Supplementary data.

Clonal conversion of B-lymphoid leukemia reveals cross-lineage transfer of malignant states.

Rajesh Somasundaram[1], Josefine Åhsberg[1], Kazuki Okyama[1], Jonas Ungerbäck[1], Henrik Lilljebjörn[2], Thoas Fioretos[2], Tobias Strid[1] and Mikael Sigvardsson[1,3*].

[1]Department of Clinical and Experimental Medicine, Experimental Hematopoiesis Unit, Faculty of Health Sciences, Linköping University, Sweden.

[2]Department of Laboratory Medicine, Division of Clinical Genetics, Lund University, Sweden

[3]Division of Molecular Hematology Lund University Sweden

*Corresponding author: Mikael Sigvardsson, Faculty of Health Sciences, Linköping, Sweden

University Lab1, Level 13, Linköping. E-mail: mikael.sigvardsson@liu.se

Tel: 46-13-227887, Fax: 46-13-224314.

Running title: B-lineage plasticity in leukemic disease

Supplemental tables

Table S1-4

RNA-seq of $CD19^+$ LN cells (pMIG), converted $CD11b^+$ (C/EBPα–pMIG), and $Thy1.2^+$ cells (ICN1-pMIG). Genes ≥ 2-fold up- or down-regulated with a statistic significance (adj. $p < 0.05$) in $CD11b^+$ (C/EBPα -pMIG) (**Table S1** [up] and **S2** [down]) or $Thy1.2^+$ cells (ICN1-pMIG) (**Table S3** [up] and **S4** [down]) compared to the control are shown. RefSeq IDs, peak positions, gene annotations, peak values (RPKM) of individual samples, fold-change (vs pMIG $CD19^+$ cells), and statistic values are as indicated.

Table S5-8

Genes ≥ 2-fold up- or down-regulated with a statistic significance (adj. $p < 0.05$) in $CD11b^+$ (C/EBPα -pMIG) (**Table S5** [up] and **S6** [down]) or $Thy1.2^+$ cells (pMIG-ICN) (**Table S7** [up] and **S8** [down]) compared to control cells, were analyzed by GO enrichment analysis with ontology of biological process (complete) (from **Table S1-S4**). Significantly enriched GO terms are shown. Counts of annotated genes from reference (*Mus musculus* [n = 22322]) and RNA-seq data, expected counts, fold enrichment, and statistic values (Bonferroni) are indicated. *: $p < 0.05$, **: $p < 0.005$, ***: $p < 0.0005$

Population	Tumour	pMIG(%)	ICN1-pMIG(%)	P Values
$CD19^{+}Thy1.2^{-}$	LN-$Ebf1^{+/-}$294	64.64 ± 1.70 (n=3)	31.99 ± 1.28 (n=2)	na
	LN-$Ebf1^{+/-}$449	1.48 ± 0.79 (n=5)	14.35 ± 7.42 (n=6)	ns
	LN-$Pax5^{+/-}$233	23.71 ± 2.34 (n=6)	12.33 ± 2.42 (n=6)	0.0070**
	LN-$Pax5^{+/-}$369	51.90 ± 0.30 (n=3)	9.46 ± 0.17 (n=3)	< 0.0001****
$CD19^{+}Thy1.2^{+}$	LN-$Ebf1^{+/-}$294	20.39 ± 0.39 (n=3)	36.11 ± 0.75 (n=2)	na
	LN-$Ebf1^{+/-}$449	97.88 ± 0.91 (n=5)	83.27 ± 7.5 (n=6)	ns
	LN-$Pax5^{+/-}$233	73.60 ± 2.65 (n=6)	86.75 ± 2.52 (n=6)	0.0049**
	LN-$Pax5^{+/-}$369	30.74 ± 0.24 (n=3)	3.14 ± 0.27 (n=3)	< 0.0001****
$CD19^{-}Thy1.2^{+}$	LN-$Ebf1^{+/-}$294	≤0.05 (n=3)	32.28 ± 0.58, n=2	na
	LN-$Ebf1^{+/-}$449	0.23 ± 0.07 (n=5)	0.68 ± 0.17 (n=6)	ns
	LN-$Pax5^{+/-}$233	1.57 ± 0.18 (n=6)	1.00 ± 0.21 (n=6)	ns
	LN-$Pax5^{+/-}$369	6.14 ± 0.05 (n=3)	76.00 ± 1.13 (n=3)	< 0.0001****
$CD3^{+}Thy1.2^{+}$	LN-$Ebf1^{+/-}$294	≤0.05 (n=3)	≤0.05 (n=2)	na
	LN-$Ebf1^{+/-}$449	≤0.05 (n=5)	≤0.05 (n=6)	na
	LN-$Pax5^{+/-}$233	≤0.05 (n=6)	≤0.05 (n=6)	na
	LN-$Pax5^{+/-}$369	≤0.05 (n=3)	≤0.05 (n=3)	na

Supplementary table 9.

