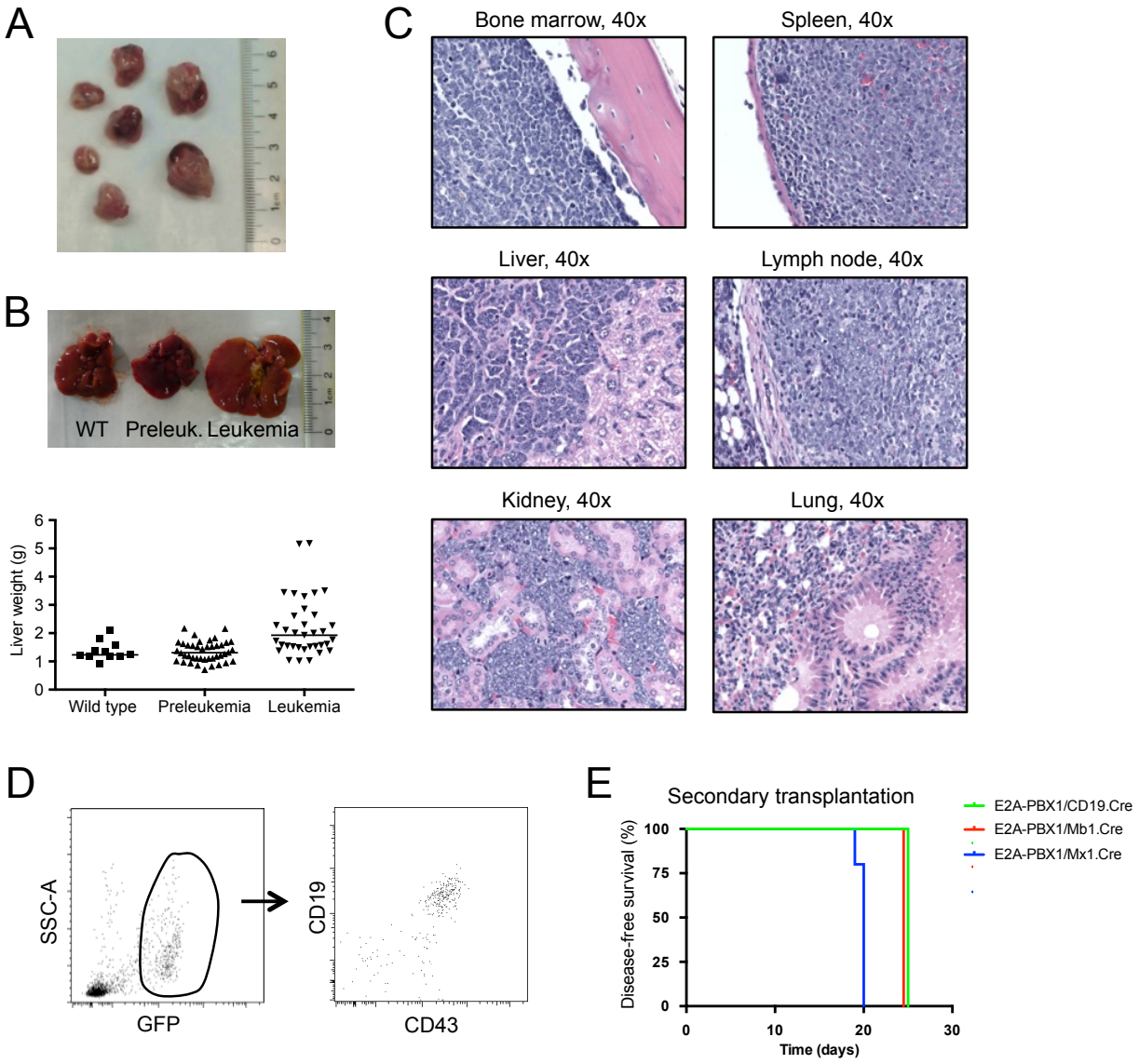
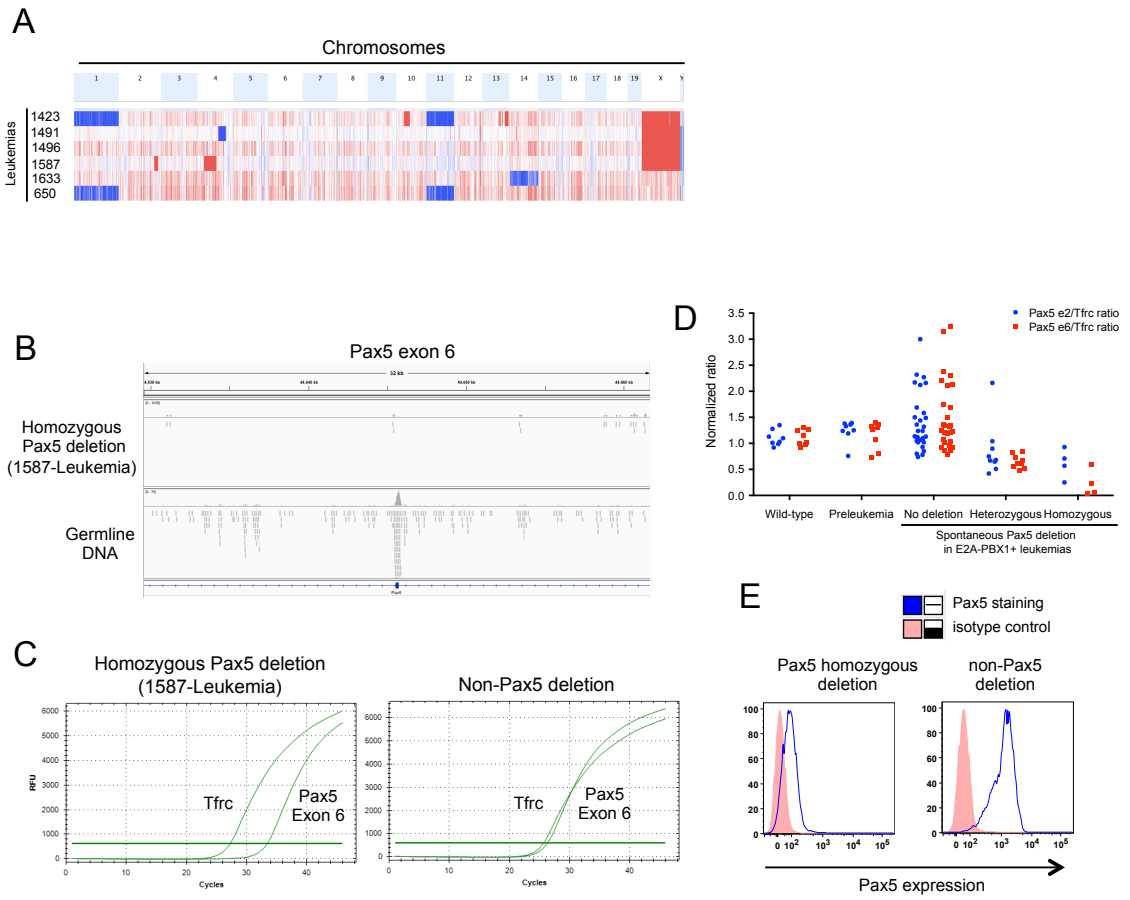
Supplemental Information

