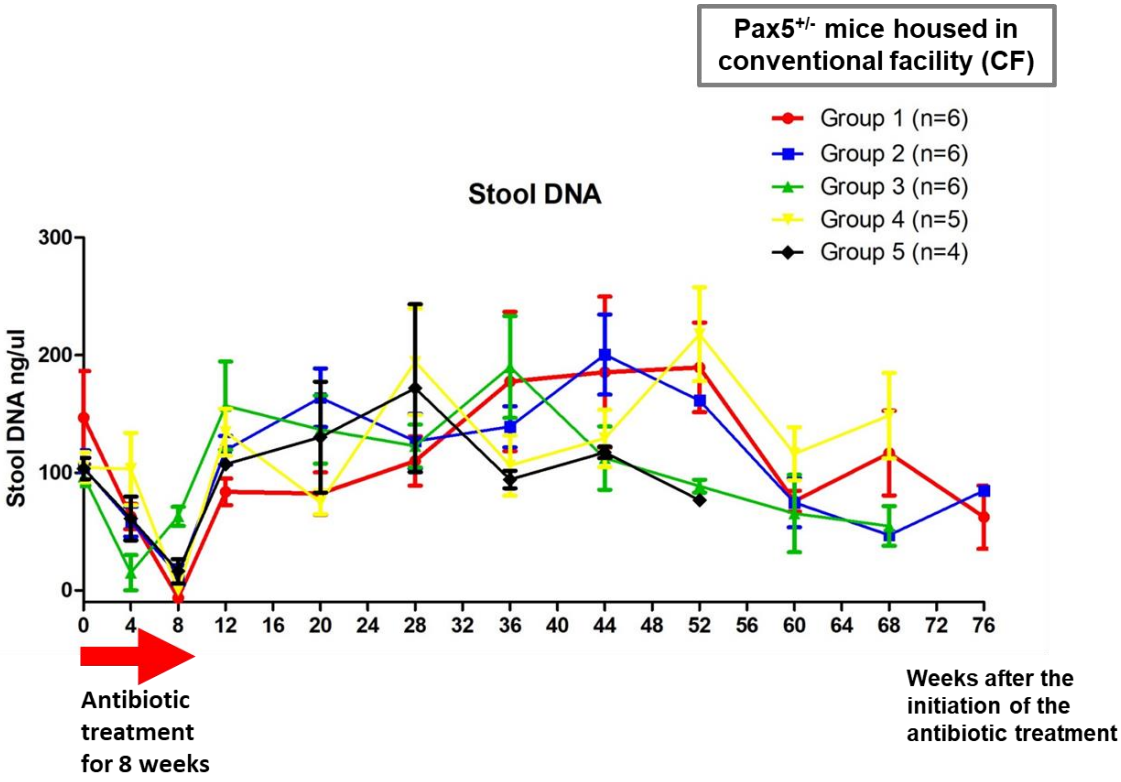



**Supplemental Figure 5. Experimental design of the "abx" cohort.** Mice of the *Pax5*$^{+/-}$ (gray) and WT (white) genotype were born in SPF (light blue). Lifespan of a mouse is indicated with a horizontal bar. The green bar indicates that the mouse remained healthy throughout the experiment and red if pB-ALL occurred at the end of the mice's life. At the age of 10 weeks, approximately six mice were co-housed per cage. 27 *Pax5*$^{+/-}$ and 18 *WT* mice were transferred from SPF to CF (gray). All mice were treated with an antibiotic cocktail introduced in their drinking water for 8 weeks (golden box). Fecal pellets for microbial profiling were collected every ~5 weeks as indicated by vertical lines. Dotted vertical lines correspond to samples that had to be dropped out due to a quality control failure. Abx indicates antibiotic treatment.

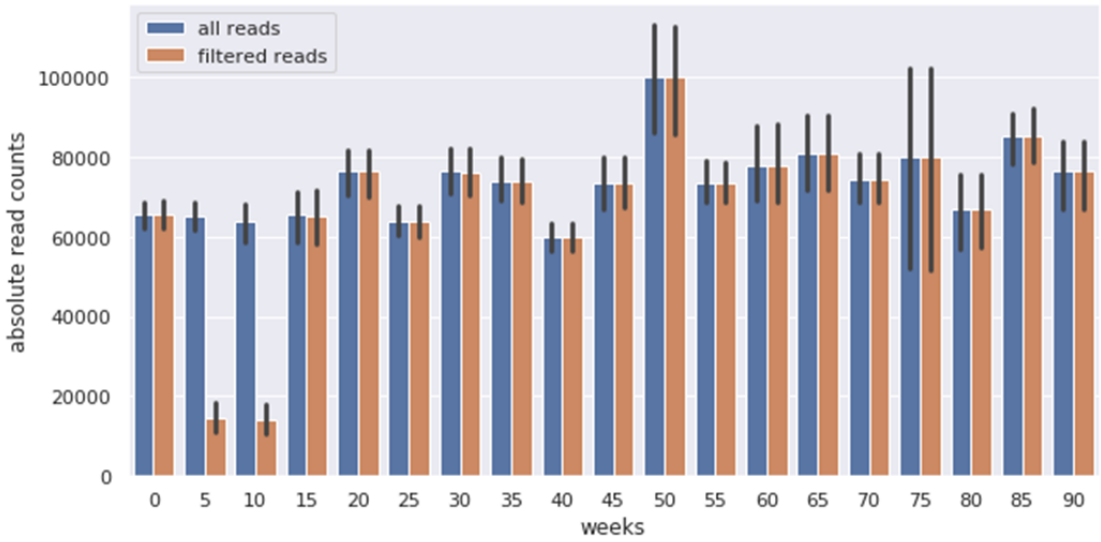



**Supplemental Figure 6. ASVs differentially abundant between WT and *Pax5*$^{+/-}$ DNA stool samples.** All mouse gut samples from the "cohousing"-cohort, except those from genotype-mixed cages, and all mouse gut samples from the "Abx" cohort, except those receiving antibiotic treatment, made up the 329 and 173 samples of *Pax5*$^{+/-}$ and WT genotype animals, respectively. Using dsFDR, we identified 1013 of all 3983 16S V4 bacterial features as statistically significantly differentially abundant between microbial profiles of both genotype groups. Of those 1013 features, Supplementary Figure S6 lists 40 ASVs of those 1013 ASVs, with a minimal mean relative abundance of at most 0.1%, sorted by mean relative abundance, that were also identified by a random forest approach. The left panel uses box-plots to visualize the relative abundance across samples for each feature in *Pax5*$^{+/-}$ samples (blue) and WT samples

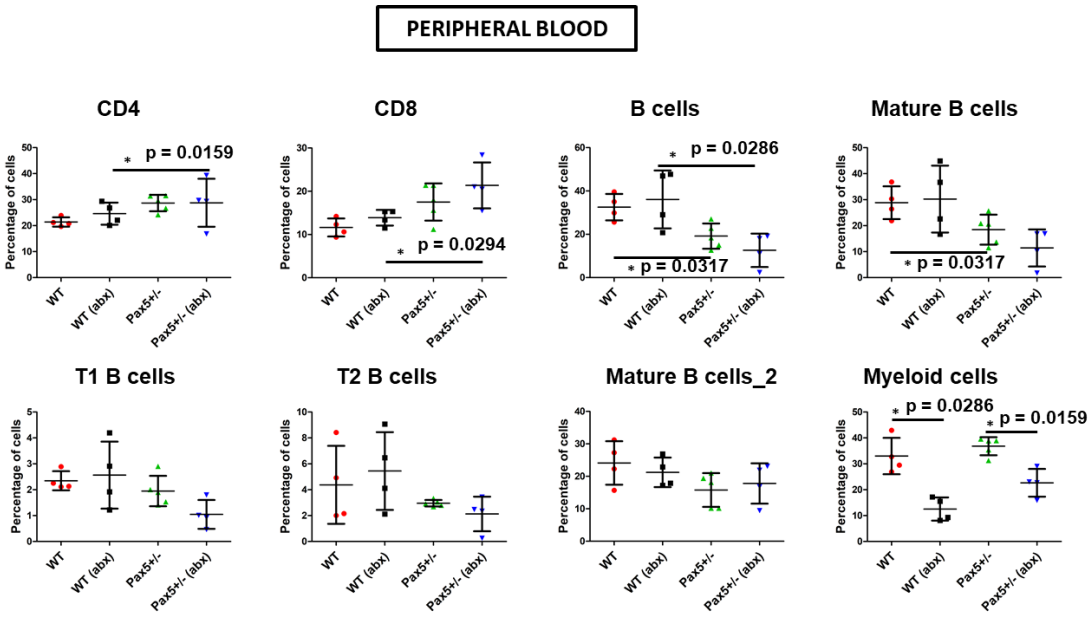
(orange). The middle panel indicates the mean relative abundance shift between *Pax5*$^{+/-}$ and WT. The bar points to the left, if the mean relative abundance of the feature in WT samples is higher compared to *Pax5*$^{+/-}$ and to the right otherwise. The right panel shows the same difference, but as a fold-change (natural logarithm) to highlight strong changes for features with little relative abundance in both genotypes. The y-axis of the right panel reports the two most detailed predicted taxonomic ranks of each feature. 'g' means that the reference database (Greengenes 13.8) holds a genus label for this feature, but its name is empty (unknown).



