Leukemia (2017) **31,** 2503–2514

www.nature.com/leu

ORIGINAL ARTICLE

The T-ALL related gene *BCL11B* regulates the initial stages of human T-cell differentiation

VL Ha[1], A Luong[1], F Li[2], D Casero[3], J Malvar[1], YM Kim[1,4], R Bhatia[5], GM Crooks[3,6,7,8] and C Parekh[1,4]

The initial stages of T-cell differentiation are characterized by a progressive commitment to the T-cell lineage, a process that involves the loss of alternative (myelo-erythroid, NK, B) lineage potentials. Aberrant differentiation during these stages can result in T-cell acute lymphoblastic leukemia (T-ALL). However, the mechanisms regulating the initial stages of human T-cell differentiation are obscure. Through loss of function studies, we showed BCL11B, a transcription factor recurrently mutated T-ALL, is essential for T-lineage commitment, particularly the repression of NK and myeloid potentials, and the induction of T-lineage genes, during the initial stages of human T-cell differentiation. In gain of function studies, BCL11B inhibited growth of and induced a T-lineage transcriptional program in T-ALL cells. We found previously unknown differentiation stage-specific DNA binding of BCL11B at multiple T-lineage genes; target genes showed BCL11B-dependent expression, suggesting a transcriptional activator role for BCL11B at these genes. Transcriptional analyses revealed differences in the regulatory actions of BCL11B between human and murine thymopoiesis. Our studies show BCL11B is a key regulator of the initial stages of human T-cell differentiation and delineate the BCL11B transcriptional program, enabling the dissection of the underpinnings of normal T-cell differentiation and providing a resource for understanding dysregulations in T-ALL.

Leukemia (2017) **31,** 2503–2514; doi:10.1038/leu.2017.70

INTRODUCTION

T-cell differentiation is initiated in the human thymus by multilineage CD34+ progenitors that have arrived from the bone marrow (BM).[1] The earliest stages of thymopoiesis are characterized by a gradual commitment to the T-lineage (loss of alternative lineage potentials).[2,3] Dysregulations during these stages can result in T-cell acute lymphoblastic leukemia (T-ALL);[4–8] thus delineation of mechanisms underlying the early stages of human thymopoiesis is relevant to understanding T-cell leukemogenesis. Functional studies have yielded key insights about the mechanisms underlying murine T-lineage commitment.[9] However, critical immunophenotypic and regulatory differences between murine and human thymopoiesis have been recognized,[2,10–12] and there are few studies of the molecular processes mediating human T-lineage commitment. Thus, our understanding of the regulation of the initial stages of human thymopoiesis is incomplete.

CD34+ progenitors comprise $< 1\%$ of all human thymocytes.[3] The earliest thymic progenitors (CD34+CD7 − CD1a −) are multipotent, possessing myelo-erythroid as well as full lymphoid (B, T and NK) potential. Successive stages of T-lineage commitment are marked by the sequential upregulation of CD7 and CD1a, and a progressive loss of alternative (non T) lineage potentials, resulting in the generation of CD34+CD7+CD1a+ cells, the earliest fully T-lineage-committed progenitors, which produce double positive (CD4+CD8+) thymocytes.[1–3]

Bcl11b is a transcription factor whose expression during murine hematopoiesis is restricted to the T and innate lymphoid lineages.[13,14] Homozygous *Bcl11b* deletion induces a differentiation arrest during murine thymopoiesis.[13,15] A germ line *BCL11B* mutation was recently reported in a human T-cell immunodeficiency syndrome.[16] Somatic *BCL11B* mutations or deletions have been associated with T-ALL (9–16% of patients), and murine studies suggest a tumor suppressor role for *Bcl11b*.[17–19] *BCL11B* expression is reduced in Early T-cell precursor ALL (ETP-ALL), a T-ALL subtype with a differentiation arrest at an early stage of thymopoiesis, suggesting *BCL11B* insufficiency may underlie the pathogenesis of ETP-ALL.[7,20] However, the function of *BCL11B* in the initial stages of human T-cell differentiation has not been defined. Moreover, the DNA binding targets of BCL11B during thymopoiesis, and thereby its role in the regulation of T-lineage transcription networks, are largely undelineated. As understanding the regulation of the earliest stages of T-cell differentiation is important for the elucidation of T-cell leukemogenesis, we investigated the role of *BCL11B* in the initial stages of human T-cell differentiation. We show *BCL11B* is critical for human T-lineage commitment, particularly the induction of T-lineage genes and the repression of NK and myeloid potentials. Important differences were observed between the regulatory actions of *BCL11B* in humans and mice. Genome-wide analyses demonstrated previously unknown stage-specific BCL11B binding at multiple T-lineage genes, revealing a previously undescribed BCL11B transcriptional program.