Comparative genomics reveals multistep pathogenesis of E2A-PBX1 acute lymphoblastic leukemia

Jesús Duque-Afonso, Jue Feng, Florian Scherer, Chiou-Hong Lin, Stephen H.K. Wong, Zhong Wang, Masayuki Iwasaki and Michael L. Cleary

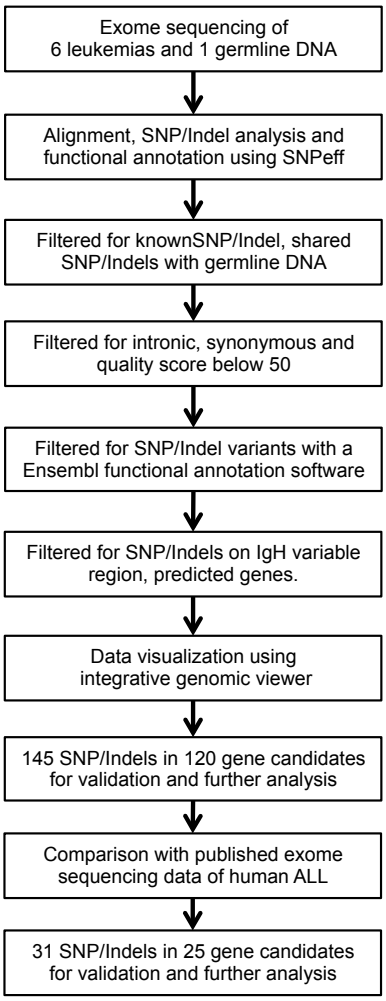


Supplementary Figure 1. Features of leukemias in conditional E2A-PBX1 transgenic mice. **(A)** Lymph nodes of representative E2A-PBX1 leukemic mice. **(B)** Comparison of liver size (upper) and weight (lower) for wild type (n=11), healthy preleukemic (n=42), and leukemic (n=35) mice. Each dot represents a mouse; horizontal bars represent the mean. **(C)** Histologic analysis shows infiltrating leukemia cells in the indicated tissues. **(D)** Flow cytometry analysis shows GFP+ cells expressing CD19 and CD43 surface markers in the central nervous system of a leukemic mouse. **(E)** Kaplan Meier plot displays survival of mice (n=3-5 in each group) after secondary transplantation of E2A-PBX1 leukemia cells.



Supplementary Figure 2. Acquired deletions of *Pax5* in E2A-PBX1 leukemias. **(A)** Copy number variation (CNV) analysis of exome sequences of six leukemias showing chromosomal gains (blue) and losses (red). Of note, 1587-leukemia harbors a partial deletion of chromosome 4, where *Pax5* gene is located. **(B)** Data visualization of exome sequencing reads in *Pax5* exon 6 locus of 1587-leukemia with spontaneous homozygous *Pax5* deletion. **(C)** Genomic DNA-qPCR on *Pax5* exon 6 and *Tfrc* gene as control, confirming *Pax5* deletion detected in exome sequencing. **(D)** Analysis of *Pax5* exon2 and exon 6 deletion compared to *Tfrc* gene (control) by genomic qPCR of total bone marrow of wild type (n=8), sorted lin-CD19+CD43+GFP+ preleukemias (n=8) and leukemias (n=41). Leukemias with *Pax5* deletions were classified as heterozygous (ratio below 0.70) or homozygous (ratio below 0.30) in at least one fragment. Underestimation of *Pax5* deletions is possible due to selective genomic qPCR. **(E)** Pax5 expression by flow cytometry with intracellular staining of a representative sample with homozygous *Pax5* deletion and with non-*Pax5* deletion from 14 analyzed leukemia samples. Red, isotype-stained cells; blue, antibody-stained cells.