Constitutively active Notch signals in *Pax5*$^{+/-}$ or *Ebf1*$^{+/-}$ pro-B leukemia cells cause lineage conversion from pro-B to early-T cell stage in vitro.

Pro-B cells from 2 *Ebf1*$^{+/-}$ and 2 *Pax5*$^{+/-}$ tumors were *in vitro* differentiated for 14 days on OP9 stroma cells, either with (ICN1-pMIG) or without (pMIG) a constitutive expressed Notch1 signal. The table displays the percentage of (% of live) $CD19^{+}Thy1.2^{-}$, $CD19^{+}Thy1.2^{+}$, $CD19^{-}Thy1.2^{+}$ and $CD3^{+}Thy1.2^{+}$ cells generated. All percentages were calculated from the total number of $PI^{-}GFP^{+}$ cells. Statistical analysis was performed using unpaired student's t test. **, P < 0.01; ****, P < 0.000, Non-significant (ns) and for sample for which statistical analysis cannot be performed is indicated as Not-applicable (na).

Table S10-13

ATAC-seq of $CD19^+$ LN cells (pMIG), converted $CD11b^+$ (C/EBPα -pMIG), and cells cultured on OP9 ($CD19^+$) or OP9-DL1 ($Thy1.2^+$). ATAC-peaks ≥ 2-fold increased or decreased with a statistic significance (adj. $p < 0.05$) in $CD11b^+$ (C/EBPα -pMIG) (**Table S10** [up] and **S11** [down]) or $Thy1.2^+$ cells (culture on OP9-DL1) (**Table S12** [up] and **S13** [down]) compared to the control are shown. Peak IDs, positions, scores, regional annotations, distances from transcription start sites, promoter IDs, gene and transcript IDs (Entrez, Unigene, RefSeq, Ensembl and Gene Symbl), peak values, and fold-change values (vs control) are as indicated.

Supplementary figure legends.

Supplementary figure 1: Characterization of a representative lymph node sample from Pax5$^{+/-}$Ebf1$^{+/-}$ mouse #484.

Representation of the gating strategy used to characterise and sort the Pax5$^{+/-}$Ebf1$^{+/-}$ (*TH*), *Pax5*$^{+/-}$ or *Ebf1*$^{+/-}$ tumor samples.

Supplementary figure 2: Tumor cells from *Ebf1*$^{+/-}$ or *Pax5*$^{+/-}$ are sensitive to extra cellular Notch signaling and can switch from pro-B cell leukemia cells to T-lineage cells *in vitro*.

Q-PCR analysis was performed on lymph node cells from *Ebf1*$^{+/-}$ or *Pax5*$^{+/-}$ mice cultured either on OP9 or OP9-DL1. RNA was extracted from CD19^{+}Thy1.2^{-} and CD19^{+}Thy1.2^{+} cells cultured on OP9 and CD19^{-}Thy1.2^{+} cells cultured on OP9-DL1 after 14 days culture. B cell genes such as *Igl1l, Cd79a* and *Cd19* were either down-regulated or non-detectable (ND), whereas T-cell genes such as *Lck and Gata3* were significantly up-regulated by cells cultured on OP9-DL1. Each dot indicates a biological replicate analyzed by triplicate in Q-PCR reactions and the colour indicate independently generated tumor. Statistical analysis was performed using unpaired student's *t* test. Astricks in the same colour as the tumor depicted in the figure indicates significance; *, $P < 0.05$; ***, $P < 0.001$; and Non-detectable (ND).

Supplementary figure 3: Ectopic expression of PAX5 is incompatible with phenotypic swtich to T-lineage cells. Representative FACS plots from lineage conversion experiments on OP9 or OP9-DL1 cells with control virus (pMIG) or PAX5 expressing (Pax5-pMIG) transduced tumor cells from two different *Ebf1*$^{+/-}$

Pax5$^{+/-}$ tumors (683, 1103). The lower panel summarize the data. Each dot represents one well of cells analyzed.

