Deletion of genes encoding PU.1 and Spi-B leads to B cell acute lymphoblastic leukemia associated with driver mutations in Janus Kinases

Carolina R. Batista^{*†}, Michelle Lim^{*†}, Anne-Sophie Laramée^{*†}, Faisal Abu-Sardana^{*†}, Li S. Xu ^{*‡}, Rajon Hossain^{*†}, and Rodney P. DeKoter^{*†}

^{*}Department of Microbiology & Immunology and the Centre for Human Immunology, Schulich School of Medicine & Dentistry, Western University, London, Ontario, Canada.

[†]Division of Genetics and Development, Children's Health Research Institute, Lawson Research Institute, London, Ontario, Canada.

This work was supported by the Canadian Institutes of Health Research Grants MOP-10651 and MOP-137414 (to R.P.D.), a grant from the Leukemia and Lymphoma Society of Canada (to R. P. D.), and an Ontario Trillium Scholarship (to C.R.B.)

Corresponding author: Rodney P. DeKoter, Department of Microbiology & Immunology, Schulich School of Medicine & Dentistry, Western University, London, Ontario, Canada N6A 5C1. Phone: (519) 661-2084; Fax: (519) 661-3499; E-mail address: rdekoter@uwo.ca

20 **Abstract** (200 words)

21 Precursor B-cell acute lymphoblastic leukemia (B-ALL) is associated with recurrent mutations that occur
 22 in cancer-initiating cells. There is a need to understand how spontaneous driver mutations influence clonal
 23 evolution in leukemia. The ETS-transcription factors PU.1 and Spi-B (encoded by *Spi1* and *Spib*) execute
 24 a critical role in B cell development and serve as complementary tumour suppressors by opposing the
 25 proliferative events mediated by IL-7R signaling. Here, we used a mouse model to conditionally delete
 26 *Spi1* and *Spib* genes in developing B cells. These mice developed B-ALL with a median time to euthanasia
 27 of 18 weeks. We performed RNA and whole-exome sequencing (WES) on leukemias isolated from Mb1-
 28 CreΔPB mice and identified single-nucleotide variants (SNVs) in *Jak1*, *Jak3* and *Ikzf3* genes, resulting in
 29 amino acid changes and in the gain of early stop-codons. JAK3 mutations resulted in amino acid
 30 substitutions located in the pseudo-kinase (R653H, V670A) and in the kinase (T844M) domains.
 31 Introduction of these mutations into wild-type pro-B cells conferred survival and proliferation advantages.
 32 We conclude that mutations in Janus kinases represent secondary drivers of leukemogenesis in the absence
 33 of Spi-B and PU.1 transcription factors. This mouse model represents an useful tool to study clonal
 34 evolution and tumour heterogeneity in B-ALL.

35 Introduction

36 Acute lymphoblastic leukemia is the most common type of childhood cancer, with approximately
 37 6000 new cases diagnosed in the United States each year¹. Most leukemias originate within the B cell
 38 rather than the T cell lineage^{2,3}. Precursor B cell acute lymphoblastic leukemia (pre-B-ALL) is a disease
 39 that is revealed by the presence of transformed precursor B cells in the blood, bone marrow, and tissues;
 40 and is most common in 1-5 year old patients⁴. Most pre-B-ALL cases are associated with genetic
 41 abnormalities that include chromosomal translocations or point mutations. In pre-B-ALL, up to two thirds
 42 of genes with point mutations encode transcriptional regulators such as Pax-5, Ikaros, or EBF1³. Pre-B-
 43 ALL cells are frequently arrested at an early stage of development, express interleukin-7 receptor (IL7R),
 44 and have high levels of Janus Kinase (JAK)-STAT signaling to sustain survival and proliferation⁵⁻⁷.
 45 Activating mutations of the *IL7R* gene have been described in human pre-B-ALL⁸. *JAK* and *IL7R*
 46 mutations are frequent in several subtypes of pre-B-ALL including the recently described disease Ph-like
 47 leukemia^{9,10}. In summary, mutations that activate cytokine signaling, and impair differentiation, function
 48 as driver mutations in pre-B-ALL.

49 PU.1 (encoded by *SP11*) and Spi-B (encoded *SPIB* in mice) are transcription factors of the E26-
 50 transformation-specific (ETS) family¹¹. These two proteins share a conserved DNA binding domain and
 51 interact with an overlapping set of DNA binding sites within the genome¹². PU.1 and Spi-B complement
 52 one another's function, and activate multiple genes involved in B cell receptor signaling¹²⁻¹⁵. Lack of these
 53 factors in developing B cells results in a block to B development at the small pre-B cell stage associated
 54 with impaired *Ig* light chain rearrangement^{15,16}. Importantly, conditional deletion of Spi-B and PU.1 in
 55 developing B-cells leads to high incidence of B-ALL in mice, but the mechanisms of leukemogenesis in
 56 the absence of these transcription factors are still undetermined¹⁷. In conclusion, PU.1 and Spi-B are
 57 required for B-cell development, and function as complementary tumour suppressors in the B cell lineage.

58 B cell neoplasms, like all cancers, are thought to be diseases in which there is clonal evolution
 59 from a common precursor, in which acquired gene mutations drive an evolutionary natural selection
 60 process^{18,19}. The mechanisms by which cancer-initiating cells respond to selection pressures during clonal
 61 evolution have been classified into a number of common hallmarks²⁰. In response to selection pressure,
 62 the genetic makeup of cancer-initiating cells changes during the course of disease due to acquired
 63 mutation. Mutations can be broadly classified as drivers or passengers^{18,21}. Driver mutations are those that
 64 provide a growth advantage to a cancer clone, whereas passenger mutations do not provide a growth
 65 advantage. Pediatric B-ALL is less curable upon relapse due to clonal evolution of the leukemia, resulting
 66 in driver mutations inducing a more aggressive disease²². High levels of intratumoral heterogeneity of
 67 mutations is a poor prognostic marker for leukemia²³. Whole-exome or whole-genome sequencing of pre-
 68 B-ALL cases is expected to lead to a deeper understanding of the genetic causes of this disease, ultimately
 69 permitting molecular targeted therapy for individual patients².

70 In this study, we investigated the molecular features of leukemogenesis in a spontaneous model of
 71 B-ALL induced by deletion of genes encoding PU.1 and Spi-B. We generated *Mb1^{+/Cre}Spi1^{lox/lox}Spib^{-/-}*
 72 mice, called here Mb1-CreΔPB mice¹⁶. We found that Mb1-CreΔPB mice developed pre-B-ALL
 73 characterized by the high expression of IL-7R, with a median time to euthanasia of 18 weeks. Using
 74 whole-exome sequencing (WES) and RNA-seq, we identified single-nucleotide variants (SNVs), most of
 75 which were predicted to have a role in the control of cell proliferation, communication and metabolism.
 76 Strikingly, we identified recurrent SNVs in genes encoding Aiolos, Jak1, and Jak3 in mouse leukemias.
 77 Further analysis revealed that SNVs located in *Jak3* resulted in three different types of amino acid
 78 substitutions within the pseudo-kinase domain (R653H, V670A) and kinase domain (T844M). We
 79 confirmed the ability of these mutations to provide survival and proliferation advantages to normal pro-B
 80 cells. In summary, this study shows that *Jak3* mutations are secondary drivers of leukemogenesis in the

- 81 absence of Spi-B and PU.1. This mouse model may be useful to determine the effects of molecular targeted
82 therapies on intratumoral heterogeneity and clonal evolution in B-ALL.