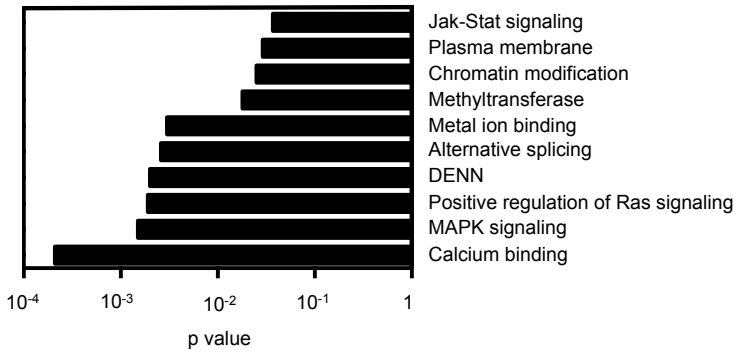
A



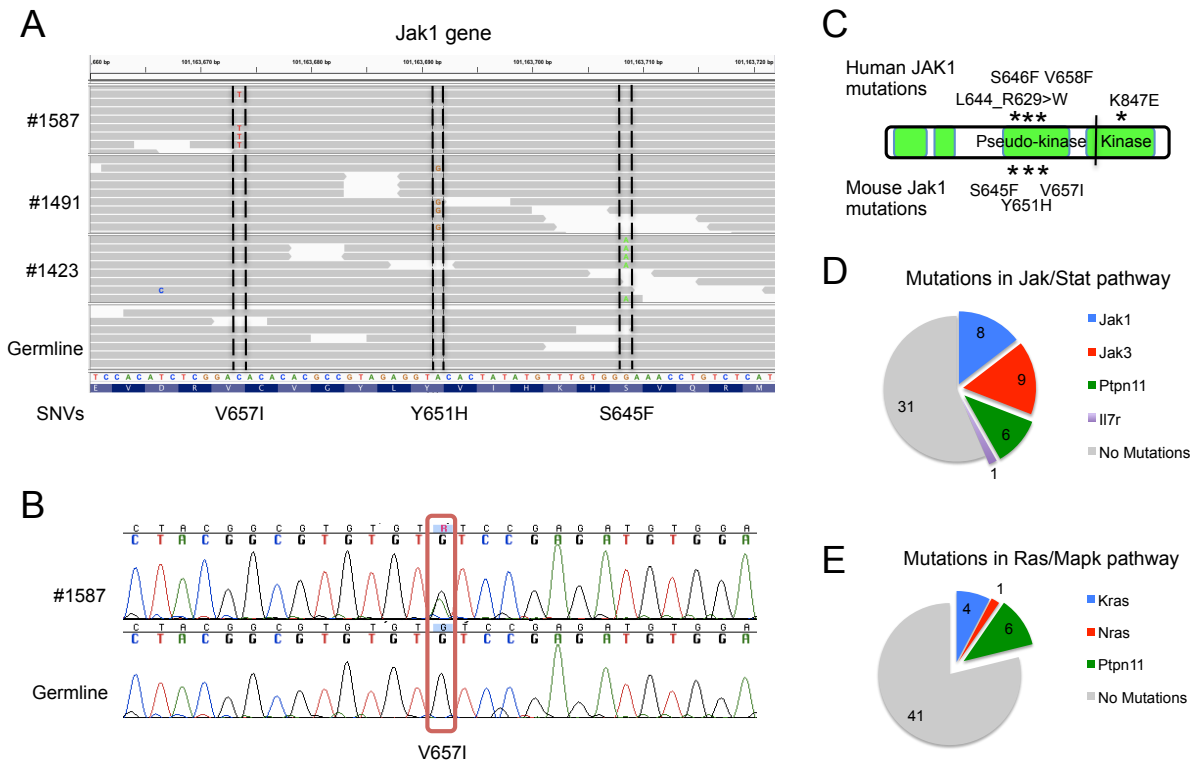
B

ID sample	Total reads (Gpb)	Raw coverage	Total SNV/ Indels	SNV	Missense SNV	Nonsense SNV	Indel
650	10.61	213	25	16	15	1	9
1423	12.41	250	26	21	20	1	5
1491	12.92	260	14	11	11	0	3
1496	12.31	248	35	18	17	1	17
1587	11.76	237	25	19	18	1	6
1633	11.11	204	8	7	6	1	1
Total			133	92	87	5	41
Mean (range)			22.2 (8-35)	15.3 (7-21)	14.5 (6-20)	0.83 (0-1)	6.8 (1-17)
Validated (%)			19/6 (76.0)	19/1 (95.0)	17/1 (94.4)	2/0 (100)	0/5 (0)