Supplementary figure 4: A majority of *Pax5*$^{+/-}$*Ebf1*$^{+/-}$ lymph node cells switched from a CD19^{+}Thy1.2^{-} to CD19^{-}Thy1.2^{+} cells are KIT positive. Dot plots representing the percentage of KIT^{+} cells calculated from the total number of CD19^{-} Thy1.2^{+} events. Each dot represents one well from cell culture and the colour represents the tumors.

Supplementary figure 5: Conversion of B-lineage ALL cells into T-lineage cells is an efficient process. Diagrams displaying the relative cloning frequency of three different primary tumors generated from individual mice. The cloning frequency on OP9 and OP9-DL1 was determined by limiting dilution experiments and analysis of the data using the L-calc software. FACS analysis was used to determine the cellular composition od the generated clones. The fraction of clones containing CD19^{-}cells are indicated in red.

Supplementary figure 6: C/EBPβ expression on B-ALL cells from *Ebf1*$^{+/-}$ or *Pax5*$^{+/-}$ tumors results in lineage conversion from pro-B to monocyte like, CD11b expressing cells.

Ebf1$^{+/-}$ or *Pax5*$^{+/-}$ pro-B cells were transduced with C/EBPβ-pMIG or pMIG followed by 5 days of *in vitro* differentiation on OP9 stroma cell. The cells were analyzed by flow cytometry and percentage of CD19^{+}CD11b^{-}, CD19^{+}CD11b^{+} and CD19^{-}CD11b^{+} (% of PI^{-}GFP^{+}) were plotted. Each dot represents one well and the each tumour is indicated by individual colour. Statistical analysis was performed

using unpaired t-test. Asterisks in the same color as the tumour depicted in the figure indicates significance; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

Supplementary figure 7: C/EBPβ expression on B-ALL cells results in lineage conversion into LPS responsive monocytic cells. Q-PCR analysis of TH-LN cells transduced either with C/EBPβ-pMIG or pMIG, *in vitro* differentiated by co-culture with OP9 for 5 days, followed by with or without LPS (100ng/ml; 18h) stimulation and the level of of *IL6, IL1β, TNFα and IL12p40* expression were quantified. Each dot indicates a biological replicate analyzed in triplicate Q-PCR reactions. Significant differences are shown with asterisks. Statistical analysis was performed using unpaired student's t test. ****, $P < 0.0001$and Non-detectable (ND).

Supplementary figure 8: Normal or transformed *Ebf1*$^{+/-}$*Pax5*$^{+/-}$ pro-B cells display minimal changes in chromatin accesibility before lineage conversion.
Box-plots displaying the relative chromatin accesibility at EBF1, PAX5, GATA3 or CEBPα binding sites as assayed by ATAC-seq analysis of in vitro expanded Wt or *Ebf1*$^{+/-}$*Pax5*$^{+/-}$ pro-B cells and primary Wt in vitro differentiated CD19^{-} Thy1.2^{+} T- or CD19^{-}CD11b^{+} myeloid- cells. The upper panel display the same analysis comparing non transformed *Ebf1*$^{+/-}$*Pax5*$^{+/-}$ pro-B cells with two independently generated tumor populations as well as the Wt, T and CD11b^{+} cells. ATAC tags were annotated on peaks from ChIP-seq data retrieved from GEO mapped to mm10 and reanalyzed for peakfinding using HOMER; EBF1 (GSE69227), PAX5 (GSE38046), GATA3 (GSE31235), TCF7 (TCF1) (GSE46662) and C/EBPα (GSM537983). Boxplots of Log2 tags (means from duplicates) on these peaks were quantified. Statistics was tested using one-way-Anova with Sidak´s multiple comparison test.

Significantly up and down regulation were shown with red and blue astricks respectively; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

Supplementary figure 9: EBF1 or PAX5 are frequently deleted in T-ALL.

(a) A heat map displaying copy number status for *EBF1* and *PAX5* together with *NOTCH1* mutation status for 17 human T-ALL cell lines. **(b)** A heat map indicating copy number status for *EBF1* and *PAX5* together with *NOTCH1* mutation status for 67 human T-ALLs. Blue indicates loss of material and red indicates gain of material. Mutation status for *NOTCH1* is indicated on the right side of each heatmap. The exact location of *EBF1* and *PAX5* is indicated below the heatmaps.