83 **Materials and Methods**

84 **Mice and breeding**

85 Mb1-Cre mice have been described previously²⁴. For this study, Mb1-Cre mice were crossed with
 86 *Spi1*^{lox/lox} *Spib*^{-/-} to generate *Mb1*^{+/-Cre} *Spi1*^{lox/lox} *Spib*^{-/-} mice (also referred to as Mb1-CreΔPB mice).
 87 C57BL/6 mice were purchased from Charles River Laboratories (Saint-Constant, QC, Canada).
 88 *Mb1*^{+/-Cre}*Spi1*^{lox/lox} *Spib*^{+/+} (referred to as Mb1-CreΔP) and *Mb1*^{+/-Cre}*Spi1*^{+/+} *Spib*^{-/-} or *Mb1*^{+/+}*Spi1*^{+/+} *Spib*^{-/-}
 89 (referred to as Mb1-CreΔB) mice were used as experimental controls. Mice were fed with regular chow
 90 and tap water ad libitum and housed with a 12-h light–dark cycle. Mice showing signs of illness were
 91 euthanized and examined for spleen, thymus and lymph node enlargement and analyzed as described
 92 below. All experiments were performed in compliance with the Western University Council on Animal
 93 Care.

95 **Histology and microscopic analysis**

96 Mb1-CreΔPB mice showing signs of disease were euthanized by CO₂ and spleen and thymus were
 97 removed for histological analysis. Spleen and thymus removed from Mb1-CreΔB mice were used as
 98 control. Organs were fixed in 10% buffered formalin. Tissues were paraffin embedded, sectioned, and
 99 stained with hematoxylin and eosin. High-resolution micrographs were captured using a Q-Color3 digital
 100 camera (Olympus, Markham, ON, Canada).

102 **Flow Cytometry**

103 Single cell suspensions were prepared from enlarged spleen and thymus from Mb1-CreΔPB mice.
 104 Red blood cells were removed from single-cell suspensions using hypotonic lysis. Flow cytometric
 105 analyses were performed using a FACSCanto or LSRII instrument (BD Immunocytometry Systems, San

Jose, CA). Antibodies were purchased from eBioscience (San Diego, CA), BioLegend (San Diego, CA), or BD Biosciences (Mississauga, ON, Canada) and included PE-anti-CD19 (1D3), FITC-anti-BP-1 (6C3), APC-anti-IgM (II/ 41), PE-anti-Igk (187.1), FITC-anti-IL-7R α (A7R34), BV421-anti-B220 (RA3-6B2), PE-anti-BP-1 (6C3). Data were analyzed using FlowJo 9.1 software (Tree Star, Ashland, OR).

Whole-Exome Sequencing

Three matched-tumour and tail mouse samples (853, 854 and 857) had genomic DNA isolated using Wizard Genomic DNA Purification Kit (Promega Corporation, Madison WI). Whole-Exome Sequencing was performed by McGill University and Génome Québec Innovation Centre (Montreal, Canada). The SureSelectXT Mouse All Exon kit (Agilent Technologies, Mississauga, Canada) was used to perform exome target capture. Paired-end DNA library preparation was performed using SureSelectXT Target Enrichment System. An Illumina HiSeq4000 instrument was used to perform the whole-exome sequencing. BAM files were converted to Fastq format on Galaxy suite²⁵ using “SAMtoFastq” tool. Fastq reads had Illumina Adaptors were removed using TrimGalore!. Trimmed reads were aligned to the mouse reference genome (mm10) using Bowtie. BAM aligned files were used as input for single somatic variant callers. Somatic variants of three matched-tumour mouse samples were called by using two independent methods, Strelka and Mutect^{26,27}. For both methods, variants were called using the standard settings recommend by the package documentation. VCF files containing the somatic nucleotide variants (SNVs) and small insertions and deletions (INDELs) generated by Strelka and Mutect were further filtered following the criteria ‘passed’ or ‘keep’ for the Strelka and Mutect outputs, respectively. Filtered VCF files outputted by Strelka and Mutect were intersected using VCFtools and annotated using the SnpEff²⁸ package using the mouse genomic annotation (mm10). Annotated variants were classified according the variant impact in HIGH, MODERATE, MODIFIER and LOW, and overlapping variants among the three

samples were determined. Independently, outputs generated by Strelka and Mutect were also annotated using the SnpEff tool for further analysis.

RNA Sequencing

RNA was extracted from thymic tumours 853, 854 and 857 using RNeasy Kit (QIAGEN, ON, Canada). RNA-sequencing was performed by McGill University and Génome Québec Innovation Centre (Montreal, Canada). Paired-end (mRNA-sequencing stranded) libraries were prepared using Truseq stranded total RNA library prep kit. Libraries were sequenced using an Illumina HiSeq4000 PE100 instrument. Data analysis was performed as described previously¹⁶. In summary, Standard Illumina sequencing adaptors were removed using Trim Galore!. Trimmed Fastq reads were aligned to the mouse reference genome version mm10 using TopHat2²⁹ in mate-paired mode using Galaxy suite. Assembled transcript abundance was determined as previously using Cufflinks²⁹.

Sanger Sequencing

RNA extracted from 20 thymic leukemias was reversed transcribed to complementary DNA using iScript cDNA synthesis kit (Bio-Rad, Hercules, CA). Primers flanking the *Ikzf3*, *Jak3* and *Jak1* mutations identified by WES were designed using Primer-BLAST. cDNA was used as a template for PCR reactions using specific primers (Sequences described in **Supplementary Information**). PCR reactions were run in a 1.5% agarose gel and the PCR products were excised and purified from the gel using QIAquick Gel Extraction Kit (QIAGEN, ON, Canada). Purified DNA was submitted for Sanger sequencing at London Regional Genomics Centre, Western University (London, ON, Canada). Sequencing chromatograms and DNA sequences were analyzed in the software 4Peaks (Mekentosj, Amsterdam).

DNA Constructs and Site-Directed Mutagenesis

The mouse MSCV-IRES-GFP-JAK3 plasmid was kindly provided by Dr. Kevin D. Bunting. T844M, R653H, and V670A mutations were generated by site-directed mutagenesis using Q5® Site-Directed Mutagenesis Kit (New England BioLabs, Ipswich, MA). Site-directed mutagenesis primers for each one of the specific mutations were designed using NEBaseChanger tool. Mutations were confirmed by Sanger Sequencing before retroviral production and spinoculation.

Retroviral production and infection

Retroviral supernatants for wild-type Jak3, and T844M, R653H, V670A mutants were generated Platinum-E retroviral packaging cells³⁰ using PEIPro transfection reagent (PolyPlus, New York, NY). 10 µg of MSCV plasmid were cotransfected with 4 µg pECO plasmid (Clontech, Mountain View, CA). Plasmid DNA was mixed with polyethyleneimine (PEI) at a 2:1 PEI/DNA ratio and applied to 3×10^6 cells in a 10-cm plate. The medium was changed after overnight incubation, and virus-containing supernatant was collected 48 hours after medium change. Wild-type pro-B cells were infected by spinoculation with 1 mL of viral supernatant containing 10ul of polybrene at the concentration of 10mg/ml. Cells were centrifuged at $3000 \times g$ for 2 hours at 30°C and allowed to recover for 24 hours before medium change and infection was confirmed by flow cytometric analysis for GFP.