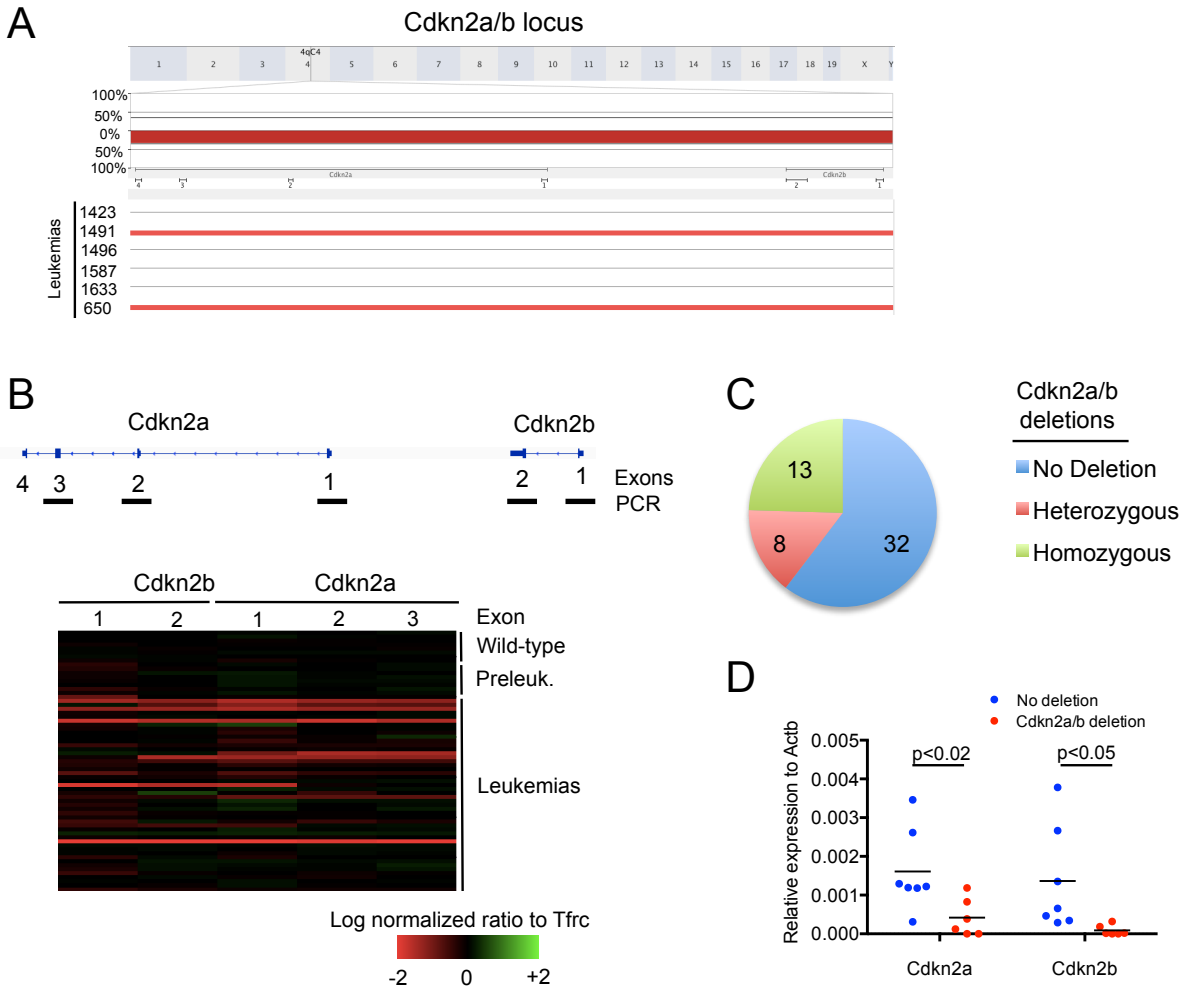
C



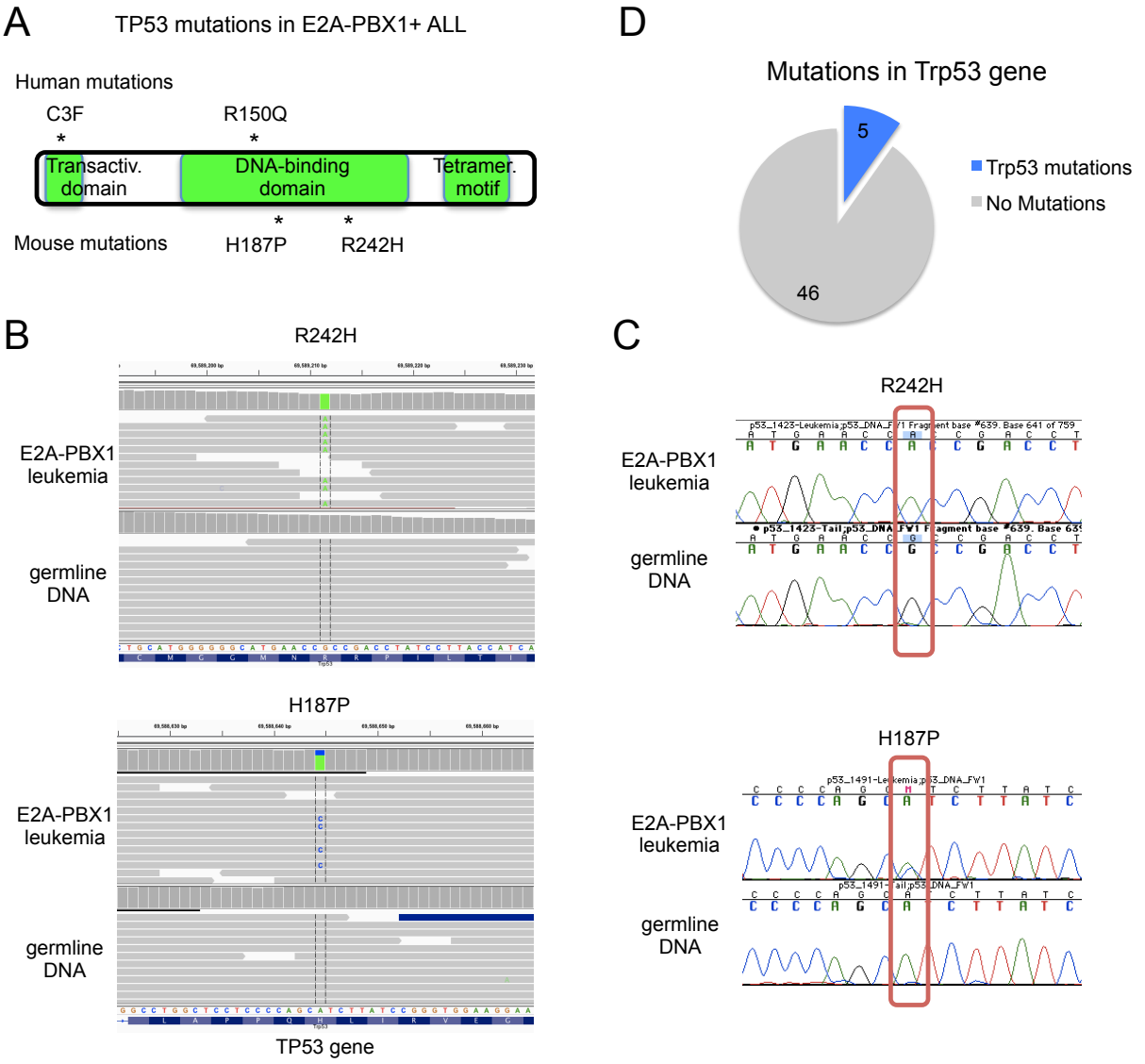
Supplementary Figure 3. Workflow for analysis of SNVs and Indels. (A) Schematic shows workflow for analysis of SNVs/Indels identified by whole exome sequencing of bone marrow leukemia cells (n=6). **(B)** Total number of SNV and Indels identified in exome sequencing as well as validated by Sanger sequencing. **(C)** Functional annotation clustering of enriched pathways and biological processes of genes affected by SNVs in mouse E2A-PBX1 leukemias using the DAVID functional annotation tool (1, 2).