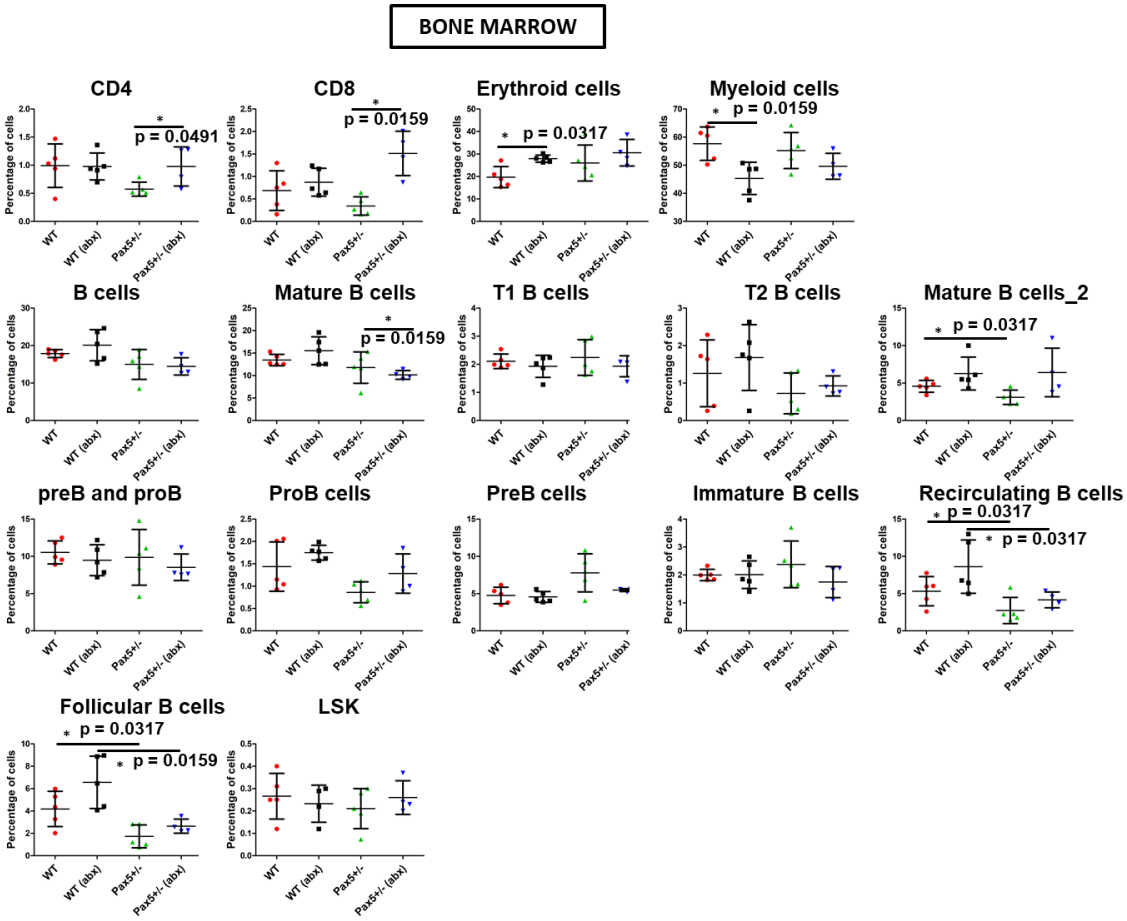

**Supplemental Figure 7. Antibiotics alter the gut microbiota composition.** Representation of the stool DNA concentration from each group of *Pax5*$^{+/-}$ mice treated with antibiotics over time. At the time of antibiotic treatment, bacterial DNA was nearly completely absent from stool samples due to the treatment. One month later the amount of bacterial DNA from the gut was restored.



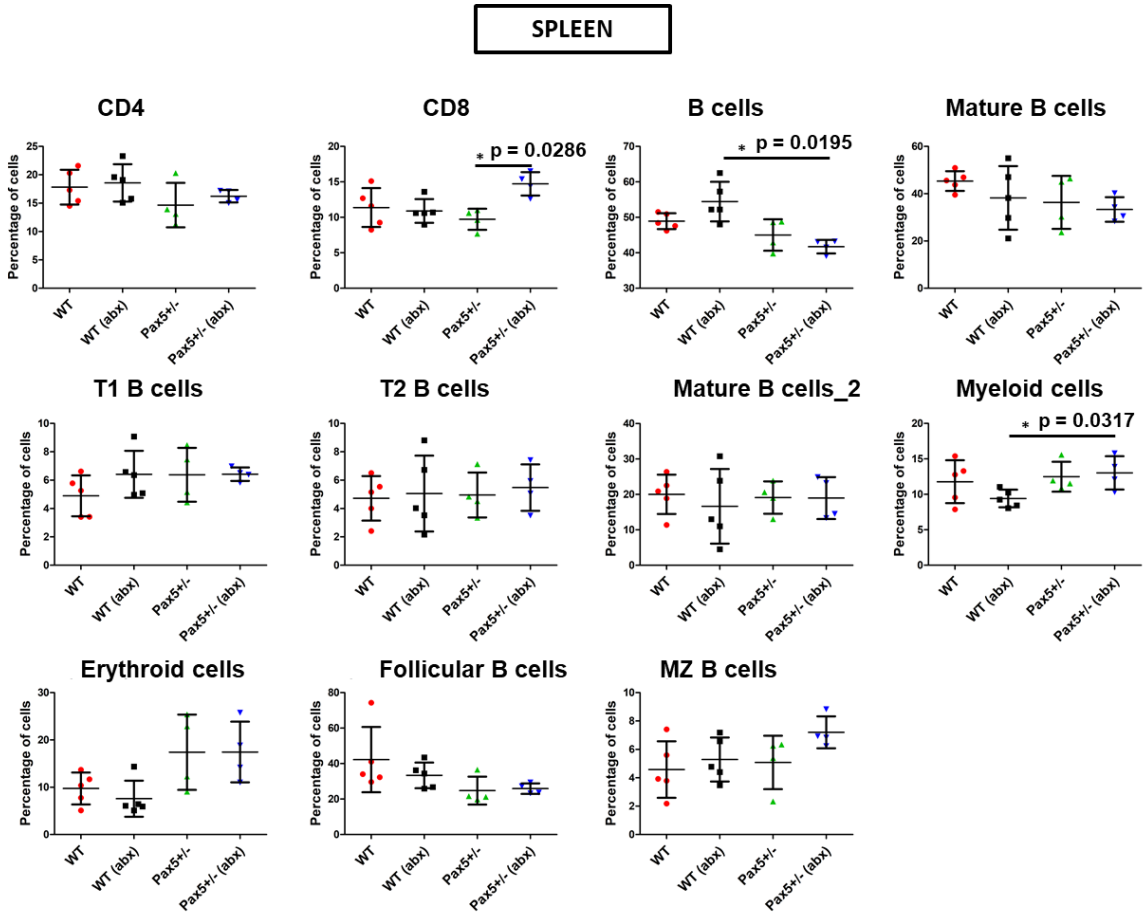

**Supplemental Figure 8. Impact of removing mitochondrial and chloroplast raw reads.** Samples were stratified by time points (x-axis in weeks). Y-axis shows the total number of raw reads per sample in blue. Orange bars indicate total number of raw reads for the same sample, after ASVs and their corresponding reads were removed that are assigned to the class “chloroplast” or the family “mitochondria” and thus most likely originate from plant material (diet) or the mouse host. This removal only affects samples of the antibiotics treatment phase and thus shows its impact on the microbiome. Plant or host reads in all other samples were so minor such that read remove did not lead to significant differences.



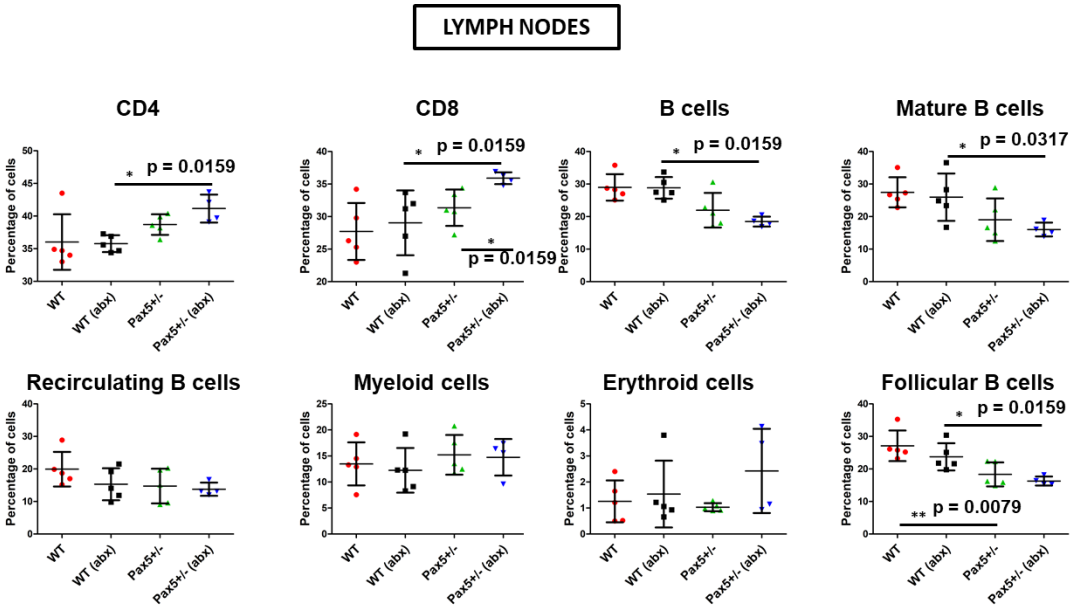

**Supplemental Figure 9. Peripheral blood FACS analysis of *Pax5*$^{+/-}$ mice after antibiotic treatment.** Antibiotic treatment decreased peripheral blood myeloid cells in *Pax5*$^{+/-}$ and *WT* mice. Immune subsets in peripheral blood were analysed by FACS in mice treated with antibiotics (5 WT and 4 *Pax5*$^{+/-}$ mice) and untreated mice (5 WT and 5 *Pax5*$^{+/-}$ mice). See Supplemental Table S4 for cells markers used in the analysis. All mice were analysed just after 8 weeks of antibiotic treatment, at 4 months of age. Error bars represent the mean and SD. Unpaired *t* test *p*-values are indicated. ("Mature B cells" represent CD19$^{+}$B220$^{+}$ and "Mature B cells_2" represent IgD$^{+}$ IgM$^{-}$ B cells.)



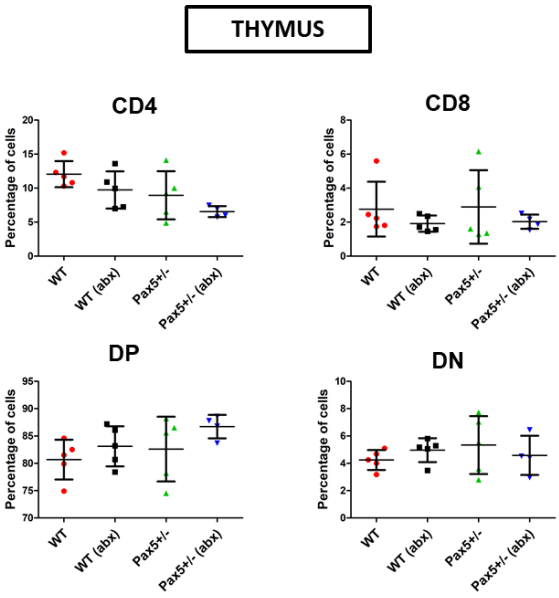

**Supplemental Figure 10. Bone marrow FACS analysis of *Pax5*$^{+/-}$ mice after antibiotic treatment.** Antibiotic treatment increased T cells and reduced mature B cells in the bone marrow of *Pax5*$^{+/-}$ mice. The treatment also reduced myeloid cells and increased erythroid cells in the bone marrow of antibiotic-treated WT mice. Immune subsets in bone marrow were analysed by FACS in mice treated with antibiotics (5 WT and 4 *Pax5*$^{+/-}$ mice) and untreated mice (5 WT and 5 *Pax5*$^{+/-}$ mice). See Supplemental Table S4 for cells markers used in the analysis. All mice were analysed just after 8 weeks of antibiotic treatment, at 4 months of age. Error bars represent the mean and SD. For the significant differences, unpaired t-test p-values are indicated in each case. ("Mature B cells" represent CD19$^{+}$B220$^{+}$ and "Mature B cells_2" IgD$^{+}$ IgM$^{-}$ B cells.)