Cell Culture

Wild-type and Jak3 mutant-infected pro-B cell lines used in this study were cultured in IMDM (Wisent, QC, Canada) containing 10% FBS (Wisent), 1X penicillin/streptomycin/L-glutamine (Lonza, Shawinigan, QC, Canada), and 5×10^{-5} M β -mercaptoethanol (Sigma-Aldrich St. Louis, MO). Media also

174 contained 5% or 0.5% conditioned medium from the IL-7 producing cell line J558-IL-7³¹. Cell lines were
175 maintained in 5% CO₂ atmosphere at 37°C.

176

177 **Availability of data**

178 WES data is available from the Sequence Read Archive, accession SRP136503
179 (<https://www.ncbi.nlm.nih.gov/sra/SRP136503>). RNA-seq data is available from the Gene Expression
180 Omnibus, accession GSE112506 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112506>).

181

182 **Statistical Analysis**

183 All data reported in this study were graphed as mean ± SEM. Statistical analysis was performed
184 using Prism 5.0 (Graphpad Software, La Jolla, CA) using ANOVA or Student's *t* test. Values with $p \leq$
185 0.05 were considered significant, where $p \leq 0.05$ (*), $p \leq 0.01$ (**), $p \leq 0.001$ (***), $p \leq 0.0001$ (****).

186 Results

187 Deletion of genes encoding PU.1 and Spi-B leads to B-cell acute lymphoblastic leukemia

188 We recently reported that deletion of the genes encoding both PU.1 and Spi-B in B cells under control
 189 of the *Cd79a* (*Mb1*) promoter (*Mb1^{+/-}Cre Spi1^{lox/lox} Spib^{-/-}* mice, abbreviated to Mb1-CreΔPB) resulted in a
 190 severe impairment to B cell development at the large pre-B to small pre-B cell transition in the bone marrow
 191 of 6-10 week-old mice¹⁶. Mb1-CreΔPB mice that express Cre recombinase and are deleted for Spi-B but are
 192 homozygous for the wild type *Spi1* allele were fertile and healthy. In contrast, Mb1-CreΔPB mice had a
 193 median survival of 18 weeks, at which point they required euthanasia due to signs of illness, including
 194 lethargy and laboured breathing (**Fig. 1A, 1B**). Dissection of euthanized mice revealed enlargement of the
 195 spleen and thymus (**Fig. 1A, 1C, 1D**). Histological analysis revealed that normal spleen and thymus
 196 organization was completely effaced in moribund Mb1-CreΔPB mice compared to the controls (**Fig. 1E**).

197 Flow cytometry analysis was used to determine the stage in which leukemic cells begin to infiltrate
 198 the spleen and thymus of pre-leukemic Mb1-CreΔPB mice. We observed the presence of B220⁺CD19⁺ cells
 199 in the spleen and thymus of Mb1-Cre ΔPB mice aged 11 weeks, but not in either C57BL/6 or Mb1ΔB mice
 200 (**Fig. 2A and 2B**). At the time of euthanasia of leukemic Mb1-Cre ΔPB mice aged >15 weeks there were
 201 high frequencies of B220⁺CD19⁺ cells in both organs (**Fig. 2A and 2B**). Further analysis seeking to
 202 determine the phenotype of cells infiltrating the spleen and thymus showed that these organs contained high
 203 frequencies of B220⁺ and CD19⁺ cells that also expressed c-Kit, CD43, BP-1, and IL-7Rα (**Fig. 2D** and data
 204 not shown). This analysis suggested that these cells were a pre-B cell-like acute lymphoblastic leukemia
 205 similar to that previously reported in CD19-CreΔPB mice¹⁷ or mice deleted for genes encoding PU.1 and
 206 IRF4/8³². Next, leukemic cells obtained from the spleen or thymus of Mb1-CreΔPB mice were characterized
 207 to determine expression of cell surface IgM and Igκ. Analysis of leukemias from Mb1-CreΔPB mice revealed
 208 that at least 74% of individual mouse leukemias did not express IgM or Igκ at the cell surface (**Fig. 2E**,

209 **bottom; 2F, left panel**). 26% of leukemias from Mb1-Cre Δ PB mouse spleen or thymus had cell surface
 210 expression of both IgM and Ig κ (**Fig. 2E, top; 2F, left panel**). In contrast, analysis performed on CD19-
 211 Cre Δ PB mice showed that 78% of leukemias had IgM and Ig κ on the cell surface (**Fig. 2F, right panel**).
 212 Most leukemias isolated from Mb1-Cre Δ PB mice had *IgH* rearrangements, although fewer leukemias
 213 expressed IgM at the cell surface (**Fig. 2C**). In summary, these results show that leukemia cells begin to
 214 appear in thymus and spleen of Mb1-Cre Δ PB mice after 11 weeks of age. All leukemias in Mb1-Cre Δ PB
 215 mice at 11-26 weeks of age expressed IL-7R, but most did not express Ig on their cell surface, suggesting
 216 that these cells resembled pro-B or large pre-B cells.

217

218 **Whole-exome sequencing (WES) identifies somatic nucleotide variants in leukemias generated in the** 219 **lack of Spi-B and PU.1**

220 As described above, Mb1-Cre Δ PB mice had a variable time to requirement of euthanasia as well
 221 as heterogeneity in Ig expression status of leukemias (**Fig. 1 and 2**). This variability suggested that
 222 secondary driver mutations are required to induce leukemia in addition to the initiating lesion of *Spi1/SpiB*
 223 deletion. In order to discover potential driver mutations in Mb1-Cre Δ PB mice leukemias, we performed
 224 whole-exome sequencing (WES) analysis of three mouse tumours, 853, 854, and 857, as well as on
 225 matched genomic tail DNA as a control (**Supplementary Fig. 1**, data available in **Supplementary**
 226 **Information**). Single Nucleotide Variants (SNVs) for each one of the samples were identified by
 227 comparing leukemia and control tail DNA exome sequences using the Strelka somatic variant caller²⁶.
 228 The number of SNVs identified by Strelka differed for each one of the tumours, where sample 854 showed
 229 the highest number of SNVs (3,887) (**Fig. 3A-C**). Samples 853 and 857 presented a very similar number
 230 of SNVs, totaling 2,558 and 2,465, respectively. Although the tumours had a variable number of SNVs,
 231 the distribution of the SNVs by chromosome followed a similar pattern in which chromosome 2 showed

the highest frequency of SNVs, followed by chromosome 11, chromosome 1 and chromosome 7 (Supplementary Fig. 2A-C). The majority of the nucleotide alterations in the three tumours analyzed were C•G → A•T transversions (Supplementary Fig. 2D). In order to gain insight into the nature of the mutational processes in Mb1-CreΔPB mice, we determined the mutational signature of leukemias using DeconstructSigs³³. This package compares the similarity of each tumour sample to a list of previous published signatures generated by the analysis of 40 distinct types of human cancer³⁴. This analysis confirmed enrichment of C•G → A•T transversions in leukemias 853, 854 and 857 (Supplementary Fig. 3A-C). Mutational signature analysis performed using DeconstructSigs indicated that the signatures in leukemias 853, 854, and 857 was most similar to COSMIC Signatures 18, 24, 4, and 9 (Supplementary Fig. 3A-C). Signature 18, which showed the highest inferred weight for all leukemia samples analyzed, has been also commonly observed in human childhood cancers including B-ALL and neuroblastoma, but does not have a known mechanism³⁴. We also observed an enrichment of C→A transversions flanked 5' base by either adenine (A), cytosine (C), guanine (G) or thymine (T) nucleotides. These transversions were consistently flanked at the 3' base by an adenine (A) nucleotide (Supplementary Fig. 3D).