Supplementary experimental procedures

Animal models.

Pax5$^{+/-}$ (Urbánek et al. 1994) or *Ebf1*$^{+/-}$ (Lin and Grosschedl 1995) were all on C57BL/6 (CD45.2) background whereas the control animals used in transplantation were on C57BL/6 (CD45.1). *Pax5*$^{+/-}$ and *Ebf1*$^{+/-}$ mouse were crossed to generate transheterozygote for *Pax5*$^{+/-}$*Ebf1*$^{+/-}$. Animal procedures were performed with consent from the local ethics committee at Linköping University (Linköping, Sweden).

Cell culture

Normal lineage negative Sca1^{+}Kit^{+} (LSK) cells or lymph node pro-B cells (Lin^{-} (CD11b, Gr1, Ter119 and CD11c) NK1.1^{-}IgM^{-}CD3^{-}CD19^{+}CD43^{+} were sorted and expanded *in vitro* by co-culture on OP9 stroma cells using OptiMEM supplemented with 10% heat inactivated fetal calf serum, 25mM HEPES, 50μg/ml Gentamicin, 50μM β-mercaptoethanol, 10ng/ml KIT ligand, 10ng/ml Fms-like tyrosine kinase 3 ligand (FLT3L), and 10ng/ml Interleukin-7. For constitutively active Notch1 expression experiments, ICN1-pMIG or pMIG (control) retrovirus (MSCV) (Pui et al. 1999) transduced cells were differentiated on OP9 for 14 days. The *in vitro*-differentiated cells were analysed either by flow-cytometry, live sorted for Q-PCR or used for transplantation. For extra-cellular Notch experiment control cells were co-cultured on OP9 or OP9-DL1 stroma cells for 14 days. Both for extracellular Notch and the constitutive Notch1 signal experiments, the above-mentioned media and the cytokines were used. During *in vitro* differentiation experiments, the cells were re-supplemented with 10ng/ml KIT ligand, 10ng/ml Fms-like tyrosine kinase 3 ligand (FLT3L), and 10ng/ml Interleukin-7 after 7 days. For pro-B to myeloid

differentiation experiments, C/EBPβ-pMIG or pMIG (control) retrovirus transduced pro-B cells were co-cultured on OP9 in OptiMEM supplemented with 10% heat inactivated fetal calf serum, 25mM HEPES, 50μg/ml Gentamicin, 50μM β-mercaptoethanol, 10ng/mL Fms-like tyrosine kinase 3 ligand (FLT3L), 10ng/ml Interleukin-7, 10ng/ml IL3, 10ng/ml SCF and 10ng/mL M-CSF. At day five of co-culture, cells were analysed by flow cytometry or sorted for Q-PCR or stimulated with LPS (100ng/ml; 18h) or transplanted. During C/EBPα transduction experiment C/EBPα-ER-pMIG or pMIG (Control) retrovirus were transduced into pro-B cells and cultured either with or without 4-HydroxyTamoxifen (1μM; Sigma; H7904) supplemented with 10% heat inactivated Charcoal stripped fetal calf serum, 25mM HEPES, 50μg/ml Gentamicin, 50μM β-mercaptoethanol, 10ng/mL Fms-like tyrosine kinase 3 ligand (FLT3L), 10ng/ml Interleukin-7, 10ng/ml IL3, 10ng/ml SCF 10ng/mL M-CSF and analysed by flow cytometry. For transplantation experiments C/EBPα-ER-pMIG or pMIG transduced cell were expanded for one week, followed by in vitro differentiation in the presence of 4-OHT, $CD11b^{+}CD19^{-}$ (from C/EBPα-ER-pMIG) and $CD11b^{-}CD19^{+}$(from pMIG) cells were transplanted. All retroviral transductions were performed using spin-infection method as described (Zandi et al. 2012). All mouse cytokines were bought from Peprotech (Peprotech, Rocky Hill, NJ).

FACS staining.