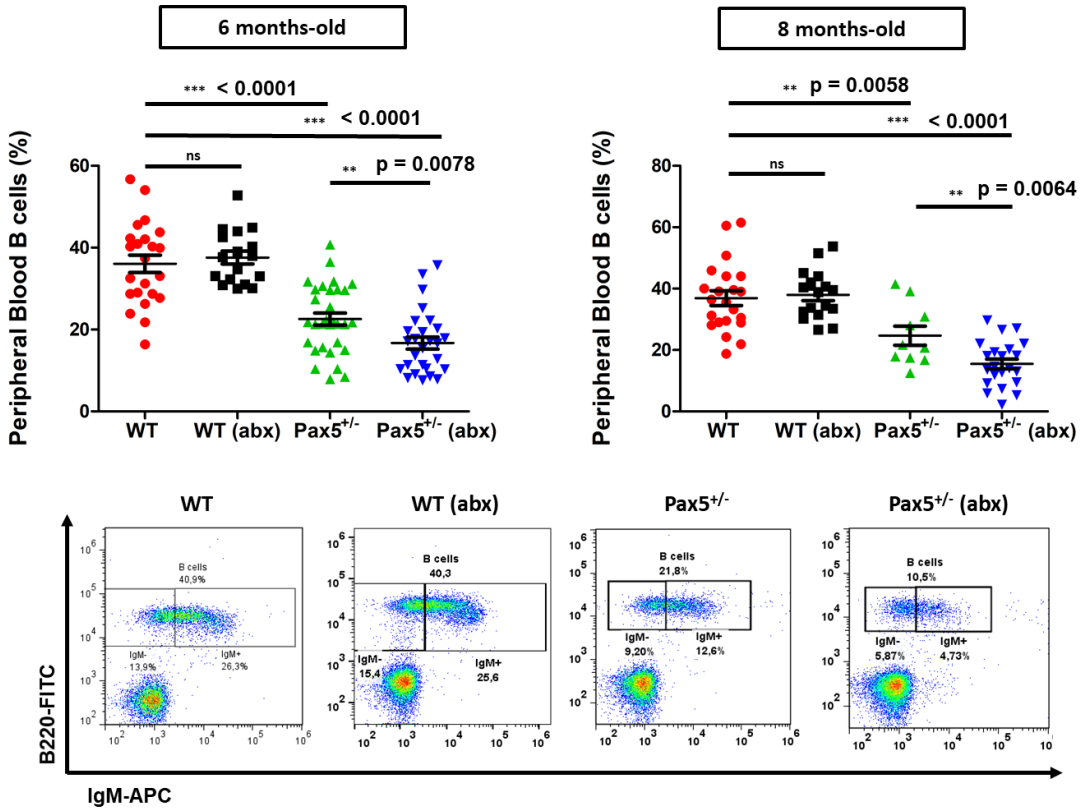

**Supplemental Figure 11. Spleen FACS analysis of *Pax5*$^{+/-}$ mice after antibiotic treatment.** Antibiotic treatment increased splenic CD8$^{+}$ T cells in *Pax5*$^{+/-}$ mice. Immune subsets in the spleen were analysed by FACS in mice treated with antibiotics (5 WT and 4 *Pax5*$^{+/-}$ mice) and untreated mice (5 WT and 4 *Pax5*$^{+/-}$ mice). See Supplemental Table S4 for cells markers used in the analysis. All mice were analysed just after 8 weeks of antibiotic treatment, at 4 months of age. Error bars represent the mean and SD. For the significant differences, unpaired t-test p-values are indicated in each case. ("Mature B cells" represent CD19$^{+}$B220$^{+}$, "Mature B cells_2" IgD$^{+}$ IgM$^{-}$ B cells.)



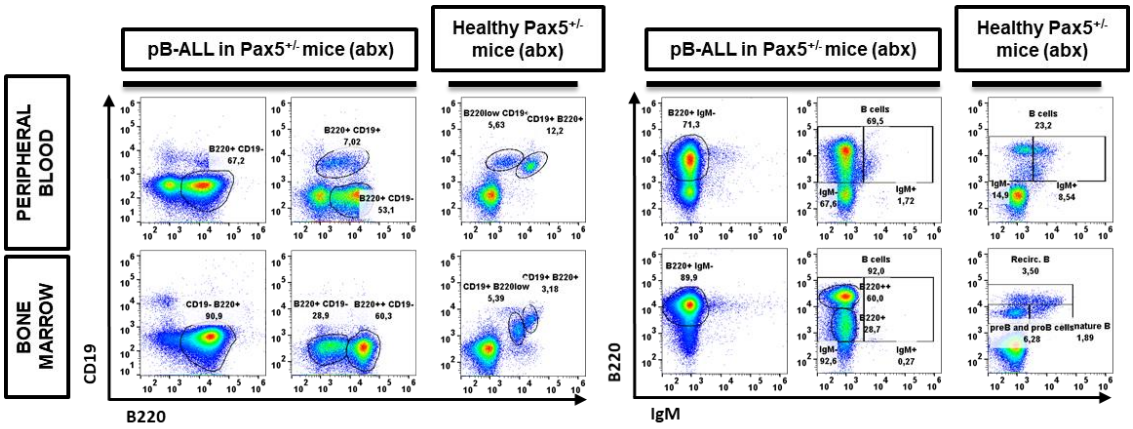

**Supplemental Figure 12. Lymph node FACS analysis of *Pax5*$^{+/-}$ mice after antibiotic treatment.** Antibiotic treatment increased CD8$^{+}$ T cells in the lymph nodes of *Pax5*$^{+/-}$ mice. Immune subsets in the lymph nodes were analysed by FACS in mice treated with antibiotics (5 WT and 4 *Pax5*$^{+/-}$ mice) and untreated mice (5 WT and 5 *Pax5*$^{+/-}$ mice). See Supplemental Table S4 for cells markers used in the analysis. All mice were analysed just after 8 weeks of antibiotic treatment, at 4 months of age. Error bars represent the mean and SD. For the significant differences, unpaired t-test p-values are indicated in each case. ("Mature B cells" represent CD19$^{+}$B220$^{+}$ B-cells.)



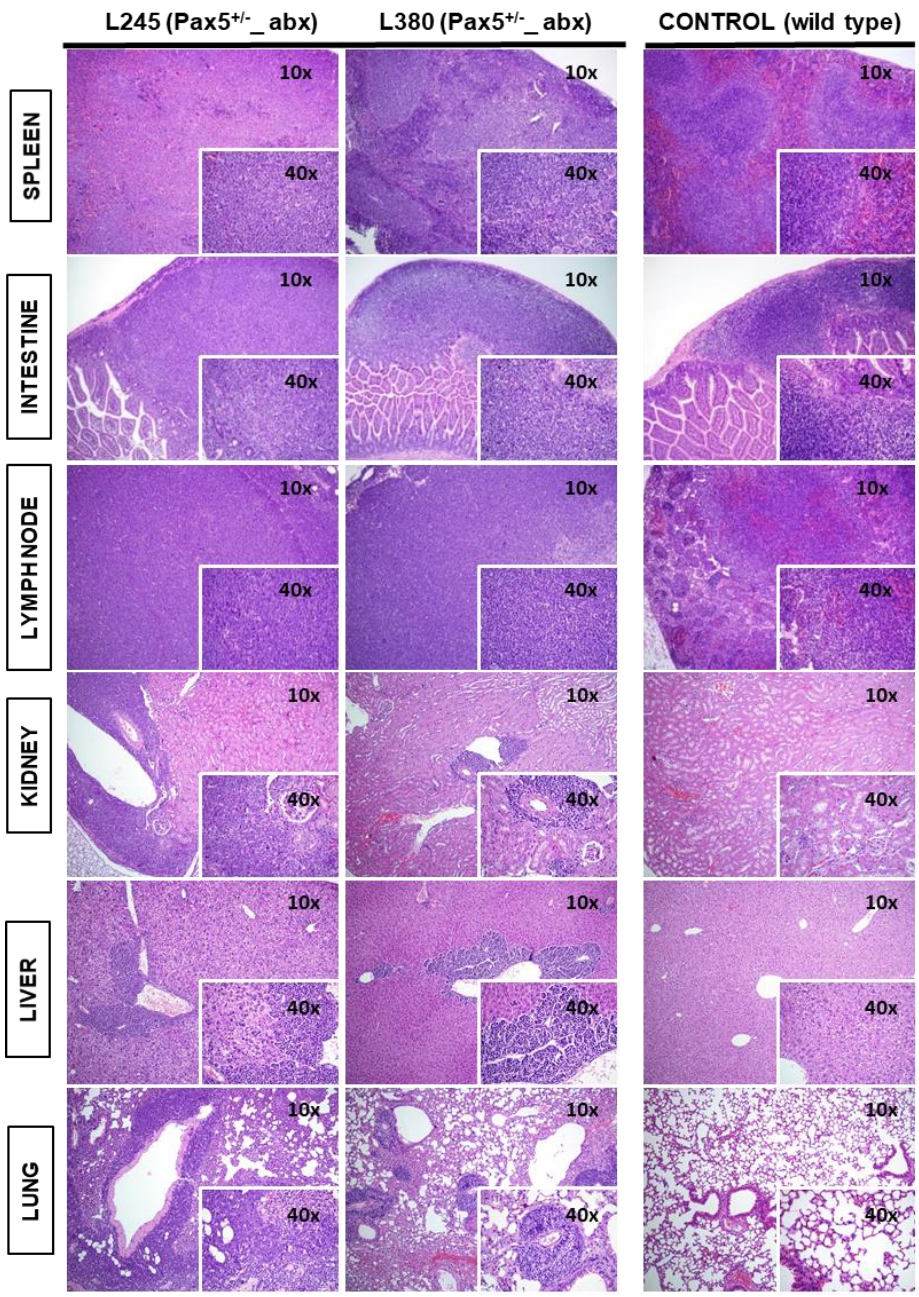

**Supplemental Figure 13. Thymus FACS analysis of *Pax5*$^{+/-}$ mice after antibiotic treatment.** Antibiotic treatment did not affect thymic CD8$^{+}$ T cells. Immune subsets in the thymus were analysed by FACS in mice treated with antibiotics (5 WT and 4 *Pax5*$^{+/-}$ mice) and untreated mice (5 WT and 5 *Pax5*$^{+/-}$ mice). See Supplemental Table S4 for cells markers used in the analysis. All mice were analysed just after 8 weeks of antibiotic treatment, at 4 months of age. Error bars represent the mean and SD. For the significant differences, unpaired t-test p-values are indicated in each case.