In order to increase the confidence in the SNVs identified by Strelka and to reduce the number of SNVs without a potential biological function, somatic variants were called with a second variant discovery method, Mutect²⁷. Mutect detected fewer SNVs than Strelka (Fig. 3A-C). The overlap of SNVs identified by both methods was used to determine high confidence SNVs. We found that by using this strategy the number of SNVs was reduced, resulting in 183 SNVs for sample 853, 293 SNVs for sample 854, and 199 SNVs for sample 857. High confidence SNVs were annotated using the mouse genome (*mm10*) using the SnpEff tool, which also enables the prediction of the biological effect of the variants²⁸. We found that the majority of the high confidence SNVs were classified by as having or a MODIFIER or MODERATE deleterious impact while a few SNVs were classified as HIGH (Fig. 3D). We next looked at the effect of

SNVs classified as MODERATE and HIGH, and found that most of the SNVs classified as MODERATE resulted in missense variants; while the few SNVs predicted as having HIGH impact caused the gain or loss of stop codons, or alterations in splice acceptor or donor sites, defined by the two bases before exon start or ends, respectively (**Fig. 3E**). We next investigated whether there were common SNVs among the three mouse tumours sequenced. Samples 854 and 857 showed common SNVs in the genes *Sntg1*, *Brd8*, *Lrp5* and *Slco1b2* (**Fig. 3F**). Samples 853 and 854 showed common SNVs in the genes *Ryr2* and *Skint6*. Finally, samples 853 and 857 showed common SNVs in the genes *Ikzf3* and *Jak3* (**Fig. 3F**). Interestingly, *Ikzf3* and *Jak3* are well-characterized driver mutations in human pre-B cell acute lymphoblastic leukemia³⁵. The identification of common *Ikzf3* and *Jak3* SNVs suggests that this approach is effective at identifying potential driver mutations.

Mutations in Janus Kinase 1 and 3 (*Jak1/Jak3*) and Aiolos (*Ikzf3*) genes are potential secondary drivers of leukemogenesis in Mb1-Cre Δ PB mice

As cancer driver genes would be expected to contain mutations and be expressed²¹, we also performed RNA-sequencing analysis on leukemias 853, 854, and 857 in order to assess gene expression levels. To filter data based on levels of gene expression, Fragments Per Kilobase of transcript per Million mapped reads (FPKM) was determined from RNA-seq data using Cufflinks. FPKM was plotted against VAF determined using Strelka²⁶ for three mouse leukemias 853, 854, and 857. In each of the three samples there were highly expressed genes that also had high VAF (**Fig. 4A-C**). Importantly, sample 853 had high FPKM and VAF for variants in *Jak3* and *Ikzf3* (**Fig 4A**). Sample 854 had high FPKM and VAF for a variant in *Jak1* (**Fig 4B**). Finally, sample 857 had high FPKM and VAF for variants in *Jak3* and *Ikzf3* (**Fig 4C**). The elevated expression levels added to the high VAF of *Jak1*, *Jak3* and *Ikzf3* genes in Mb1-Cre Δ PB mice leukemias supports the hypothesis that these variants represent secondary driver mutations.

Next, we sought to define the biological processes in which mutated genes could be involved by performing gene ontology analysis in genes with VAF greater than 20%, using PANTHER Gene List Analysis³⁶. We observed a pattern for all three samples where most of the genes were categorized in two major biological processes, “Cellular Process” and “Metabolic Process”. Specifically, genes within “Cellular Process” were subcategorized in sub processes as cell communication, cell proliferation and cell cycle (**Fig. 4D-F**). Genes as *Cit*, *Cdk9* and *Stag1* are related to the control of cell cycle by regulating cytokinesis, transcription and cohesion of sister chromatids after DNA replication, respectively. Importantly, *Smarchb1* encodes a core subunit protein of the ATP-dependent SWI/SNF chromatin remodeling complex, and was previously identified as a tumor suppressor³⁷. *Nap1l4*, encodes a member of the nucleosome assembly protein and has a role as a histone chaperone. Genes involved in cell communication and adhesion, as *Cntnap5b*, *Fras1*, *Sdk1* and *Magi*, were also enriched in our pathway analysis. In summary, *Jak3* and *Ikzf3* mutations were identified as potential drivers of leukemogenesis in Mb1-CreΔPB mice, based on the observation that these genes are mutated in two of three leukemias, have high VAF, and have high levels of expression.

Recurrent mutations in Janus Kinase 3 (*Jak3*) and Janus Kinase 1 (*Jak1*) in leukemias from Mb1-CreΔPB mice

Analysis of the impact of variants on protein expression showed that SNVs in leukemias 853 and 857 resulted in coding changes in *Jak3* (R653H, V670A, and/or T844M) while the SNV in *Jak1* in sample 854 resulted in a V657F substitution (**Fig. 5A, 5B**). These genes encode for Jak1 and Jak3 that are critically required for signaling through cytokine receptors IL7R and CRLF2³⁸. Comparison of Jak3 and Jak1 protein sequences among six different vertebrates showed that amino acids that underwent substitution in consequence of a SNV were highly conserved between these species (**Fig. 5A, 5B**). Sanger sequencing of samples 853, 854 and 857 confirmed the presence of SNVs identified by WES in the *Jak3*, *Jak1* and *Ikzf3*

genes (**Supplementary Fig. 3**). To determine whether these variants represent recurrent leukemia driver mutations, we performed Sanger sequencing on a total of 19 Mb1-Cre Δ PB leukemias using primers flanking the regions containing the SNVs identified by WES. Interestingly, this examination revealed that Jak3 R653H mutation was detectable in 5/19 leukemias analyzed. *Jak1* mutations V657F or V655L showed even higher recurrence than Jak3 mutation, being detectable in 10/19 leukemias (**Fig. 5D**). *Irf3* mutations were detected only in samples 853 and 857, confirming the exome sequencing (**Fig. 5C, 5D**). In summary, these data show that leukemia in Mb1-Cre Δ PB mice is accompanied by recurrent secondary driver mutations in *Jak1* and *Jak3*.