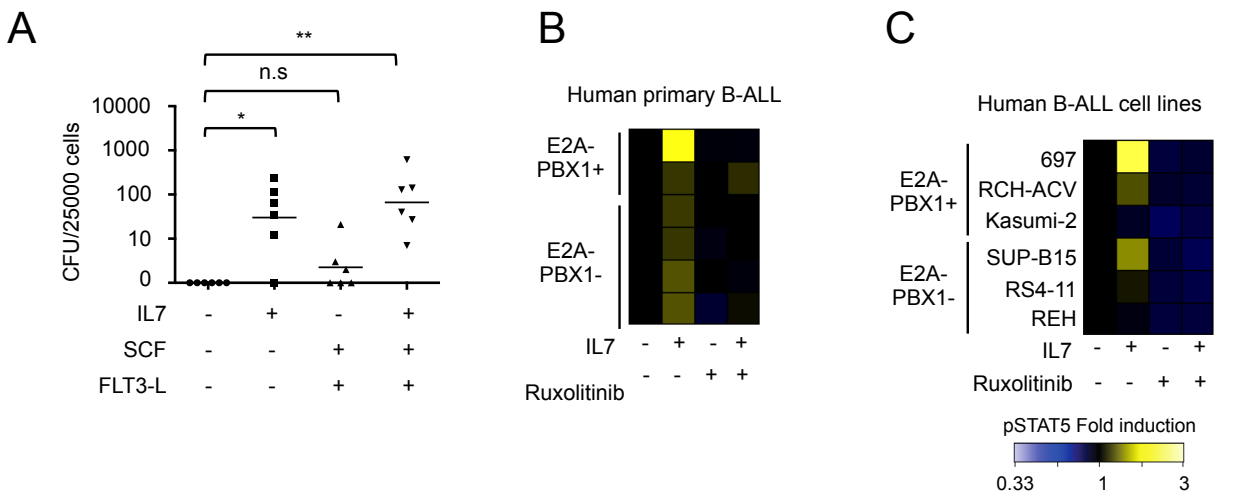
Supplementary Figure 4. Activating mutations in Jak/Stat and Ras/Mapk signaling pathways in mouse E2A-PBX1 leukemias. (A) Visualization of exome sequence data for three E2A-PBX1 leukemias compared to germline DNA at the *Jak1* gene locus. **(B)** Chromatograms from Sanger sequencing show validation of a *Jak1* mutation identified by exome sequencing. **(C)** Schematic representation of the JAK1 protein depicts locations of mutations observed in human ALL (3) and mouse E2A-PBX1 leukemias by exome sequencing. **(D, E)** Pie charts show number of mutations in the Il7r/Jak/Stat **(D)** and Ras/Mapk **(E)** signaling pathways in a cohort (n=51) of mouse E2A-PBX1 leukemias. Note: Concurrent *Jak1* and *Jak3* mutations were present in one leukemia, concurrent *Jak1* and *Ptpn11* mutations in two leukemias, concurrent *Jak3* and *Ptpn11* in one leukemia, although our data do not indicate whether concurrent mutations are in the same clones. Two leukemias carried the *Jak1* V657I mutation, which is comparable to human *JAK1* V658F and equivalent to human *JAK2* V617F (4). Of note, cells from a leukemia that grew in the absence of IL-7 also contained homozygous *Il7r* gene mutation in the transmembrane domain (Supplementary Table 1). Transmembrane domain mutations of the Il7r are known to confer self-homodimerization and cytokine-independent growth in vitro (5) and have been reported in human ALL (6). Concurrent *Kras* and *Ptpn11* mutations were found in one leukemia.