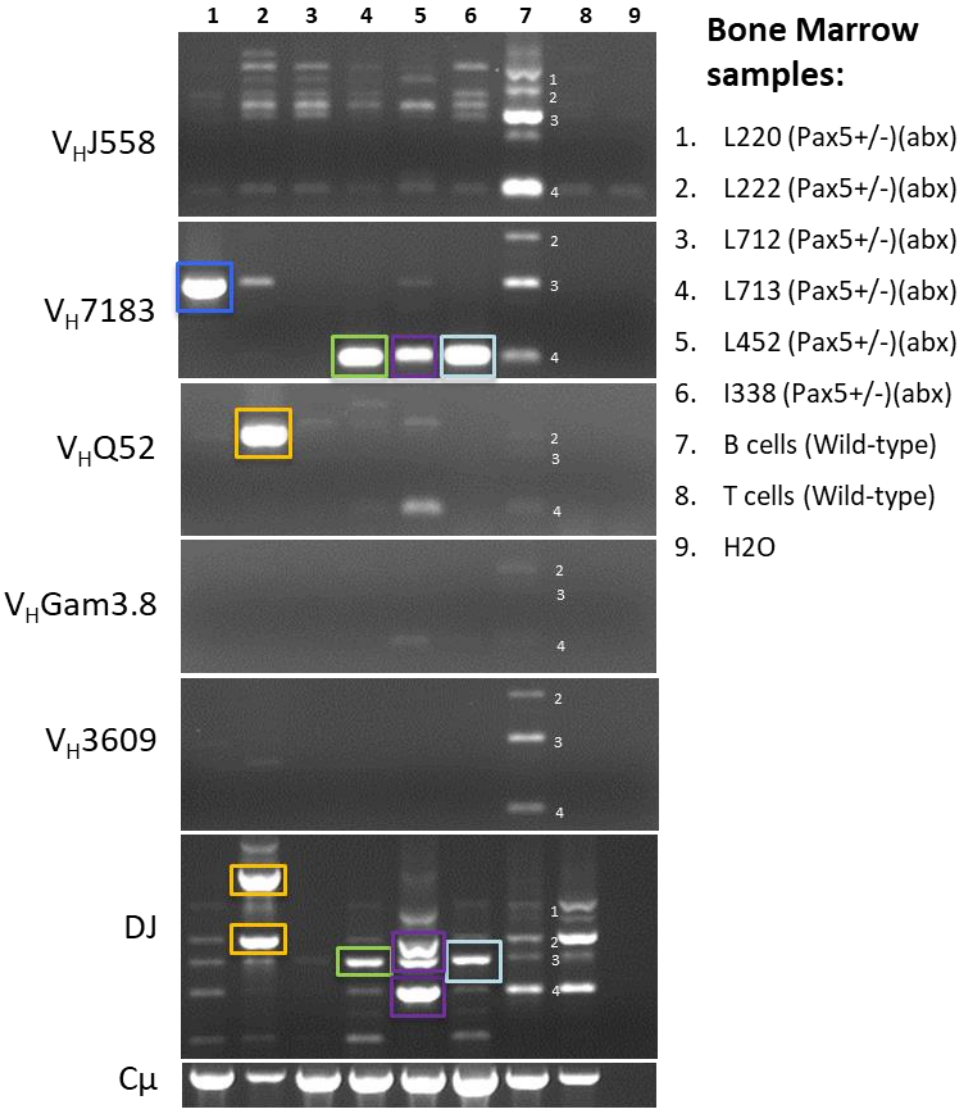

**Supplemental Figure 14. Reduction of peripheral blood B cells in *Pax5*$^{+/-}$ mice treated with antibiotics.** Peripheral blood B-cells were analysed by FACS in *Pax5*$^{+/-}$ mice treated with antibiotics (n=27), age-matched (6 months-old and 8 months-old) *WT* mice (n=23), antibiotic treated *WT* mice (n=17) and untreated *Pax5*$^{+/-}$ (n=31) mice. Error bars represent the mean and SD. Unpaired *t* test *p*-value is indicated in each case.



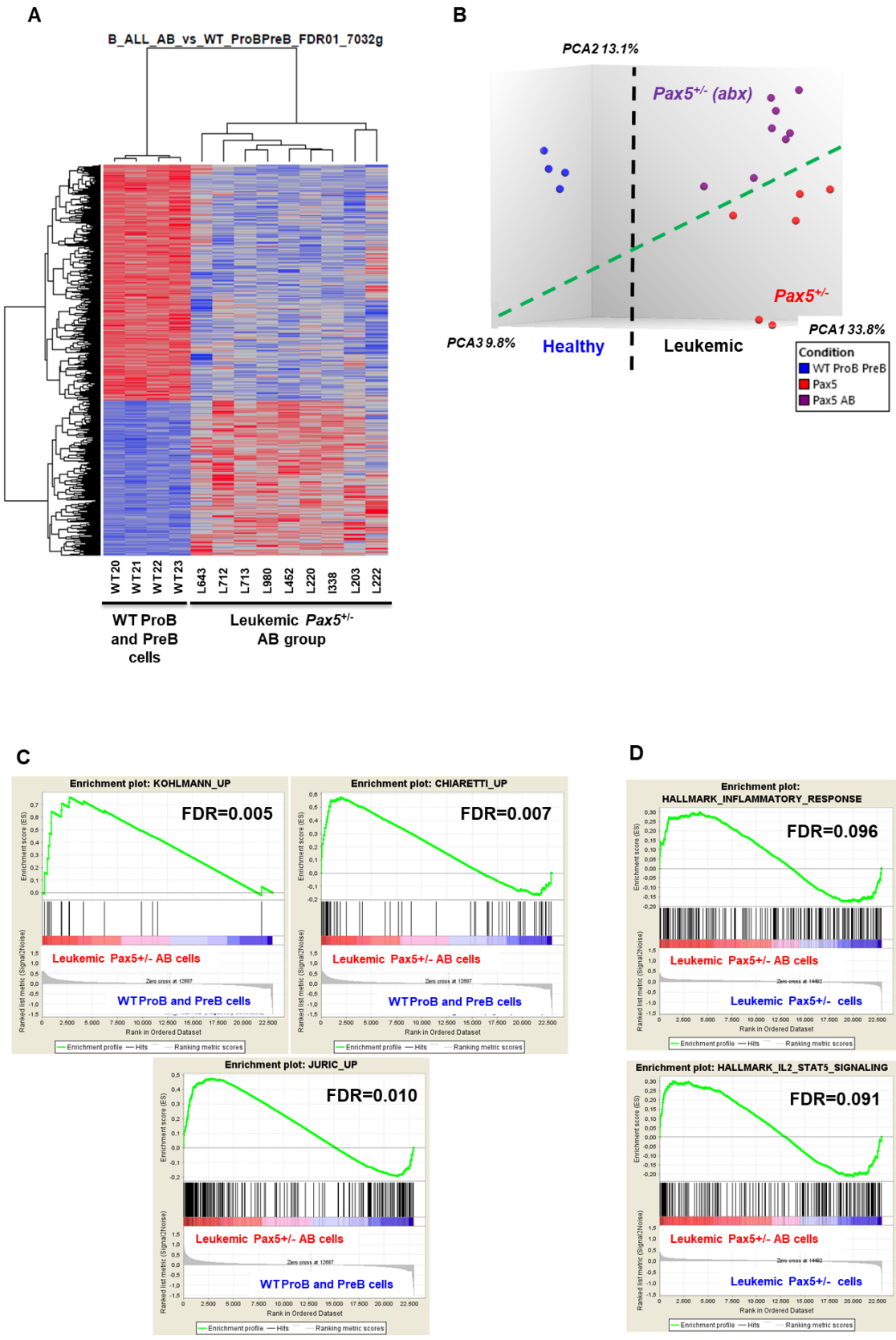

**Supplemental Figure 15. pB-ALL development in *Pax5*$^{+/-}$ mice treated with antibiotics and exposed to infectious conditions.** Representative FACS blots of B cell ($CD19^{+}B220^{+}$ and $B220^{+}IgM^{+/-}$) subsets in PB and BM from diseased *Pax5*$^{+/-}$ mice treated with antibiotics and exposed to common infections compared to an age-matched healthy *Pax5*$^{+/-}$ mouse are shown.




**Supplemental Figure 16. pB-ALL in *Pax5*$^{+/-}$ mice after depletion of gut commensal bacteria.** Hematoxylin and eosin stainings demonstrating blast infiltration of spleen, small intestine, lymph nodes, kidney, liver and lung from *Pax5*$^{+/-}$ antibiotic-treated leukemic mice (L245 and L380) are shown. Loss of normal architecture resulting in cells morphologically resembling lymphoblast can be observed. Tissues from a control littermate WT mouse are shown for reference. Abx: antibiotic treatment.