[1]Children's Center for Cancer and Blood Disease, Children's Hospital Los Angeles, Los Angeles, CA, USA; [2]MiNGS Core Laboratory, Children's Hospital Los Angeles, Los Angeles, CA, USA; [3]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine University of California, Los Angeles, Los Angeles, CA, USA; [4]Department of Pediatrics, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; [5]Division of Hematology and Oncology, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA; [6]Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, Los Angeles, CA, USA; [7]Department of Pediatrics, David Geffen School of Medicine at University of California Los Angeles, Los Angeles, CA, USA and [8]Jonsson Comprehensive Cancer Center, University of California Los Angeles, Los Angeles, CA, USA. Correspondence: Dr C Parekh, Department of Pediatric Hematology/Oncology, Children's Hospital Los Angeles, 4650 Sunset Boulevard, Mail Stop #54, Los Angeles, CA 90027, USA.
E-mail: cparekh@chla.usc.edu
Received 22 August 2016; revised 16 January 2017; accepted 15 February 2017; accepted article preview online 24 February 2017; advance online publication, 14 March 2017

MATERIALS AND METHODS

Lentiviral vectors

Polymerase III (U6/H1) or polymerase II (MNDU3) shRNA lentiviruses were used to knockdown (KD) *BCL11B* (Supplementary Figure 1b; Supplementary Table 1). The *BCL11B* expression lentiviral plasmid was made by inserting a PCR amplified *BCL11B* cDNA sequence from the *BCL11B* ORF plasmid (ThermoFisher Scientific, Waltham, MA, USA) into the MNDU3-PGK-GFP plasmid.

Primary tissues

Deidentified cord blood (CB), human thymuses, normal BM and a primary relapsed T-ALL (CD34+CD1 − CD2+CD3 − CD5+CD7+CD4 − CD8+cytoplasmic CD3+intracellular T-cell receptor beta chain+; translocation 1;9, p34; q34) BM sample were obtained via University of California Los Angeles, Hollywood Presbyterian Hospital, Children's Hospital Los Angeles (CHLA), or the Children's Oncology Group (informed consent obtained as per CHLA IRB approved protocols). Primary T-ALL cells were expanded via serial transplantations in NOD/SCID/IL2Rγ − / − (NSG) mice (primagraft cells).[21]

Transduction and culture of CB and thymic CD34+ cells

CD34+ cells were enriched from CB or human thymuses using magnetic-activated cell sorting (MACS, Miltenyi Biotec, San Diego, CA, USA). CD34+ CB cells were transduced with shRNA (multiplicity of Infection (MOI) = 10) lentivirus. CB or thymic CD34+ cells were transduced with *BCL11B* expression or control lentivirus (MOI = 1). CD34+GFP+lin − CB or CD34 +CD1a − GFP+lin − thymic cells were sorted (fluorescence activation cell sorting, FACS) and then co-cultured with OP9-DLL1[22] or MS5 stroma, respectively. In some KD experiments (experiments 1 and 3), transduced cells were co-cultured with OP9-DLL1 without prior FACS. The medium for OP9-DLL1 co-cultures medium consisted of alpha-MEM (ThermoFisher), 20% fetal bovine serum (FBS, Hyclone, lot no. AXF42576), L-glutamine (2 mM, Cellgro, Manassas, VA, USA), pencillin–streptomycin (0.5 ×, CellGro), IL-7 (5 ng/ml) and Flt3 ligand (5 ng/ml). Cells were harvested on day 11–17, counted and analyzed by flow cytometry. To assess alternative lineage potentials of T-cell precursors, populations isolated from CB-OP9-DLL1 co-cultures (CD45+GFP+CD7+CD1a − , CD45+GFP+CD7+CD1a+) or primary thymocytes by FACS were plated on the MS5 (murine BM stromal) cell line in: (1) DMEM (ThermoFisher), 10% FBS, L-glutamine, pencillin–streptomycin, 2-mercaptoethanol (50 μM, ThermoFisher), IL-3 (5 ng/ml), Tpo (50 ng/ml), FLT-3 ligand (5 ng/ml), SCF (5 ng/ml), GM-CSF (5 ng/ml) and EPO (4 ng/ml) (myelo-erythroid conditions); (2) RPMI 1640 (ThermoFisher), 5% FBS, L-glutamine, pencillin–streptomycin, 2-mercaptoethanol (50 μM, ThermoFisher), Tpo (5 ng/ml), FLT-3 ligand (5 ng/ml) and IL-7 (5 ng/ml; B/NK conditions); and (3) DMEM, 10% FBS, L-glutamine, pencillin–streptomycin, 2-mercaptoethanol (50 μM), IL-15 (10 ng/ml) and SCF (20 ng/ml) (NK conditions). Cytokines were purchased from Milteyni (except IL-15, GM-CSF and Epo (Peprotech, Rocky Hill, NJ, USA)). Supplementary Table 2 lists FACS antibodies.