R653H, V670A, and T844M amino acid substitutions in Jak3 confers survival and proliferation advantages

We next tested whether R653H, V670A, and T844M amino acid substitutions in Jak3 were able to confer growth advantage to normal cells. Site-directed mutagenesis was used to introduce these mutations to the wild-type mouse *Jak3* coding region in an MSCV-IRES-GFP vector (a kind gift from Dr. Kevin D. Bunting). Pseudovirus was generated to spin-infect wild-type pro-B cells that grow in cultures containing IL-7 and ST2 stromal cells. Pro-B cells were also infected with MSCV-Jak3 and a MSCV- empty vectors as controls. Following infection, the cell lines were cultured at high (5% conditioned medium) and low IL-7 (0.5%). Flow cytometry was performed every two days for 14 days to determine the frequency of GFP⁺ cells. Pro-B cells infected with *Jak3* mutations R653H, V670A, or T844M outgrew uninfected or control infected pro-B cells at low and high IL-7 concentration (**Fig. 6A, 6B**). Next, in order to verify whether *Jak3* mutations resulted in greater proliferation, cells infected with mutant Jak3 were plated at the initial concentration of 1x10⁵ cells/ml in conditions of low and high IL-7. Cell numbers and the percentage of viable GFP⁺ cells were determined after cells were kept in culture for the period of four days. At low IL-7 concentration, pro-B cells

324 infected with *Jak3* R653H, V670A, and T844M mutations showed a significant increase in the number of
 325 GFP⁺ cells after four days in culture (**Fig. 6C, 6D**). Therefore, we conclude that T844M, R653H and V670A
 326 amino acids substitutions were able to increase the proliferation of pro-B cells in conditions of low IL-7.
 327 Taken together, these data indicate that *Jak3* mutations identified in this analysis confer survival and growth
 328 advantages to pro-B cells, consistent with a role as driver mutations.

329 Discussion

330 In this study, we showed that Mb1-Cre Δ PB mice develop precursor B-ALL that is detectable in
 331 mice by 11 weeks of age and results in a requirement for euthanasia in 100% of mice by median 18 weeks.
 332 The latency of 11-18 weeks, and the variable frequency of leukemias expressing Ig, show that there is
 333 clonal variability, suggesting that secondary driver mutations are required for disease progression. By
 334 characterizing the mutation landscape of Mb1-Cre Δ PB leukemias using whole exome sequencing coupled
 335 with RNA-sequencing, we discovered that mutations in *Jak1* and *Jak3* represent recurrent secondary
 336 driver mutations. Introduction of mutant *Jak3* into wild type cells was sufficient to promote growth in
 337 response to low IL-7 concentrations. Taken together, these results show that deletion of genes encoding
 338 PU.1 and Spi-B result in leukemia by predisposition to additional *de novo* driver mutations in genes
 339 encoding Janus kinases.

340 Dysregulation of PU.1 and/or Spi-B expression are known to be involved in human leukemia and
 341 have been established as tumor suppressors in mouse models^{17,39}. Minimal reductions in PU.1 expression
 342 are sufficient to induce a preleukemic condition in hematopoietic stem cells⁴⁰. Whole exome sequencing
 343 and whole genome sequencing studies have revealed that inactivating mutations in genes encoding PU.1
 344 and Spi-B are detectable and infrequent in human leukemia^{41,42}. In contrast, PU.1 levels are repressed by
 345 frequently occurring chromosomal abnormalities such as FLT3 internal tandem duplication (FLT3-ITD)⁴³
 346 and RUNX1-ETO fusion⁴⁴. Spi-B levels are repressed by ETV6-RUNX1 chromosomal fusion³⁹. It is not
 347 known how reduced PU.1 and/or Spi-B might lead to acquisition of secondary driver mutations. However,
 348 PU.1 and Spi-B are involved in attenuating IL-7R signaling in developing pro- and pre-B cells. Target
 349 genes of PU.1 and Spi-B involved in regulating IL-7-induced proliferation include *Btk* encoding the
 350 tumour suppressor Bruton tyrosine kinase¹⁵, and *Blmk* encoding B Cell Linker Protein⁴⁵. Btk and Blmk
 351 work together to attenuate IL-7R signaling in developing B cells⁴⁶. We speculate that mutagenesis in Mb1-

Cre-ΔPB is associated with genomic instability due to uncontrolled cell proliferation. Interestingly, a similar disease is induced by combined deletion of PU.1 and IRF4, suggesting a common axis of gene regulation by PU.1/Spi-B and PU.1/IRF4³².

We identified a number of single nucleotide variants (SNVs) in our analysis, which suggest that several rounds of mutational events contributed to leukemogenesis. To select the potential leukemia initiator lesions, we focused our attention in genes that were highly expressed and had high variant allelic frequencies (VAF), reasoning that genes within this category are likely to have originated earlier in leukemia progression. This analysis identified Aiolos (*Ikzf3*), Jak1 (*Jak1*) and Jak3 (*Jak3*) as potential secondary driver mutations. These genes are involved in signaling downstream of cytokine receptors during B cell development. Cytokine receptor genes are commonly mutated in human B-cell acute lymphoblastic leukemias⁴⁷, and mutations in *Jak3* and *Ikzf3* have recently emerged as novel mutated genes in high-risk B-ALL⁴⁸. In particular, our analysis focused on R653H, V670A, and T844M mutations in *Jak3*. Jak3 R653H and V670A mutations were recently observed in B-ALL generated in activation-induced cytidine deaminase (AID) deficient mice⁴⁹. Human equivalents of Jak3 R653H and V670A have been described in human ALL (R657Q, V674F), and furthermore have been shown to function by activation of proliferation in response to interleukin-7⁵⁰. Importantly, human JAK3 V674F is sufficient to induce T-cell acute lymphoblastic leukemia when retrovirally delivered to hematopoietic precursors⁵⁰. *Jak1* missense mutations V655L or V657F were detectable in 10 of 19 leukemias in our analysis. *Jak1* V657F also corresponds to a human *JAK1* mutation that is a frequent driver in human B-ALL (V658F)⁵¹. Interestingly *JAK1* V658F mutation has been previously observed in T-cell acute lymphoblastic and acute myeloid leukemias⁵². V658F is also thought to be paralogous to the JAK2 V617F mutation that functions as a driver in more than 95% of cases of polycythemia vera⁵³. In summary, the mutations identified in this study in *Jak1* and *Jak3* are relevant to human leukemia.

375 Taken together, this study confirms that leukemia in Mb1-Cre Δ PB mice is accompanied by *de*
376 *novo* recurrent secondary driver mutations in the Janus Kinase signaling pathway. Mutations in the JAK
377 signaling pathway are recurrent in human leukemia including the recently discovered Ph-like
378 classification that represents a high priority for discovering new therapies^{9,10}. In order to study the genetic
379 clonal evolution that underlies diseases such as Ph-like leukemia, there is still a need for mouse models.
380 In order to advance understanding, mouse models should have the following characteristics: 1) They
381 should develop leukemias with high penetrance and reproducibility, 2) they should replicate the genetic
382 and molecular heterogeneity of tumors and involve *de novo* mutations 3) they should occur in immune
383 competent mice, and 4) they should mimic that clinical behaviour of human disease^{54,55}. The Mb1-Cre Δ PB
384 mouse model develops B-ALL with 100% penetrance by 18 weeks of age that is driven by heterogeneous
385 *de novo* driver mutations. This mouse model may be useful to determine the effects of molecular targeted
386 therapies on intratumoral heterogeneity and clonal evolution in B-ALL.
387

388 **Acknowledgements**

389 We thank Michelle Ho for assistance with genotyping of mice. We thank Michael Reth (Freiburg,
390 Germany) for providing Mb1-Cre mice. We thank Dr. Kevin D. Bunting (Winship Cancer Institute,
391 Atlanta, USA) for providing the MSCV-IRES-GFP-JAK3 vector. We also thank scientists and staff of the
392 McGill University and Genome Quebec Innovation Centre for performing library construction and next-
393 generation sequencing. We also thank Kristin Chadwick and the London Regional Flow Cytometry Core
394 Facility for assistance with flow cytometric analysis.