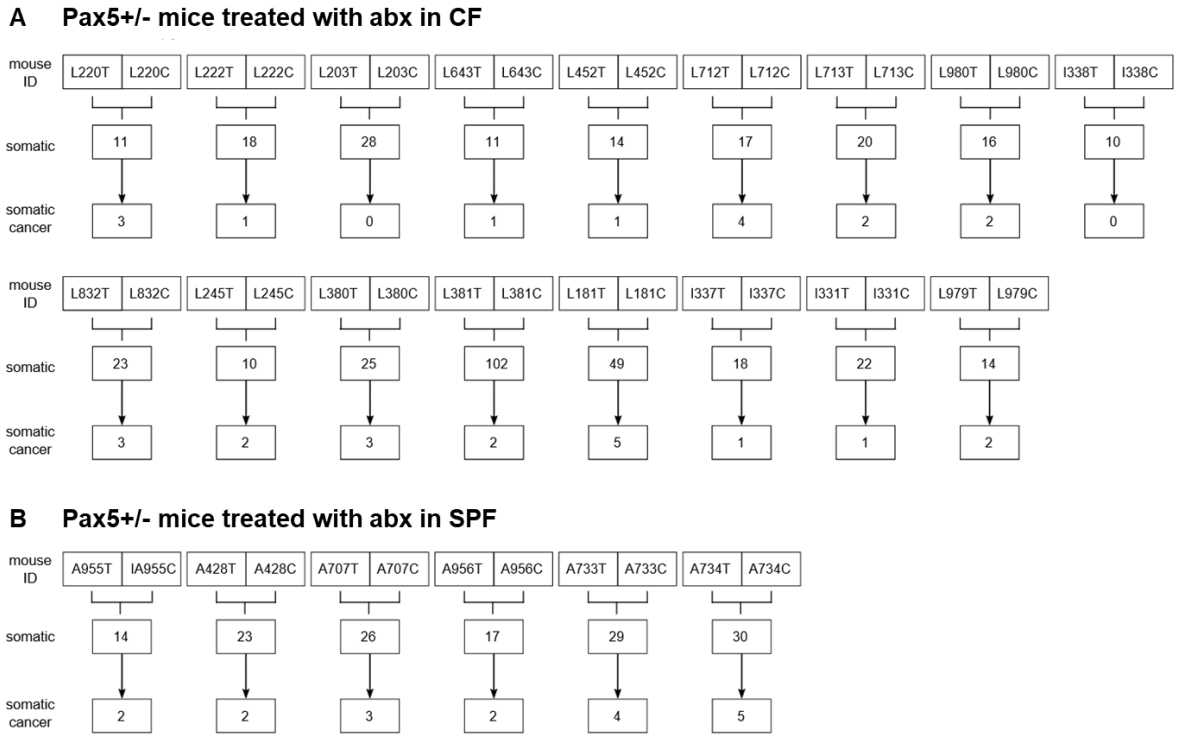



**Supplemental Figure 17. Analysis of BCR clonality of leukemias arising in *Pax5*$^{+/-}$ mice treated with antibiotics.** PCR analysis of BCR gene rearrangements in bone marrow of diseased mice is shown. Sorted CD19$^{+}$ splenic B cells (B cells) of healthy mice served as a control for polyclonal BCR rearrangements. CD8$^{+}$CD4$^{+}$ T cells from the thymus of healthy mice served as a negative control. Bone marrow leukemic cells showed an increased clonality within their BCR repertoire (indicated by the code number of each mouse analysed).

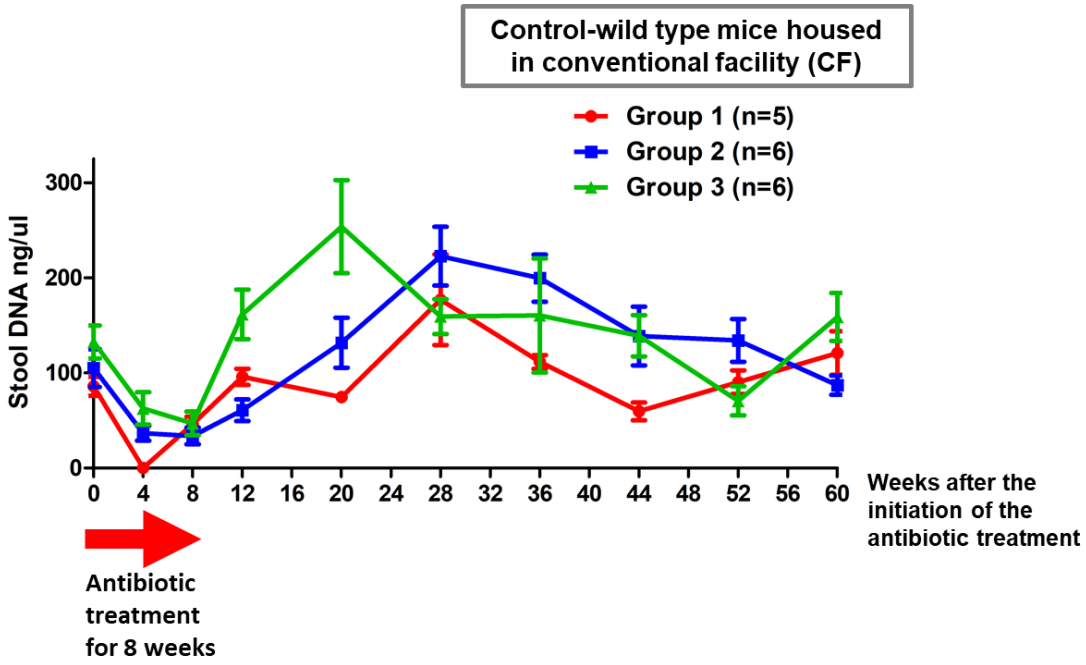



**Supplemental Figure 18. Gene expression analysis of leukemic *Pax5*$^{+/-}$ pB-ALL with an altered gut microbiome. A)** Unsupervised heatmap showing the

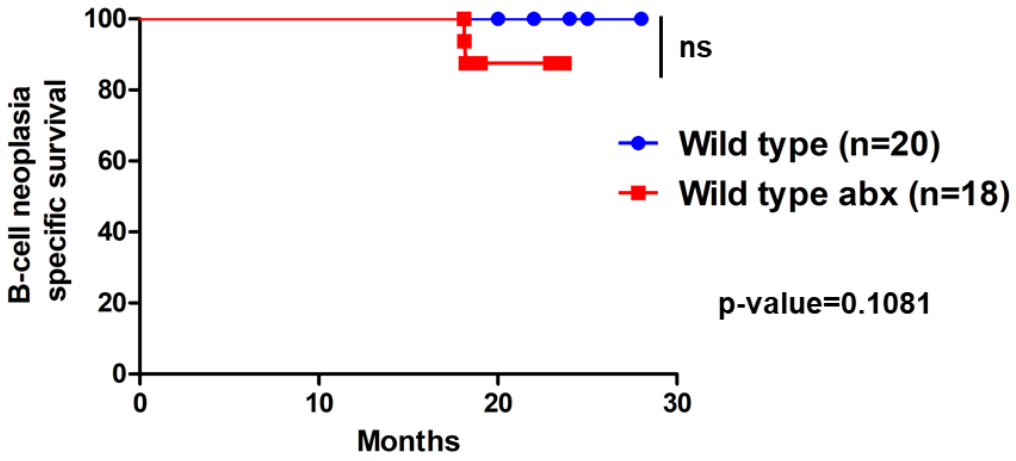
differentially expressed genes between leukemic cells from *Pax5*$^{+/-}$ mice treated with antibiotics (n=9) and proB and preB cells from WT mice (n=4) is shown. The significance analysis of microarrays was defined by a FDR =0.01%. **B)** Principal component analysis plot showing the differences in gene expression of leukemic cells from *Pax5*$^{+/-}$ mice treated with antibiotics (purple dots, n=8), untreated *Pax5*$^{+/-}$ leukemic mice (red dots, n=6) and healthy pro-B and pre-B cells from WT heathy mice (blue dots, n=4). **C)** GSEA plots showing that leukemic cells from *Pax5*$^{+/-}$ mice treated with antibiotics are enriched in specific genesets built from the lists of genes upregulated in human pB-ALL reported by Juric et al.[19] (54 cases of adult BCR-ABL positive and negative ALL), Kohlmann et al.[20] (34 cases of adult ALL) and Chiaretti et al.[18] (128 cases of adult ALL). **D)** GSEA plots showing a partial, bilateral enrichment of inflammatory responses and IL2-STAT5 signaling pathways in leukemic cells derived from *Pax5*$^{+/-}$ mice treated with antibiotics compared to leukemic cells from untreated *Pax5*$^{+/-}$ mice.



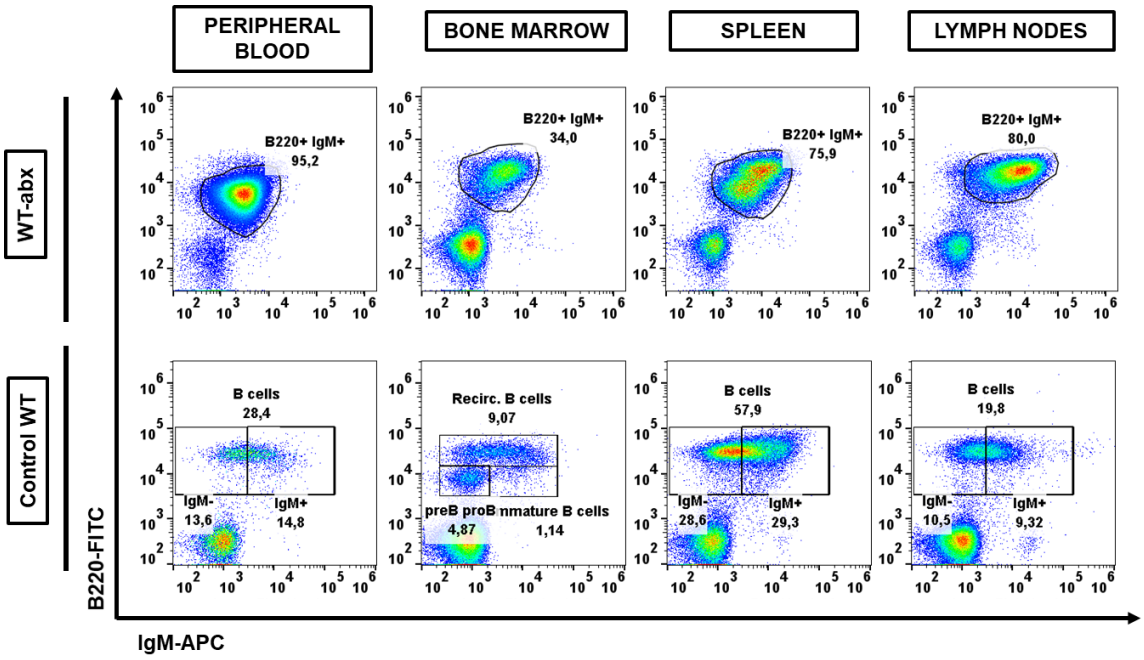

**Supplemental Figure 19. Number of called somatic mutations in *Pax5*$^{+/-}$ pB-ALL of mice treated with antibiotics.** Whole exome sequencing analysis of tumor (T) (leukemic cells) and control samples (tail DNA from the same mouse) from mice exposed to infections **A**) and mice not exposed to infections **B**) was performed by employing the tools *mutect* and *varscan*. The Cancer Gene Consensus list by Cosmic was used to identify somatic hits in cancer genes.