Frozen lymph node cells were CD16/CD32 (Fc)–blocked, stained and sorted based on the following markers (Lin^{-} (CD11b,Gr1,Ter119 and CD11c) $NK1.1^{-}CD3^{-}$ $CD19^{+}CD43^{+}$ and expanded *in vitro* by co-culture on OP9 stroma cells. For *in vitro* differentiation (pro-B to T cells) experiments, CD16/CD32 (Fc)–blocked cells were stained with antibodies against CD3, CD19 and Thy1.2. For *in vitro* (pro-B to

Myleoid) differentiation analysis CD16/CD32 (Fc)–blocked cells were stained with antibodies against CD11b/Mac1 and CD19. For the T-cell developmental-block stage analysis, CD16/CD32 (Fc)–blocked cells were stained with antibodies against CD19, Thy1.2, CD3, Kit, CD25 and CD44. Analysis and cell sorting was performed on a BD FACSAria^{TM} (BD Biosciences, San Jose, California) using propidium iodide (PI, In vitrogen, Paisly UK) as viability marker. From ICN1-pMIG or pMIG transplantation, recipient mouse organs were CD16/CD32 (Fc)–blocked prior to staining, followed by staining with Thy1.2, CD19, CD45.2, and CD3. For C/EBP transplantation, recipient mouse organs were analysed by CD16/CD32 (Fc)–blocked, followed by staining with CD19, CD45.2, CD45.1, Thy1.2 and CD11b. Analysis and cell sorting was performed on a BD FACSAria^{TM} (BD Biosciences, San Jose, California) using propidium iodide (PI, Invitrogen, Paisly UK) as viability marker. The following antibodies and clones were used CD16/CD32 (Fc-block) – (93, eBioscience), CD11b/Mac1 (M1/70, Biolegend), Gr1 (RB6-8C5, eBioscience), Ter119 (Ter119, Biolegend), CD11c (N418, Biolegend), NK1.1 (PK136, Biolegend), CD3 (17A2, Biolegend), CD19 (ID3, eBioscience), Thy1.2 (53-2.1, eBioscience) CD43 (S7, BD Biosciences), kit (2B8, eBioscience), CD25 (PC61, eBioscience), CD44 (IM7, eBioscience), CD45.1(A20, Biolegend), and CD45.2 (A20, Biolegend).

RNA-sequencing and data analysis.

Total RNA was isolated from control CD19^{+} LN cells (pMIG), converted CD11b^{+} (C/EBPα-ER-pMIG), Thy1.2^{+} cells (ICN1-pMIG) and cells cultured on OP9 or OP9-DL1 by use of RNAeasy Micro Kit (Qiagen, Hilden, Germany) according to manufacturer's recommendations. Libraries were constructed using NuGEN's Ovation Ultralow Library systems (NuGEN Technologies, San Calros, CA) and were

subsequently subjected to 76 cycles of NextSeq500 sequencing (Illumina, San Diego, CA). Each biological sample was processed and sequenced in duplicate or triplicate. For analysis of RNA-Seq experiments the reads were aligned to mouse reference genome (mm10 / GRCm38) using TopHat (Kim et al. 2013). If not indicated, further analyses were performed using the HOMER platform (Heinz et al. 2010) using normalization to 10M mapped reads by the *analyzeRepeats.pl* command with the option *–count exons condeseGenes*. For analysis of statistically significance among differently expressed genes, the data was analyzed using *analyzeRepeats.pl* with the *–noadj* option, and filtered by ≥ 50 tags repeats followed by the *getDiffExpression.pl* command using edgeR (Robinson et al. 2010). Data normalized to 10M mapped reads per experiment in log scale was visualized by Java Treeview (Saldanha 2004) following hierarchical clustering in Cluster3 (centering genes on mean, average linkage cluster) (de Hoon et al. 2004). Gene lists (2-fold up- or down-regulated in converted cells, with a statistic significance [$p < 0.05$]) from RNA-seq experiment were uploaded to PANTHER Overrepresentation Test (release 2016/7/15) (http://geneontology.org/), and enrichment analyses were run with Gene ontology database released on 2016/9/24. *Mus musculus* genes from the database (22322 genes) were employed as reference.

Quantitative RT-PCR.