395

396 **Conflict of Interest**

397 The authors have no financial conflicts of interest.

References

- 1 Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin* 2014; **64**: 9–29.
- 2 Hunger SP, Mullighan CG. Redefining ALL classification: toward detecting high-risk ALL and implementing precision medicine. *Blood* 2015; **125**: 3977–3987.
- 3 Inaba H, Greaves M, Mullighan CG. Acute lymphoblastic leukaemia. *Lancet* 2013; **381**: 1943–1955.
- 4 Pui CH, Robison LL, Look AT. Acute lymphoblastic leukaemia. *Lancet* 2008; **371**: 1030–1043.
- 5 Buchner M, Swaminathan S, Chen Z, Muschen M. Mechanisms of pre-B-cell receptor checkpoint control and its oncogenic subversion in acute lymphoblastic leukemia. *Immunol Rev* 2015; **263**: 192–209.
- 6 Muschen M. Rationale for targeting the pre-B-cell receptor signaling pathway in acute lymphoblastic leukemia. *Blood* 2015; **125**: 3688–3693.
- 7 Mangolini M, de Boer J, Walf-Vorderwulbecke V, Pieters R, den Boer ML, Williams O. STAT3 mediates oncogenic addiction to TEL-AML1 in t(12;21) acute lymphoblastic leukemia. *Blood* 2013; **122**: 542–549.
- 8 Shochat C, Tal N, Bandapalli OR, Palmi C, Ganmore I, te Kronnie G *et al*. Gain-of-function mutations in interleukin-7 receptor-alpha (IL7R) in childhood acute lymphoblastic leukemias. *J Exp Med* 2011; **208**: 901–908.
- 9 Jain N, Roberts KG, Jabbour E, Patel K, Eterovic AK, Chen K *et al*. Ph-like acute lymphoblastic leukemia: a high-risk subtype in adults. *Blood* 2017; **129**: 572–581.
- 10 Roberts KG, Yang Y-L, Payne-Turner D, Lin W, Files JK, Dickerson K *et al*. Oncogenic role and therapeutic targeting of ABL-class and JAK-STAT activating kinase alterations in Ph-like ALL.

- 421 *Blood Adv* 2017; **1**: 1657–1671.
- 422 11 Ray D, Bosselut R, Ghysdael J, Mattei MG, Tavitian A, Moreau-Gachelin F. Characterization of
423 Spi-B, a transcription factor related to the putative oncoprotein Spi-1/PU.1. *Mol Cell Biol* 1992; **12**:
424 4297–4304.
- 425 12 Solomon LA, Li SKH, Piskorz J, Xu LS, DeKoter RP. Genome-wide comparison of PU.1 and Spi-
426 B binding sites in a mouse B lymphoma cell line. *BMC Genomics* 2015; **16**: 76.
- 427 13 Garrett-Sinha LA, Su GH, Rao S, Kabak S, Hao Z, Clark MR *et al.* PU.1 and Spi-B are required
428 for normal B cell receptor-mediated signal transduction. *Immunity* 1999; **10**: 399–408.
- 429 14 Li SK, Abbas AK, Solomon LA, Groux GM, DeKoter RP. Nfkb1 activation by the E26
430 transformation-specific transcription factors PU.1 and Spi-B promotes Toll-like receptor-mediated
431 splenic B cell proliferation. *Mol Cell Biol* 2015; **35**: 1619–1632.
- 432 15 Christie DA, Xu LS, Turkistany SA, Solomon LA, Li SK, Yim E *et al.* PU.1 Opposes IL-7-
433 Dependent Proliferation of Developing B Cells with Involvement of the Direct Target Gene Bruton
434 Tyrosine Kinase. *J Immunol* 2015; **194**: 595–605.
- 435 16 Batista CR, Li SKH, Xu LS, Solomon LA, DeKoter RP. PU.1 Regulates Ig Light Chain
436 Transcription and Rearrangement in Pre-B Cells during B Cell Development. *J Immunol* 2017; **198**:
437 1565–1574.
- 438 17 Sokalski KM, Li SK, Welch I, Cadieux-Pitre HA, Gruca MR, DeKoter RP. Deletion of genes
439 encoding PU.1 and Spi-B in B cells impairs differentiation and induces pre-B cell acute
440 lymphoblastic leukemia. *Blood* 2011; **118**: 2801–2808.
- 441 18 Greaves M, Maley CC. Clonal evolution in cancer. *Nature* 2012; **481**: 306–313.
- 442 19 Nowell PC. The clonal evolution of tumor cell populations. *Science (80-)* 1976; **194**: 23–28.
- 443 20 Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011; **144**: 646–674.

- 444 21 Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LAJ, Kinzler KW. Cancer genome
445 landscapes. *Science* 2013; **339**: 1546–1558.
- 446 22 Ferrando AA, Lopez-Otin C. Clonal evolution in leukemia. *Nat Med* 2017; **23**: 1135–1145.
- 447 23 Hu Z, Sun R, Curtis C. A population genetics perspective on the determinants of intra-tumor
448 heterogeneity. *Biochim Biophys Acta* 2017; **1867**: 109–126.
- 449 24 Hobeika E, Thiemann S, Storch B, Jumaa H, Nielsen PJ, Pelanda R *et al.* Testing gene function
450 early in the B cell lineage in mb1-cre mice. *Proc Natl Acad Sci U S A* 2006; **103**: 13789–13794.
- 451 25 Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M *et al.* The Galaxy platform
452 for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res*
453 2016; **44**: W3–W10.
- 454 26 Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: Accurate somatic
455 small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 2012; **28**: 1811–
456 1817.
- 457 27 Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C *et al.* Sensitive detection
458 of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013; **31**:
459 213–219.
- 460 28 Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L *et al.* A program for annotating and
461 predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of
462 *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012; **6**: 80–92.
- 463 29 Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR *et al.* Differential gene and transcript
464 expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012; **7**: 562–
465 578.
- 466 30 Morita S, Kojima T, Kitamura T. Plat-E: an efficient and stable system for transient packaging of

retroviruses. *Gene Ther* 2000; **7**: 1063–1066.

31 Winkler TH, Rolink A, Melchers F, Karasuyama H. Precursor B cells of mouse bone marrow express two different complexes with the surrogate light chain on the surface. *Eur J Immunol* 1995; **25**: 446–450.

32 Pang SHM, Minnich M, Gangatirkar P, Zheng Z, Ebert A, Song G *et al.* PU.1 cooperates with IRF4 and IRF8 to suppress pre-B-cell leukemia. *Leukemia* 2016; **30**: 1375–1387.

33 Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol* 2016; **17**: 31.

34 Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin A V *et al.* Signatures of mutational processes in human cancer. *Nature* 2013; **500**: 415–421.

35 Iacobucci I, Mullighan CG. Genetic Basis of Acute Lymphoblastic Leukemia. *J Clin Oncol* 2017; **35**: 975–983.

36 Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* 2013; **8**: 1551–1566.