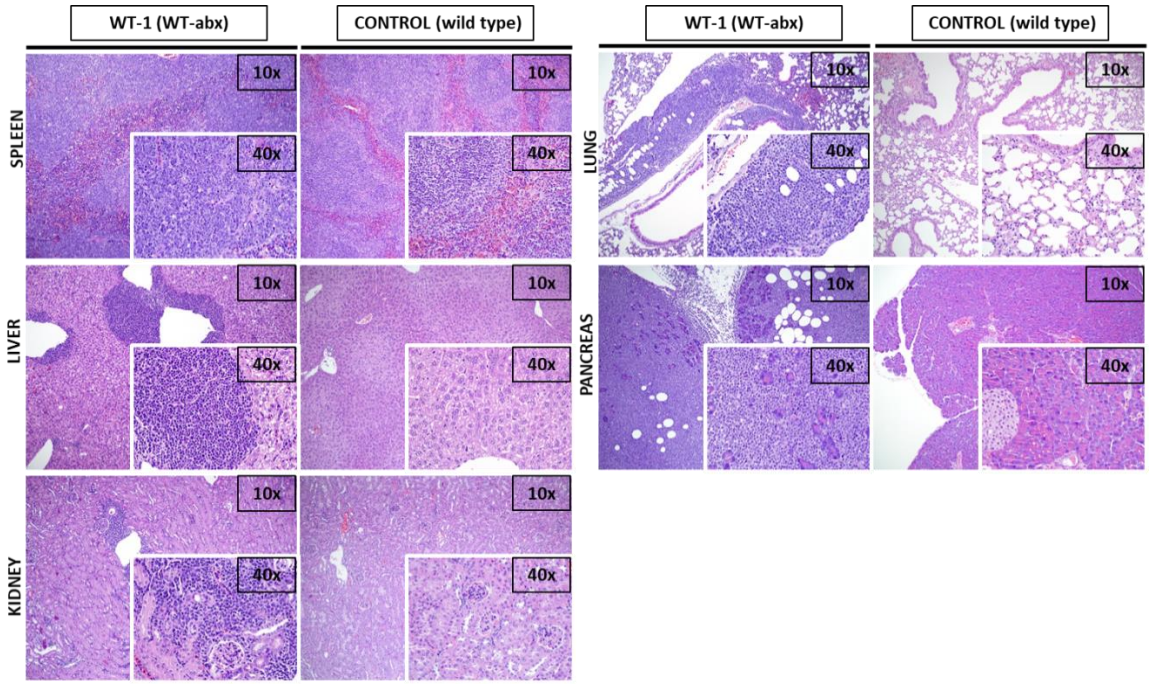

**Supplemental Figure 20. Antibiotics treatment alters the gut microbiota composition in control WT mice.** Representation of the Stool DNA concentration from each group of WT mice treated with antibiotics over time. At the time of antibiotic treatment, bacterial DNA content was nearly completely absent from stool samples due to the treatment. One month later the amount of bacterial DNA from the gut was restored.



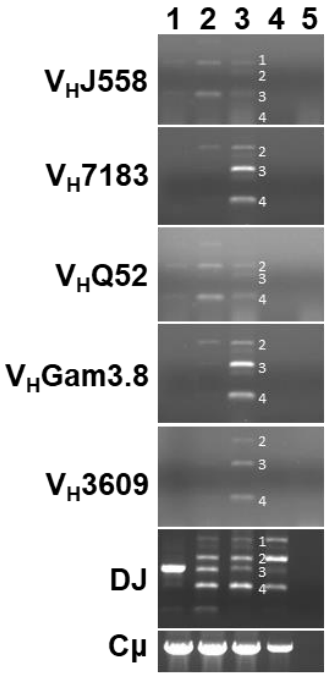

**Supplemental Figure 21. WT mice treated with antibiotics develop B-cell neoplasias.** B cell neoplasia-specific survival curve of *WT* mice treated with antibiotics (red line, n=18), showing a statistically insignificant shortened lifespan compared to untreated *WT* mice (blue line, n=20) as a result of B cell neoplasia development. Log-rank (Mantel-Cox) p-value = 0.1081.




**Supplemental Figure 22. B cell neoplasia in WT mice following depletion of gut commensal bacteria.** FACS analysis of B cells in diseased *WT* antibiotic-treated mice is shown. Representative FACS blots of B cell subsets in the PB, BM, spleen and lymph nodes compared to age-matched control littermate *WT* mice are shown.



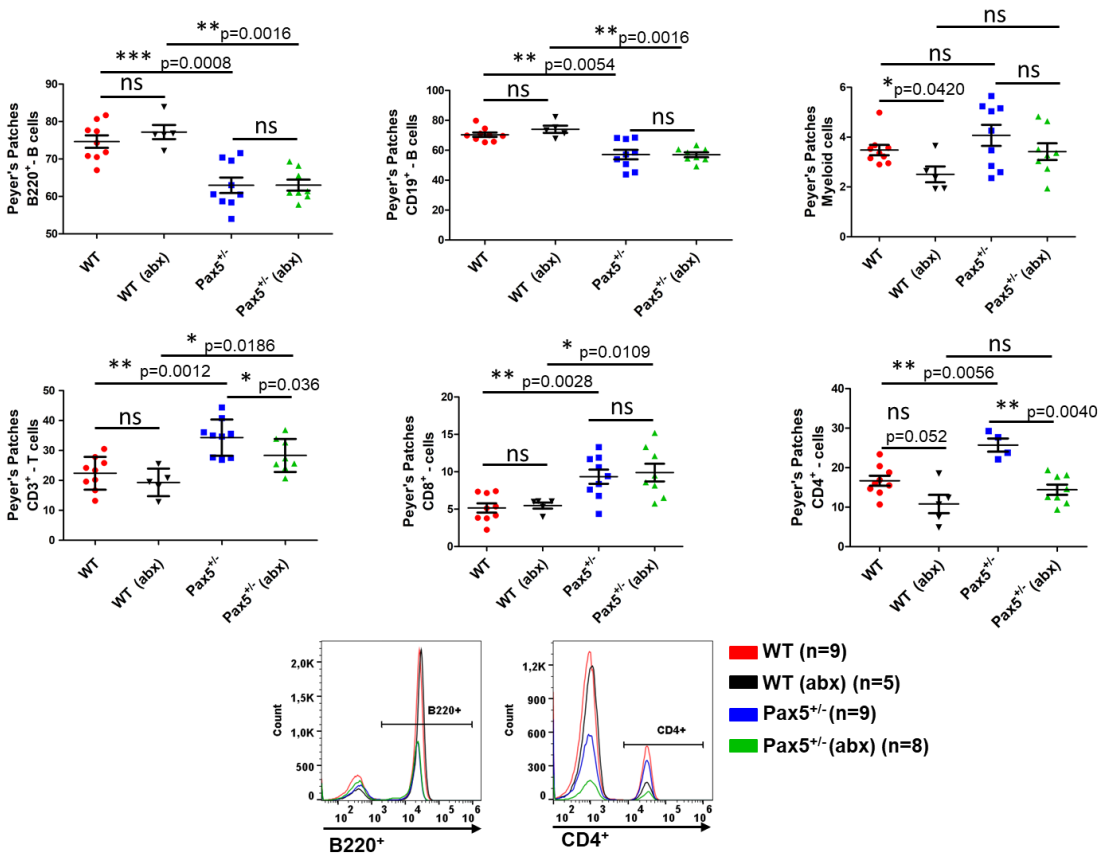

**Supplemental Figure 23. Histological images of B cell neoplasia in WT mice following depletion of gut commensal bacteria.** Hematoxylin and eosin stainings of spleen, liver, kidney, lung, and pancreas harvested from a tumor-bearing *WT* mouse treated with antibiotics and compared to a healthy WT control mouse are shown. All tissues from the diseased mouse are characterized by a loss of normal architecture resulting from effacement with cells morphologically resembling lymphoblasts.