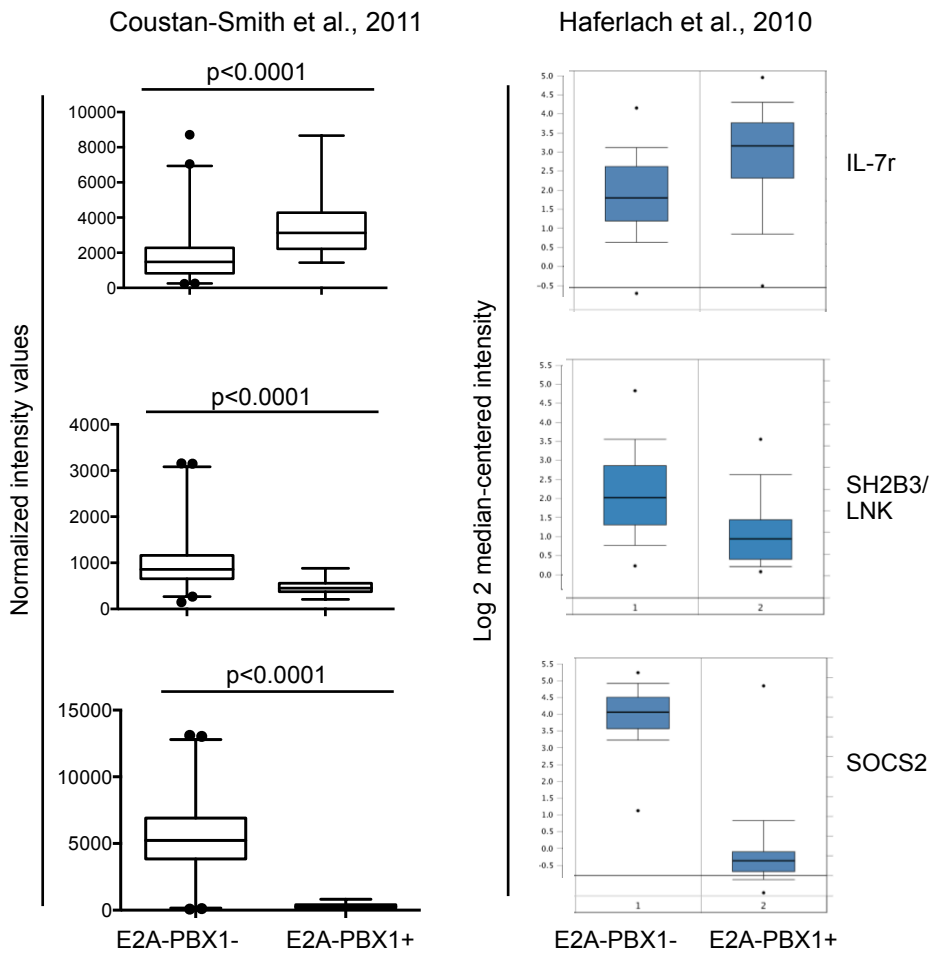
Supplementary Figure 5. *Cdkn2a/b* deletions in E2A-PBX1 mouse leukemias. (A) CNV analysis of *Cdkn2a* and *Cdkn2b* genes in six leukemias after whole exome sequencing. Red denotes loss of copy number in two leukemias. **(B)** Analysis of *Cdkn2a* (exon 1, 2 and 3) and *Cdkn2b* (exon 1 and 2) gene deletions by genomic qPCR of total bone marrow of wild type (n=8), Lin-CD19+CD43+GFP+ sorted preleukemic B cell progenitors (n=8) and E2A-PBX1 leukemia (n=53). **(C)** Pie chart showing frequency of *Cdkn2a/b* locus deletions in bone marrow of E2A-PBX1 leukemias. Copy number variation was defined as heterozygous and homozygous in regions with a normalized ratio to control *Tfrc* gene of <0.7 and <0.3, respectively, in at least 2 adjacent regions. **(D)** *Cdkn2a* and *Cdkn2b* gene expression of lin-CD19+CD43+GFP+ sorted bone marrow leukemia cells were analyzed by RT-qPCR. Expression levels were compared depending on *Cdkn2a/b* locus deletions (n=7, no deletion; n=6, *Cdkn2a/b* deletion). Statistical analysis by Student's t-test was performed between groups.



Supplementary Figure 6. ***Trp53*** **mutations in E2A-PBX1 leukemias.** **(A)** Schematic representation of *TP53* mutations found in human E2A-PBX1 patients (7) and in mouse E2A-PBX1 leukemias by exome sequencing. **(B)** Data visualization of two *Trp53* mutations of mouse E2A-PBX1 leukemias detected by exome sequencing and **(C)** their validation by Sanger sequencing. **(D)** Pie chart shows proportion of leukemias with *Trp53* mutations as assessed by targeted Sanger sequencing in a larger cohort of mice (n= 51).



Supplementary Figure 7. Aberrant activation of signaling pathways in E2A-PBX1 leukemia cells. (A) Colony-forming assay of bone marrow leukemia cells (n=6) cultured with the indicated cytokines. Colonies were enumerated after seven days. Horizontal bar denotes mean. Statistical analysis was performed by Student's t-test. * p-value <0.05, ** p value <0.01, n.s., not significant. **(B)** Human primary ALL cells and **(C)** cell lines were pre-treated with 1μM ruxolitinib or vehicle (DMSO) and then stimulated with 10 ng/ml IL7. Heat map shows pSTAT5 MFI comparing stimulated to unstimulated cells of a representative from three independent experiments. Basal levels of E2A-PBX1+ were higher compared to E2A-PBX1- primary ALLs. MFI fold induction of pSTAT5 by IL-7 in the 697 (p-value<0.02), RCH-ACV (p-value, 0.02) and SUP-B15 (p-value<0.02) cell lines was statistically significant.