Q-PCR analysis of sorted cells was performed as previously described (Mansson et al. 2008). Assays-on-Demand™ probes (Applied Biosystems, Foster City, CA) used were: *Hprt*; Mm00446968_m1, *Cd79α/Mb-1;*Mm00432423_m1, *Cd19* Mm00515420_m1, *Lck* Mm00802897_m1, *Cd3* Mm00599683_m1, *Gata3* Mm00484683_m1, Igll1 (Mansson et al. 2010), *IL-6;*Mm00446190_m1, *IL1-*

*β;*Mm00434228_m1, *TNFα;* Mm00443258_m1, *Mpo*; Mm01298424_m1 and *IL12p40;*Mm00434174_m1

Transplantation procedures

Adoptive transfers were performed by tail vein injection. Non-irradiated CD45.1 animals were transplanted with approximately 250.000 lymph node cells transduced either with retrovirus for ICN1-pMIG or pMIG for pro-B to T- cell experiment. For pro-B to Myleoid experiments, 100.000 of $CD19^{-}CD11b^{+}$ from C/EBPβ-pMIG/ C/EBPα-ER-pMIG or $CD19^{+}CD11b^{-}$ from pMIG transduced cells were transplanted into non-irradiated CD45.1 hosts.

Transduction of lympnode tumor cells with intracellular Notch1 or C/EBP.

For the generation of lymph node pro-B cells with constitutively active Notch1, tumor cells derived from $Pax5^{+/-}Ebf1^{+/-}$ animals were transduced with retroviruses (MSCV) encoding ICN1-pMIG or pMIG as a control (Pui et al. 1999). The transduced cells were sorted for GFP expression and expanded on OP9 stroma cells prior to flow cytometric, Q-PCR analysis or transplantation. For the generation of lymphnodes cells with consitutively overexpressing C/EBPβ proteins, tumor cells were transduced with retrovirus encoding C/EBPβ-pMIG or pMIG as control. And for C/EBPα transduction experiment C/EBPα-ER-pMIG or pMIG (Control) retrovirus were transduced into tumorous pro-B cells and cultured either with or without Tamoxifen (1μM; Sigma; T5648).

Ig and TCR recombination analysis.

Cells were sorted and DNA was extracted using Qiagens-Allprep DNA/RNA microkit according to manufacturers instruction. The TCRβ-VDJ and DJ assays were adopted from (Levin et al. 1993; Hamrouni et al. 2003) and the recombination events were quantified by real-time quantitative PCR using FastStart Universal SYBR green Master (ROX) (Roche). For TCR-VDJ recombination assay, we used Vβ12–5' agttacccagacacccagacatga; Jβ2-3' tgagagctgtctcctactatcgatt; and for TCR- DJ recombination assay, Dβ1-5' cagccccttcagcaaagat and Jβ1-3'cctaagttcctttccaagaccat. Normalization to albumin amplification was performed as in (Jensen et al. 2016).Amplification of rearranged IgH DNA segments was based on the use of degenerated VH-gene family primers and a primer located just downstream of J_3 (Schlissel et al. 1991). The obtained PCR products were futher analyzed by southern blot and the correct products were detected by hybridization to a ^{32}P labeled internal J_3 oligonucleotide.

Assay for Transposase Accessible Chromatin (ATAC-seq).

Eighty thousand cells from the cultures of $CD19^+$ LN cells on OP9 or OP9-DL1, and control pMIG, C/EBPα-ER-pMIG or ICN1-pMIG transduced $CD19^+$ LN cells were washed in ice cold PBS prior to Assay for Transposase Accessible Chromatin (ATAC-seq) library preparation as described in (Buenrostro et al. 2013). Libraries were single-end sequenced on a NexSeq500. Each biological sample was processed for ATAC-Seq and sequenced in triplicate. The data was mapped to mm10 using Bowtie2 (Langmead et al. 2009; Langmead and Salzberg 2012) with standard settings. Tag directories with reads mapped to the mitochondrial chromosome filtered out and UCSC BedGraph files normalized to 10M total mapped reads were created

using the HOMER platform (Heinz et al. 2010) (*makeTagDirectory, makeUCSCfile)*. BedGraph files were up-loaded to the UCSC-genome browser (Kent et al. 2002) for visualization. ATAC-seq peaks were identified using *findPeaks.pl* in HOMER (Heinz et al. 2010) with the settings: *-style histone -size 75 -minDist 75 -minTagThreshold 6 -L 8 -F 8*. Published ChIP-seq data of EBF1-ChIP in pro-B cells (GSE69227), PAX5 (GSE38046), GATA3-ChIP in T-cell progenitors (FLND2b) (GSE31235) and C/EBPα-ChIP in Thioglycolate elicited macrophages (GSM537983) were retrieved from GEO, re-mapped to mm10 using Bowtie2 and peaks identified using HOMER with the –factor setting. Abundance of ATAC-tags on the ChIP-seq peaks was analyzed in HOMER using the annotatePeaks command.