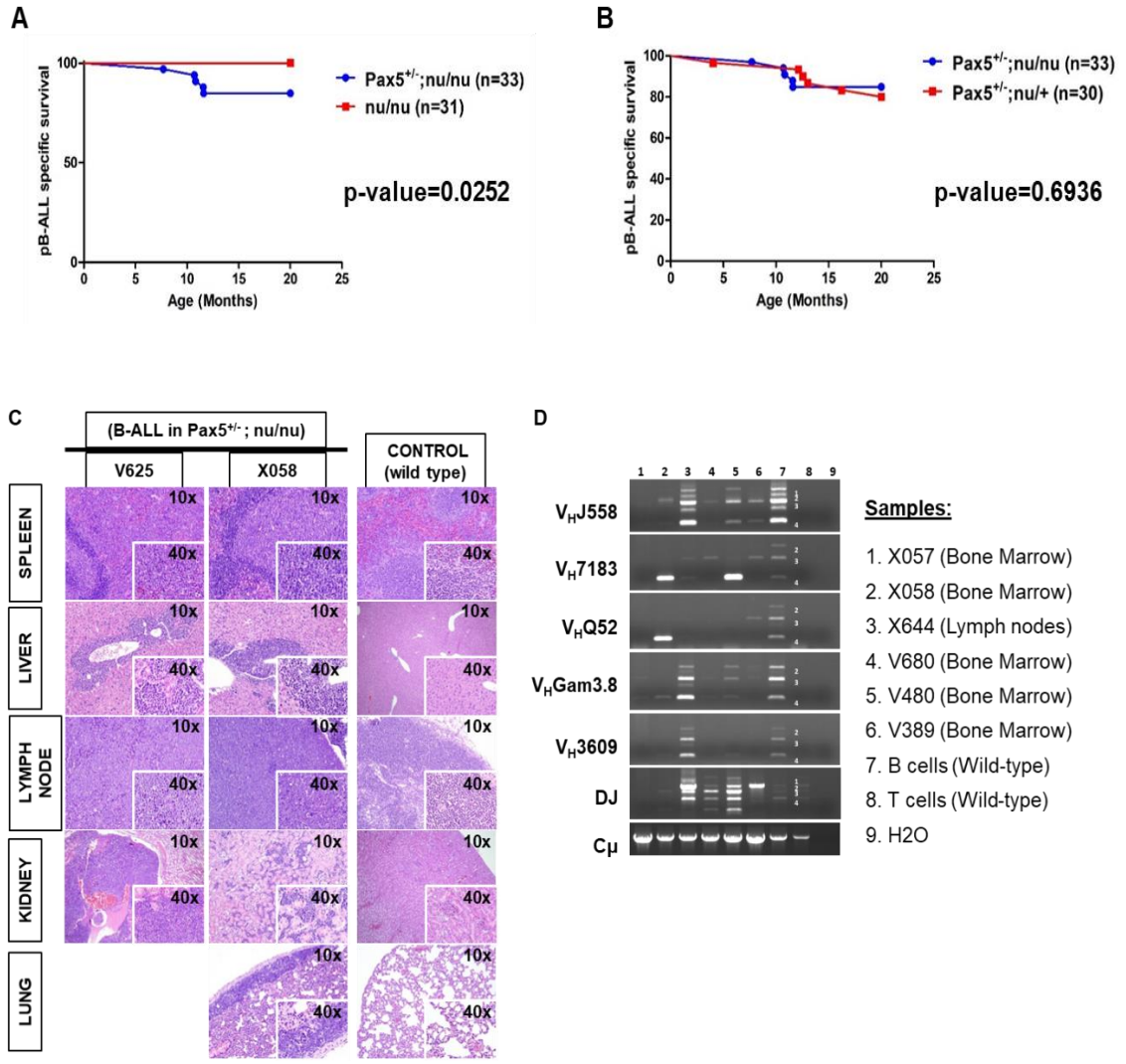

**Samples:**

1. WT-1 (WT- abx) (Peripheral Blood)
2. WT-2 (WT- abx) (Spleen)
3. B cells (Wild-type)
4. T cells (Wild-type)
5. H2O

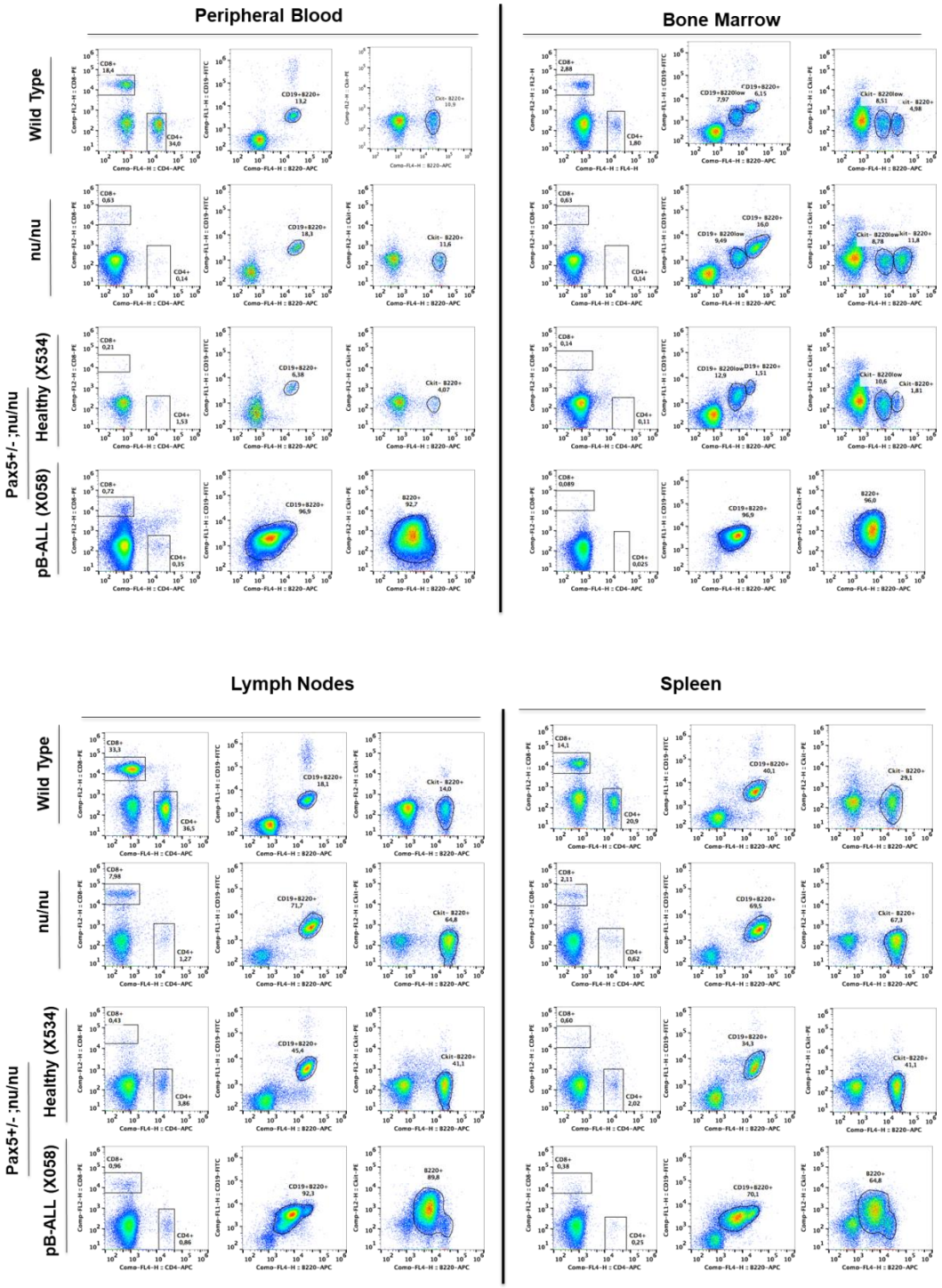
**Supplemental Figure 24. BCR clonality in WT mice treated with antibiotics that developed B-cell neoplasias**. PCR analyses of BCR gene rearrangements in infiltrated tissues (peripheral blood and spleen) of diseased mice are shown. Sorted $CD19^+$ splenic B cells (B cells) of healthy mice served as a control for polyclonal BCR rearrangements. DP ($CD8^+CD4^+$ T) cells from the thymus of healthy mice served as a negative control. Peripheral blood from WT-1 showed an increased clonality within the BCR repertoire.



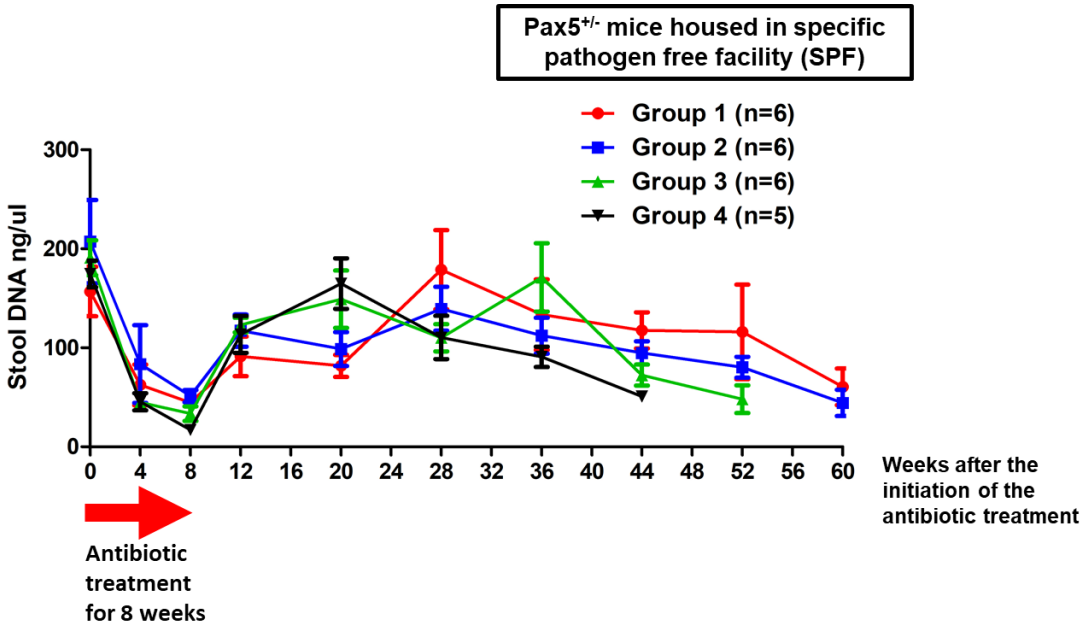

**Supplemental Figure 25. Numbers of T cells in Peyer's patches are decreased following depletion of gut commensal bacteria by antibiotic treatment**. Cellular composition of Peyer's Patches was analysed by FACS in mice treated with antibiotics (5 WT mice and 4 *Pax5*$^{+/-}$ mice) and untreated mice (9 WT mice and 9 *Pax5*$^{+/-}$ mice). The treatment with antibiotics in *Pax5*$^{+/-}$ mice did not enhance the reduction of B cells in Peyer's Patches but decreased the number of CD3$^{+}$ and CD4$^{+}$T cells in *Pax5*$^{+/-}$ antibiotic treated mice. Error bars represent the mean and SD. For the significant differences, unpaired *t*-test *p*-values are indicated in each case.