Supplementary Figure 8. Transcriptional regulation of JAK/STAT pathway in human E2A-PBX1 patients. Publicly available data sets from two independent cohorts of ALL patients (8, 9 through Oncomine) were interrogated for the expression of key target genes of the IL7/JAK/STAT pathway in E2A-PBX1+ compared to E2A-PBX1- patients. Genes with consistently statistically significant variable expression level are shown. Statistical analyses were performed using Student's t-test.

Supplementary Table 1. Validated SNVs in mouse E2A-PBX1 leukemias.[$]								
Sample ID	**Chrom #**	**Pos**	**Ref**	**Alt**	**VAF (%)**	**Gene name**	**Aa change**	**Human ALL aa change**
1496	X	12048468	T	A	100	Bcor	K1179*	W172I
1423	18	34607411	C	A	51.5	Brd8	V918I	186+1G>T
1491	2	14981326	G	A	47.5	Cacnb2	E311K	G116R
1633	11	119081097	G	A	54.5	Cbx4	P484L	
1633	X	7010060	T	G	35.3	Ccnb3	T126P	S30Y
650	8	44953418	G	T	49.4	Fat1	E1069*	R2354W V3097I
650	8	45040741	G	A	63.5	Fat1	V3952I	
1496	4	152303049	G	C	46.6	Icmt	V234L	
1496	15	9510226	A	C	100	Il7r	I247S	L242>FPGVC T244>VKAVT
1587	4	101163673	C	T	40.4	Jak1	V657I	V658F K847E S646F L644_R629>W
1491	4	101163691	A	G	27.9	Jak1	V651H	
1423	4	101163708	G	A	44.9	Jak1	S645F	
1423	8	71684008	G	A	51.3	Jak3	R653H	S789P
1587	6	145246771	C	T	40.6	Kras	G12D	G12, E63
1496	3	103060270	C	A	42.6	Nras	Q61K	G12D, Q61K
1423	5	14714095	C	T	46.8	Pclo	R1228C	T4194M
1587	5	121143097	G	A	35.1	Ptpn11	S502L	S502L, G503R, T507K
1491	11	69588644	A	C	20.2	Trp53	H187P	R174, C176, R213 Y220, G245, R248
1423	11	69589211	G	A	100	Trp53	R242H	

$ SNVs identified by whole exome sequencing analysis of mouse E2A-PBX1 leukemias, previously identified to be affected mainly in human ALL (3, 10), were validated by Sanger sequencing. DNA was obtained from BM leukemia cells and tail DNA (except leukemia #650, in which Lin+CD19-CD43-GFP- sorted cells were used for germline DNA).
*, stop codon
Chrom #, chromosome number
Pos, position
Ref, reference allele
Alt, alternative allele
Aa, amino acid
VAF, variant allele frequency

Supplementary Table 2. Antibodies for flow cytometry analysis and FACS.			
Antigen	**Fluorochrome**	**Clone**	**Source**
CD19	APC-Cy7	1D3	BD Biosciences
CD43	APC	S7	BD Biosciences
CD43	PE	S7	BD Biosciences
CD3	PE	145-2C11	BD Biosciences
CD4	PE	RM4-5	BD Biosciences
CD8	PE	E53-6.7	eBiosciences
CD11b	PE	M1/70	eBiosciences
Gr1	PE	RB6-8C5	BD Biosciences
NK1.1	PE	PK136	eBiosciences
Ter119	PE	TER-119	eBiosciences
CD127	PECy5	A7R34	eBiosciences
CD45R/B220	PECy7	RA3-6B2	BD Biosciences
CD179a/VPreB	PE	R3	Biolegend
IgM	PerCPCy5	R6-60.2	BD Biosciences
CD24	PE	M1/69	BD Biosciences
CD25	PE	PC61	BD Biosciences
CD117	PE	2B8	BD Biosciences
Bp-1	PE	6C3	eBiosciences
TdT	PE	19-3	eBiosciences (intracellular staining)
CD79a	Alexa647	24C2.5	eBiosciences (intracellular staining)
IgM	APC	II/41	BD Biosciences (intracellular staining)

Supplementary Table 3. Conjugated antibodies for phospho-flow analysis.			
Antigen	**Fluorochrome**	**Clone**	**Source**
pSTAT3 pY705	Alexa 647	4/P-STAT3	BD Biosciences
pSTAT5 pY694	PE	47	BD Biosciences
pERK pT202/pY204	Alexa 647	20A	BD Biosciences
pAKT pS473	Alexa 647	M89-61	BD Biosciences
pPLCγ2 pY759	Alexa 647	K86-689.37	BD Biosciences

Supplementary Table 4. Primers for *Pax5* deletion and genomic qPCR.		
Gene name		**Sequence**
Pax5^{del} in conditional knock-in mice	Forward	AGGAGCTCCTGGTGGCTCAGCTGA
	Reverse	CCACAGCTACTTTGAATAGGGTGAT
Pax5 exon2	Forward	GCTGCGGGCTATCTTTTAGC
	Reverse	GGCCACTCCCAGATGTAGTC
Pax5 exon 6	Forward	GTGCTGTCTCTCAAACACGC
	Reverse	GGAGTCTCCAGTGCCGAATG
Tfrc	Forward	TGTGTCTGGTCTTGGCGTTT
	Reverse	GACAGCAGGTAGCACGGTAG