Mutation analysis of human tumors.

To explore the relationship in mutational landscapes reported mutations in B-ALL, T-ALL, AML and ALAL was extracted from COSMIC Cancer gene Census database. Mutation frequency per reported tumor (AML n=26022, T-ALL n=1846, B-ALL n=1303 and ALAL n=52) was calculated. Genes with mutations less common than a frequency of 0.1% were excluded from the list. Mutation frequency of remaining genes in B-ALL, T-ALL, AML and ALAL were subject to heriarchical clustering using Cluster 3 (de Hoon et al. 2004) (log transform, center genes, average linkage clustering) and visualized in a heat map using Java TreeView (Saldanha 2004).

The Copy number and *NOTCH1* mutation status in human T-ALL samples was based on exome sequencing data from 67 human T-cell acute lymphoblastic leukemias (T-ALLs) and 17 T-ALL cell lines (De Keersmaecker et al. 2013). The data were acquired as aligned bam-files from the European Genome-phenome Archive (EGA),

accession number EGAS00001000296. This data was produced using two different exome capture kits (Nimblegen SeqCapEZ Exome v2.0 or Agilent SureSelect 38 Mb). The read coverage for unique regions from each of the two kits was therefore used to determine which of the kits was used to produce each bam-file. Global copy number profiles were calculated and visualized using CNVkit 0.7.7 (preprint article: Talevich et al, CNVkit: Copy number detection and visualization for targeted sequencing using off-target reads. *bioRxiv* doi: http://dx.doi.org/10.1101/010876). The *NOTCH1* mutation status presented in the original article (De Keersmaecker et al. 2013) was used for both human T-ALLs and cell lines.

References.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213-1218.

de Hoon MJ, Imoto S, Nolan J, Miyano S. 2004. Open source clustering software. *Bioinformatics* **20**: 1453-1454.

De Keersmaecker K, Atak ZK, Li N, Vicente C, Patchett S, Girardi T, Gianfelici V, Geerdens E, Clappier E, Porcu M et al. 2013. Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. *Nat Genet* **45**: 186-190.

Hamrouni A, Aublin A, Guillaume P, Maryanski JL. 2003. T cell receptor gene rearrangement lineage analysis reveals clues for the origin of highly restricted antigen-specific repertoires. *J Exp Med* **197**: 601-614.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576-589.

Jensen CT, Lang S, Somasundaram R, Soneji S, Sigvardsson M. 2016. Identification of Stage-Specific Surface Markers in Early B Cell Development Provides Novel Tools for Identification of Progenitor Populations. *J Immunol* **197**: 1937-1944.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**: 996-1006.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**: R36.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

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

Levin SD, Anderson SJ, Forbush KA, Perlmutter RM. 1993. A dominant-negative transgene defines a role for p56lck in thymopoiesis. *EMBO J* **12**: 1671-1680.

Lin H, Grosschedl R. 1995. Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* **376**: 263-267.

Mansson R, Zandi S, Anderson K, Martensson IL, Jacobsen SE, Bryder D, Sigvardsson M. 2008. B-lineage commitment prior to surface expression of B220 and CD19 on hematopoietic progenitor cells. *Blood* **112**: 1048-1055.

Mansson R, Zandi S, Welinder E, Tsapogas P, Sakaguchi N, Bryder D, Sigvardsson M. 2010. Single-cell analysis of the common lymphoid progenitor compartment reveals functional and molecular heterogeneity. *Blood* **115**: 2601-2609.

Pui JC, Allman D, Xu L, DeRocco S, Karnell FG, Bakkour S, Lee JY, Kadesch T, Hardy RR, Aster JC et al. 1999. Notch1 expression in early lymphopoiesis influences B versus T lineage determination. *Immunity* **11**: 299-308.

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

Saldanha AJ. 2004. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**: 3246-3248.

Schlissel MS, Corcoran LM, Baltimore D. 1991. Virus-transformed pre-B cells show ordered activation but not inactivation of immunoglobulin gene rearrangement and transcription. *J Exp Med* **1763**: 711.

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

Zandi S, Ahsberg J, Tsapogas P, Stjernberg J, Qian H, Sigvardsson M. 2012. Single-cell analysis of early B-lymphocyte development suggests independent regulation of lineage specification and commitment in vivo. *Proc Natl Acad Sci U S A* **109**: 15871-15876.