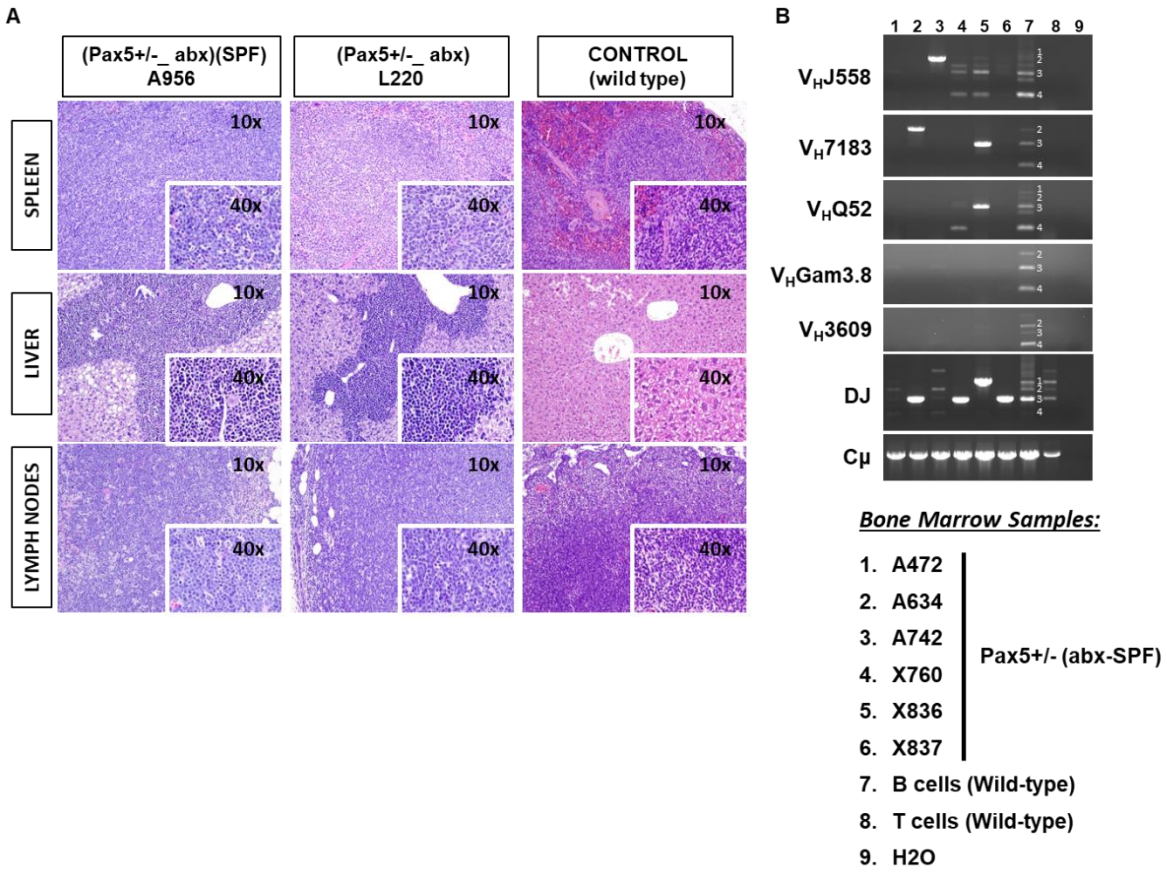

**Supplemental Figure 26. pB-ALL development in *Pax5$^{+/-}$;nu/nu* mice. (A)** *Pax5$^{+/-}$;nu/nu* mice (blue line, *n*=33) had a shorter leukemia-specific survival compared to nu/nu mice (red line, *n*=31). Log-rank (Mantel-Cox) P value = 0.0252. **(B)** *Pax5$^{+/-}$;nu/nu* mice (blue line, *n*=33), had a similar lifespan compared to *Pax5$^{+/-}$;nu/+* mice (red line, *n*=30). Log-rank (Mantel-Cox) P value = 0.6936. **(C)** Haematoxylin and eosin staining of WT mice and tumor-bearing *Pax5$^{+/-}$;nu/nu* mice demonstrating infiltrating blast cells in spleen, liver, lymph node, kidney and lung are shown. Loss of normal architecture resulting from cells morphologically resembling lymphoblasts can be observed. **(D)** BCR clonality in *Pax5$^{+/-}$;nu/nu* pB-ALL is shown. PCR analysis of BCR gene rearrangements in infiltrated BM and lymph nodes of diseased *Pax5$^{+/-}$;nu/nu* (n=3) and *Pax5$^{+/-}$;nu/+* (n=3) leukemic mice. Leukemic tissue shows an increased clonality within the BCR repertoire.



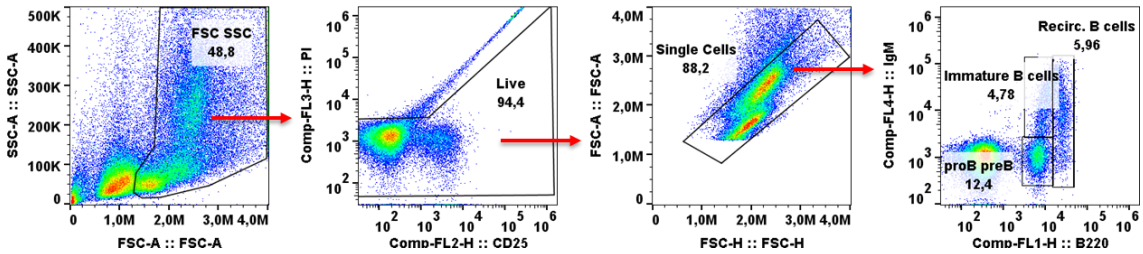

**Supplemental Figure 27. FACS analysis of pB-ALL in *Pax5$^{+/-}$;nu/nu* mice.** Representative FACS blots of hematopoietic subsets in diseased *Pax5$^{+/-}$;nu/nu* mice are shown. FACS blots from the peripheral blood, bone marrow, lymph nodes and spleen indicate an accumulation of blast B cells in a leukemic *Pax5$^{+/-}$;nu/nu* mouse (X058) (age: 7.7-11.6 months) compared to an age-matched control littermate *WT*, nu/nu and a healthy *Pax5$^{+/-}$;nu/nu* (X534) mouse.




**Supplemental Figure 28. Treatment with antibiotics alters the gut microbiota composition in *Pax5*$^{+/-}$ mice housed in SPF.** Representation of the stool DNA concentration from each group of *Pax5*$^{+/-}$ mice treated with antibiotics over time and housed in an SPF facility is shown. Bacterial DNA content from stool samples was nearly completely absent due to the antibiotic treatment. One month later, the amount of bacterial DNA from the gut was restored.




**Supplemental Figure 29. Bacterial depletion of the gut promotes pB-ALL development in genetically predisposed mice in the absence of natural infectious stimuli. A)** Haematoxylin and eosin staining of *WT* mice and tumor-bearing *Pax5*$^{+/-}$ mice with an altered gut microbiome, housed in CF and SPF conditions are shown. Infiltrating blast cells in spleen, liver and lymph nodes are evident. Loss of normal architecture due to the accumulation of cells morphologically resembling lymphoblasts can be seen in both diseased *Pax5*$^{+/-}$ mice. **B)** Analysis of BCR clonality of leukemias arising in *Pax5*$^{+/-}$ mice treated with antibiotics and housed in SPF conditions is shown. PCR analysis of BCR gene rearrangements in the bone marrow of diseased mice. Sorted CD19+ splenic B cells (B cells) of healthy mice served as a control for polyclonal BCR rearrangements. $CD8^{+}CD4^{+}$ T cells from the thymus of healthy mice served as a negative control. Bone marrow leukemic cells showed an increased clonality within the BCR repertoire (indicated by the code number of each mouse analysed).




**Supplemental Figure 30. Gating strategy used in FACS analyses.** The general gating strategy used in all flow cytometric analyses throughout the manuscript is illustrated. For each analysis, a total of at least 50,000 viable cells (PI-; Propidium iodide negative cells) were assessed.

## REFERENCES


1. Urbanek P, Wang ZQ, Fetka I, Wagner EF, Busslinger M. Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell*. 1994;79(5):901-912.

2. Rodriguez-Hernandez G, Hauer J, Martin-Lorenzo A, et al. Infection Exposure Promotes ETV6-RUNX1 Precursor B-cell Leukemia via Impaired H3K4 Demethylases. *Cancer Res*. 2017;77(16):4365-4377.

3. Martin-Lorenzo A, Hauer J, Vicente-Duenas C, et al. Infection Exposure is a Causal Factor in B-cell Precursor Acute Lymphoblastic Leukemia as a Result of Pax5-Inherited Susceptibility. *Cancer Discov*. 2015;5(12):1328-1343.

4. Cording S, Fleissner D, Heimesaat MM, et al. Commensal microbiota drive proliferation of conventional and Foxp3(+) regulatory CD4(+) T cells in mesenteric lymph nodes and Peyer's patches. *Eur J Microbiol Immunol (Bp)*. 2013;3(1):1-10.

5. Barupal DK, Zhang Y, Shen T, et al. A Comprehensive Plasma Metabolomics Dataset for a Cohort of Mouse Knockouts within the International Mouse Phenotyping Consortium. *Metabolites*. 2019;9(5).

6. Gu J, Weber K, Klemp E, et al. Identifying core features of adaptive metabolic mechanisms for chronic heat stress attenuation contributing to systems robustness. *Integr Biol (Camb)*. 2012;4(5):480-493.

7. Shim SH, Lee SK, Lee DW, et al. Loss of Function of Rice Plastidic Glycolate/Glycerate Translocator 1 Impairs Photorespiration and Plant Growth. *Front Plant Sci*. 2019;10:1726.

8. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185-193.

9. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*. 2003;31(4):e15.

10. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4(2):249-264.

11. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*. 2001;98(9):5116-5121.

12. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res*. 2001;125(1-2):279-284.

13. Team RDC. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0,. *URL* http://wwwR-projectorg/. 2010.

14. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004;5(10):R80.

15. Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;34(3):267-273.

16. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550.

17. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417-425.

18. Chiaretti S, Li X, Gentleman R, et al. Gene expression profiles of B-lineage adult acute lymphocytic leukemia reveal genetic patterns that identify lineage derivation and distinct mechanisms of transformation. *Clin Cancer Res*. 2005;11(20):7209-7219.

19. Juric D, Lacayo NJ, Ramsey MC, et al. Differential gene expression patterns and interaction networks in BCR-ABL-positive and -negative adult acute lymphoblastic leukemias. *J Clin Oncol*. 2007;25(11):1341-1349.

20. Kohlmann A, Schoch C, Schnittger S, et al. Pediatric acute lymphoblastic leukemia (ALL) gene expression signatures classify an independent cohort of adult ALL patients. *Leukemia*. 2004;18(1):63-71.

21. Caporaso JG, Lauber CL, Walters WA, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*. 2012;6(8):1621-1624.

22. Gilbert JA, Jansson JK, Knight R. The Earth Microbiome project: successes and aspirations. *BMC Biol*. 2014;12:69.

23. Walters W, Hyde ER, Berg-Lyons D, et al. Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems*. 2016;1(1).

24. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* 2011;17:10.

25. Gonzalez A, Navas-Molina JA, Kosciolek T, et al. Qiita: rapid, web-enabled microbiome meta-analysis. *Nat Methods*. 2018;15(10):796-798.

26. Amir A, McDonald D, Navas-Molina JA, et al. Correcting for Microbial Blooms in Fecal Samples during Room-Temperature Shipping. *mSystems*. 2017;2(2).

27. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37(8):852-857.

28. Jiang L, Amir A, Morton JT, Heller R, Arias-Castro E, Knight R. Discrete False-Discovery Rate Improves Identification of Differentially Abundant Microbes. *mSystems*. 2017;2(6).

29. Xu ZZ, Amir A, Sanders J, et al. Calour: an Interactive, Microbe-Centric Analysis Tool. *mSystems*. 2019;4(1).

30. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13(7):581-583.

31. Callahan BJ, Wong J, Heiner C, et al. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Res*. 2019;47(18):e103.

32. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772-780.

33. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One*. 2010;5(3):e9490.

34. Janssen S, McDonald D, Gonzalez A, et al. Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information. *mSystems*. 2018;3(3).

35. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol*. 2005;71(12):8228-8235.

36. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 2008;26:32.

37. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011;12:2825-2830.

38. Jones E, Oliphant T, Peterson P. SciPy: Open Source Scientific Tools for Python. 2001:http://www.scipy.org/

39. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30(1):207-210