37 Tokheim CJ, Papadopoulos N, Kinzler KW, Vogelstein B, Karchin R. Evaluating the evaluation of cancer driver genes. *Proc Natl Acad Sci U S A* 2016; **113**: 14330–14335.

38 Babon JJ, Lucet IS, Murphy JM, Nicola NA, Varghese LN. The molecular regulation of Janus kinase (JAK) activation. *Biochem J* 2014; **462**: 1–13.

39 Pang SH, Minnich M, Gangatirkar P, Zheng Z, Ebert A, Song G *et al.* PU.1 cooperates with IRF4 and IRF8 to suppress pre-B-cell leukemia. *Leukemia* 2016; **30**: 1375–1387.

40 Will B, Vogler TO, Narayanagari S, Bartholdy B, Todorova TI, da Silva Ferreira M *et al.* Minimal PU.1 reduction induces a preleukemic state and promotes development of acute myeloid leukemia.

490 *Nat Med* 2015; **21**: 1172–1181.

491 41 Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD *et al.* Genome-wide
492 analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 2007; **446**: 758–764.

493 42 Mullighan CG, Zhang J, Kasper LH, Lerach S, Payne-Turner D, Phillips LA *et al.* CREBBP
494 mutations in relapsed acute lymphoblastic leukaemia. *Nature* 2011; **471**: 235–239.

495 43 Mizuki M, Schwable J, Steur C, Choudhary C, Agrawal S, Sargin B *et al.* Suppression of myeloid
496 transcription factors and induction of STAT response genes by AML-specific Flt3 mutations. *Blood*
497 2003; **101**: 3164–3173.

498 44 Vangala RK, Heiss-Neumann MS, Rangatia JS, Singh SM, Schoch C, Tenen DG *et al.* The myeloid
499 master regulator transcription factor PU.1 is inactivated by AML1-ETO in t(8;21) myeloid
500 leukemia. *Blood* 2003; **101**: 270–277.

501 45 Xu LS, Sokalski KM, Hotke K, Christie D a, Zarnett O, Piskorz J *et al.* Regulation of B cell linker
502 protein transcription by PU.1 and Spi-B in murine B cell acute lymphoblastic leukemia. *J Immunol*
503 2012; **189**: 3347–54.

504 46 Clark MR, Mandal M, Ochiai K, Singh H. Orchestrating B cell lymphopoiesis through interplay of
505 IL-7 receptor and pre-B cell receptor signalling. *Nat Rev Immunol* 2014; **14**: 69–80.

506 47 Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C *et al.* Pan-cancer genome and
507 transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* 2018.
508 doi:10.1038/nature25795.

509 48 Zhang J, Mullighan CG, Harvey RC, Wu G, Chen X, Edmonson M *et al.* Key pathways are
510 frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the
511 Children ' s Oncology Group. *Blood* 2011; **118**: 3080–3087.

512 49 Auer F, Ingenhag D, Pinkert S, Kracker S, Hacein-Bey-Abina S, Cavazzana M *et al.* Activation-

induced cytidine deaminase prevents pro-B cell acute lymphoblastic leukemia by functioning as a negative regulator in Rag1 deficient pro-B cells. *Oncotarget* 2017; **8**: 75797–75807.

50 Losdyck E, Hornakova T, Springuel L, Degryse S, Gielen O, Cools J *et al.* Distinct Acute Lymphoblastic Leukemia (ALL)-associated Janus Kinase 3 (JAK3) Mutants Exhibit Different Cytokine-Receptor Requirements and JAK Inhibitor Specificities. *J Biol Chem* 2015; **290**: 29022–29034.

51 Hornakova T, Staerk J, Royer Y, Flex E, Tartaglia M, Constantinescu SN *et al.* Acute lymphoblastic leukemia-associated JAK1 mutants activate the Janus kinase/STAT pathway via interleukin-9 receptor alpha homodimers. *J Biol Chem* 2009; **284**: 6773–6781.

52 Kralovics R, Passamonti F, Buser AS, Teo S-S, Tiedt R, Passweg JR *et al.* A Gain-of-Function Mutation of JAK2 in Myeloproliferative Disorders. *N Engl J Med* 2005; **352**: 1779–1790.

53 Bandaranayake RM, Ungureanu D, Shan Y, Shaw DE, Silvennoinen O, Hubbard SR. Crystal structures of the JAK2 pseudokinase domain and the pathogenic mutant V617F. *Nat Struct Mol Biol* 2012; **19**: 754–759.

54 Jacoby E, Chien CD, Fry TJ. Murine models of acute leukemia: important tools in current pediatric leukemia research. *Front Oncol* 2014; **4**: 95.

55 Kohnken R, Porcu P, Mishra A. Overview of the Use of Murine Models in Leukemia and Lymphoma Research. *Front Oncol* 2017; **7**: 22.

532 Figure legends

533

534 **Fig. 1. Mb1-Cre Δ PB mice develop B cell acute lymphoblastic leukemia (B-ALL).** (A) Mb1-Cre Δ PB
535 mice (Δ PB) mice developed B-ALL characterized by splenic and thymic enlargement (indicated by
536 arrows). (B) Percentage survival of mice of the indicated genotypes: *Mb1*^{+/Cre}*Spi1*^{lox/lox}*Spib*^{-/-} (Mb1-
537 Cre Δ PB, n= 43); *Mb1*^{+/Cre}*Spi1*^{+/+}*Spib*^{-/-} mice (Mb1-Cre Δ B, n=36) and *Mb1*^{+/+}*Spi1*^{lox/lox}*Spib*^{-/-} (Mb1-
538 Cre Δ B, n=14). (C). Comparisons of enlarged spleens and thymuses extracted from Mb1-Cre Δ PB mice
539 compared to control Δ B mice (*Mb1*^{+/+}*Spi1*^{lox/lox}*Spib*^{-/-}). (D) Spleen (left) and thymus (right) weight in
540 grams (g) relative to the body weight in WT and Mb1-Cre Δ PB mice. WT, n=10 (spleen), 5 (thymus);
541 Mb1-Cre Δ PB, n=11 (spleen), 8 (thymus). $p \leq 0.05$ (*), mean $\pm$ SEM. (E). Histologic sections
542 (Hematoxylin & Eosin staining) of spleen and thymus illustrating the lymphocytic infiltration and loss of
543 organs normal structure in Mb1-Cre Δ PB compared to controls Mb1 Δ B (4x).