Supplementary Table 5. Primers for *Cdkn2a/b* genomic qPCR.		
Gene name		**Sequence**
Cdkn2a exon 1	Forward	TGCGGCCCTCTTCTCAAGAT
	Reverse	TTCTTGGTGAAGTTCGTGCG
Cdkn2a exon 2	Forward	CACACTCTGCTCCTGACCTG
	Reverse	TTTAGCGCTGTTTCAACGCC
Cdkn2a exon 3	Forward	GGTCTTCTTACAGCTGTGCATTTT
	Reverse	TGGGGCCTCCCTGTATCTGTAA
Cdkn2b exon 1	Forward	CTTGCCCAGCTTGTACGGAA
	Reverse	CGAAACATCCCTCGGCACT
Cdkn2b exon 2	Forward	AATCCAGGCATCAAGGCAACT
	Reverse	CCTCCAGCTCGACAAGGAAATA

Supplementary Table 6. Primers for Sanger sequencing.		
Gene name		**Sequence**
Bcor	Forward	CAAGCGAAGACGGATCTCCA
	Reverse	GTTCGTTGGCAAGGTTGTGG
Brd8	Forward	GGGGATGCCTGAAGACCAAA
	Reverse	ACCCTGAAGTGCTGGATGTG
Cacnb2	Forward	AGGCACCCTCATGTCTCTCT
	Reverse	CCCGCTTACCTTGGGAGAAG
Cbx4	Forward	AGTGACGGTAAGGCAGTTCG
	Reverse	AGCGTGAGCCCTGTAAGAAG
Ccnb3	Forward	GAGGCATGGTGCTGATAGACA
	Reverse	CAGTCTCCACCAGCAAGGTT
Fat1 1.SNP	Forward	TGTGGAAGTGGAGGTCGTTG
	Reverse	CACGGAAACATCACACCACG
Fat1 2.SNP	Forward	TTGTCTCCGTCCAGAGCATT
	Reverse	ATGTGCAGTTAGAGACGCGG
Icmt	Forward	CTAGAAGGAGTCGTGTGGGG
	Reverse	GCGGACGAAGGTTAAAGCTG
IL7r	Forward	TGGGGCTCCACAAATAACCC
	Reverse	CTTTGTCAGGGGTCATGGCT
Jak1	Forward	GTGGGAACATTGGGAGGGAG
	Reverse	AGCTCAGTCAGGTTCACACG
Jak3	Forward	CGGGATGTGGGGCTTTAACT
	Reverse	ATTCTGTCGGTGAGCACTGG
Kras	Forward	CGGATGGCATCTTGGACCTT
	Reverse	CCTTTGAGAGCCATTAGCTGC
Nras	Forward	ACCACCACCTCCTCACTCTT
	Reverse	AGATGGTCCATGGGCAAAGG
Pclo	Forward	CCTAGCGCCTTTGCAGACTT
	Reverse	GGCCAACAGGACTTAGTGGA
Ptpn11	Forward	CTGTGTCACATGGTCCCTCC
	Reverse	TGCTGACCACAGCTTACAGG
Trp53	Forward	ATGGTAAGCCCTCAACACCG
	Reverse	CGGGACTCGTGGAACAGAAA

Supplementary Table 7. Primers for VDJ rearrangement detection.		
Primer name		**Sequence**
DH	Forward	TTCAAAGCACAATGCCTGGCT
VH7183	Forward	CGGTACCAAGAASAMCCTGTWCCTGC
VHQ52	Forward	CGGTACCAGACTGARCATCASCAAGG
VH γ 3.8	Forward	CAAGGGACGGTTTGCCTTCTCTTTGGA
VH3609	Forward	KCYYTGAAGAGCCRRCTCACAATCTCC
VHJ558	Forward	CGAGCTCTCCARCACAGCCTWCATGC
Cμ5'	Forward	TGGCCATGGGCTGCCTAGCCCGGGAC
JH3	Reverse	GTCTAGATTCTCACAAGAGTCCGATAG
Cμ3'	Reverse	GCCTGACTGAGTTCACACACAAGGAG

References for supplementary information

1. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc. 2009;4(1):44-57.

2. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37(1):1-13.

3 Holmfeldt L, Wei L, Diaz-Flores E, Walsh M, Zhang J, Ding L et al. The genomic landscape of hypodiploid acute lymphoblastic leukemia. Nat Genet 2013;45:242-52.

4 Toms AV, Deshpande A, McNally R, Jeong Y, Rogers JM, Kim CU et al. Structure of a pseudokinase-domain switch that controls oncogenic activation of Jak kinases. Nat Struct Mol Biol 2013;20:1221-3.

5. Shochat C, Tal N, Bandapalli OR, Palmi C, Ganmore I, te Kronnie G et al. Gain-of-function mutations in interleukin-7 receptor-α (IL7R) in childhood acute lymphoblastic leukemias. J Exp Med 2011;208:901-8.

6. Roberts KG, Morin RD, Zhang J, Hirst M, Zhao Y, Su X et al. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. Cancer Cell 2012;22:153-66.

7. Zhang J, Mullighan CG, Harvey RC, Wu G, Chen X, Edmonson M et al. Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. Blood 2011;118:3080-7.

8. Coustan-Smith E, Song G, Clark C, Key L, Liu P, Mehrpooya M et al. New markers for minimal residual disease detection in acute lymphoblastic leukemia. Blood 2011;117:6267-76.

9. Haferlach T, Kohlmann A, Wieczorek L, Basso G, Kronnie GT, Béné MC et al. Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the International Microarray Innovations in Leukemia Study Group. J Clin Oncol 2010;28:2529-37.

10. Mar BG, Bullinger LB, McLean KM, Grauman PV, Harris MH, Stevenson K et al. Mutations in epigenetic regulators including SETD2 are gained during relapse in paediatric acute lymphoblastic leukaemia. Nat Commun. 2014;5:3469.