544

545 **Fig. 2. Most leukemias from Mb1-Cre Δ PB mice resemble pro-B cells and do not express either IgM**
546 **or Igk at the cell surface.** (A) Representative flow cytometric analysis for the presence of CD19⁺ B220⁺
547 B cells in the spleen of C57BL/6 mice; 11 week old Mb1-Cre Δ B mice and Mb1-Cre Δ PB mice (*left*
548 *panels*); and >15 week old Mb1-Cre Δ PB mice (*right panel*). (B) Representative flow cytometric analysis
549 for the presence of CD19⁺ B220⁺ B cells in the thymus of C57BL/6 mice; 11 week old Mb1-Cre Δ B mice
550 and Mb1-Cre Δ PB mice (*left panels*); and >15 week old Mb1-Cre Δ PB mice (*right panel*). (C) PCR for
551 detection of heavy chain rearrangements (J558-JH4) in leukemia B cells prepared from the thymus of
552 Mb1-Cre Δ PB mice (575-411). B cells prepared from WT mouse (C57BL/6) were used as control. The
553 *Cd79* gene was used as control for DNA quality. (D) Representative flow cytometric analysis of leukemic
554 cells from Mb1-Cre Δ PB mice gated on CD19⁺ B220⁺ cells showed that leukemias expressed IL-7R on the

cell surface (top and bottom). **(E)** Representative IL-7R⁺ leukemias expressing IgM and Igκ, (top, Ig⁺) and those not expressing IgM and Igκ (bottom, Ig⁻). **(F)** Percentage of leukemias expressing IgM and Igκ (Ig⁺) or not expressing IgM and Igκ (Ig⁻). Left panel, Mb1-CreΔPB mice, n=14 (spleen) and n=8 (thymus). Right panel, CD19-CreΔPB mice, n=9 (thymus).

Fig. 3 Identification of High Confidence Somatic Nucleotide Variants (SNVs) in Mb1-CreΔPB mouse leukemias. **(A, B and C)** SNVs identified by two different variant caller methods, Strelka and Mutect, were combined and overlapping SNVs were classified as High Confidence SNVs. **(D)** Classification of High Confidence SNVs for impact. Graph shows the number of SNVs classified according to impact. **(E)** Predicted biological effect of the High Confidence SNVs classified as having MODERATE and HIGH impact. **(F)** Venn diagram showing the overlap among High Confidence SNVs identified in the three individual mouse leukemias.

Fig. 4 Integration of whole-exome sequencing (WES) and RNA-seq. **(A, B and C).** Scatter plot correlating the levels of gene expression in Fragments Per Kilobase of Transcript per Million of reads mapped (FPKM log₁₀) and the Variant Allele Frequency (VAF) for genes in which FPKM was greater than zero. Leukemias 853, 854 and 857 are shown, respectively. **(D, E and F).** Biological Pathway Analysis in genes with VAF equal or greater than 20% was performed using Panther – Gene List Analysis. Diagram shows the number of genes enriched according the biological process. Enrichment for genes related to “Cellular Process” and “Metabolic Process” is shown for the three samples analyzed.

Fig. 5 Identified mutations in conserved regions of *Jak3*, *Jak1*, and *Irf3* genes. **(A, B and C; left)** Schematic shows the protein domains of Jak3, Jak1 and Aiolos (*Irf3*). Amino acids substitutions caused

by single nucleotide variants identified in the whole-exome sequencing in samples 853, 854 and 857 are also indicated. **(A, B and C; right)** BLAST analysis comparing Human (*Homo sapiens*), Mouse (*Mus musculus*), Rat (*Rattus norvegicus*), African Elephant (*Loxodonta africana*), Giant Panda (*Ailuropoda melanoleuca*), Atlantic Salmon (*Salmo salar*), Zebrafish (*Danio rerio*), Tropical Frog (*Xenopus tropicalis*) protein sequences for Jak1, Jak3 and Aiolos shows that amino acids which undergo substitution in consequence of a single nucleotide variation are highly conserved between the two species. **(D)** Summary of Sanger sequencing screening for the presence of amino acids substitutions in Jak1 (V657F and V655L), Jak3 (T844M and R653H), and Aiolos (R137* and H195T) in a panel of nineteen leukemias prepared from Mb1-CreΔPB mice. Filled boxes indicate samples in which mutations were identified by Sanger Sequencing.

Fig 6. Jak3 mutations confer growth advantage and increased proliferative potential in normal pro-B cells. **(A and B)** Percentage of GFP⁺ cells over the course of 14 days after infection with MSCV-empty, MSCV-JAK3, MSCV-T844M, MSCV-R653H, MSCV-V670A or MSCV-V670A/T844M in conditions of low (0.5%) and high (5%) IL-7 conditioned medium (CM), respectively. **(C and D)** Absolute number of viable GFP⁺ cells after four days in culture under conditions of low (0.5%) and high (5%) IL-7. Cells were infected with either MSCV-empty, MSCV-JAK3, MSCV-T844M, MSCV-R653H, MSCV-V670A or MSCV-V670A/T844M. Data is presented as number of cells/ml; *n*=3.

Supplementary Fig. 1. Experimental design for RNA-seq and WES-seq experiments. **(A)** DNA and RNA were isolated from leukemic cells extracted from thymus of leukemic mice. **(B)** Weeks of survival and weight of the organs extracted from representative leukemic Mb1-CreΔPB mice. Samples 853, 854, and 857 were used for sequencing experiments. **(C)** Flow cytometric characterization (IgM, x-axis; Igκ,

y-axis; gated on CD19⁺ B220⁺ cells) of leukemic mice in which DNA and RNA were isolated for sequencing. C57BL/6, control; 853, 854, 857, leukemic Mb1-CreΔPB; FMO = fluorescence minus one.

Supplementary Fig. 2. Characterization of the nucleotide somatic nucleotide variants (SNVs) identified by Strelka somatic variant caller in three Mb1-CreΔPB mice leukemias. (A, B and C) SNV distribution according to chromosome number in leukemias 853, 854 and 857, respectively. (D) Allelic change frequency for 853, 854 and 857 leukemias showing an enrichment for G>T and C>A substitutions.

Supplementary Fig. 3. Mutational signature of Mb1-CreΔPB mouse leukemias shows strong bias to C•G -> A•T transversions. (A-C) DeconstructSigs reconstruction analysis of mouse leukemias 853, 854 and 857. COSMIC mutation signatures were used to reconstruct the mutational profile of mouse leukemias and to predict similarity. (D) Heat map illustrating the mutational context of 5' and 3' nucleotides surrounding the transitions identified in samples 853, 854 and 857.

Supplementary Fig. 4. Sanger sequencing confirms the presence of SNVs identified by WES. PCR for amplification of the region containing the SNVs were performed on cDNA synthesized from RNA prepared from mouse tumours. PCR products were submitted for Sanger sequencing using primers targeting the region of interest. (A); Sanger sequencing results of region containing *Ikzf3* SNV identified in the sample 857, transition C-T; (B) *Ikzf3* SNV identified in the sample 853, transition C-T; (C) *Jak1* SNV identified in the sample 854; transition G-T; (D) *Jak3* SNV identified in the sample 853; transition G-A. (E) *Jak3* SNV identified in the sample 857; transition C-T. Chromatograms are shown for each SNV investigated. Note that most SNVs are detectable at 50% or lower allelic frequency.

Figure 1

Figure 2

Figure 3

Figure 4

A

A

m C

JAK3
T844M
JAK3
R653H
JAK1
V657F
JAK1
V655L
AIOLOS
H195T
AIOLOS
R137*

Figure 6

A

B

C

D

Supplementary Fig.1

Supplementary Fig.2

A

B

C

D

Supplementary Fig.3

A

B

C

D

Supplementary Fig.4

A. *Ikzf3* SNV identified by WES in the sample 857

B. *Ikzf3* SNV identified by WES in the sample 853

C. *Jak1* SNV identified by WES in the sample 854

D. *Jak3* SNV identified by WES in the sample 853

E. *Jak3* SNV identified by WES in the sample 857