Culture, transduction and xenotransplantation of T-ALL cells

Primagraft cells were cultured in alpha-MEM, 20% FBS, L-glutamine and penicillin–streptomycin. MOLT-16 and CCRF-CEM were purchased from ATCC (Manassas, VA, USA) and DSMZ (Braunschweig, Germany), respectively, and cultured as per their recommendations. *TP53* mutation testing was done by Sanger sequencing (Supplementary Table 3). Cells were transduced with *BCL11B* expression or control lentivirus (MOI = 1). GFP+ cells isolated by FACS 72 h post transduction were then cultured *in vitro*, transplanted into sublethally irradiated (250 cGY) 6–9-week-old female NSG mice by tail vein injection (5000–17 000 cells per mouse, age-stratified randomization, sample size not estimated *a priori*, unblinded) or used for RNA-Seq. Animals were used as per IACUC approved protocols and killed when they developed clinical signs of leukemia.

Western blot

Primary antibodies: Rabbit anti-BCL11B (D6F1, Cell Signaling, Danvers, MA, USA), anti Beta-Actin (Genetex, Irvine, CA, USA), or Anti-alpha Tubulin (EPR13478 (B), Abcam, Cambridge, MA, USA). Secondary antibodies: goat anti-rabbit or anti-mouse HRP-linked IgG (Cell Signaling).

RNA-Seq

The Ovation Human FFPE RNA-Seq System (NuGen, San Carlos, CA, USA) was used to generate amplified cDNA libraries (input = 25 ng RNA extracted using the Qiagen MIrneasy kit, Valencia, CA, USA), which were sequenced on the Illumina Nextseq500 (paired end 75 bp, average of 13 million (from T-cell precursors) or 21 million (from T-ALL cells) paired-end reads per sample, http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token = efelwcoqlbwbxcd&acc = GSE84678). Data was analyzed using Tophat,[23] HTSeq,[24] DESeq2[ref. 25] and GSEA.[26] The methods for differential expression analysis of previously generated RNA-Seq data have been described.[27]

ChIP-Seq (chromatin immunoprecipitation followed by deep sequencing)

CD34+ and CD34 − cells were isolated from human thymuses by MACS. ChIP was done using anti-Bcl11b or isotype control antibody (Cell Signaling). Libraries made with the Next Ultra DNA Library Prep Kit (NEB, Ipswich, MA, USA) (input = 5 ng ChIP DNA) were sequenced on the NextSeq500 (paired end 75 bp, average of 25 million paired end reads per sample, http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token = efelw coqlbwbxcd&acc = GSE84678). Data was analyzed using Bowtie2,[28] MACS2,[29] the MEME-CHIP suite,[30,31] a differential binding analysis strategy similar to Diffbind (http://www.bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf),[32] GREAT[33] and the Kolmogorov–Smirnov test (to analyze differences between cumulative distributions).[34]

Quantitative PCR

The RNeasy micro and the omniscript reverse transcription kits (Qiagen) were used to extract RNA and generate cDNA, respectively. QPCR TaqMan assays (ThermoFisher) used: Hs00256257_m1, (*BCL11B)*, Hs00607336_gH (*HSP90)*, Hs01060665_g1, *(ACTB*).

Statistical analysis

Linear mixed models with random effects for CB donors were utilized to analyze KD experiments. For T-ALL experiments, paired *t*-tests, ANOVA and the log-rank test were used to analyze *in vitro* data, *in vivo* data and survival, respectively.

Additional details are provided in Supplementary Methods.

RESULTS

BCL11B expression during human thymopoiesis

To assess the expression of *BCL11B* during human thymopoiesis, we analyzed previously generated RNA-Seq data from BM hematopoietic stem cells (HSC), lymphoid progenitors from the BM and thymus, and CD4+CD8+ (Thy4) cells, the predominant fully T-lineage committed population in the thymus[27,35] (developmental hierarchy of populations depicted in Figure 1a). *BCL11B* was not expressed in BM cells and first became detectable in CD34+ thymic progenitors. A progressive increase in expression was seen with successive stages of T-lineage commitment and differentiation (Figure 1b). Similar results were seen with qPCR (Supplementary Figure 1a). In summary, *BCL11B* expression is induced early on during T-cell differentiation in the human thymus, and upregulation of *BCL11B* is associated with T-lineage commitment and subsequent differentiation into CD4+CD8+ cells.

BCL11B is essential for human T-lineage commitment

The role of *BCL11B* in the initial stages of human T-cell differentiation was investigated through loss of function experiments in multi-lineage CB CD34+ progenitors using the *in vitro* OP9-DLL1 stromal co-culture system. The earliest stages of T-cell differentiation and commitment in the thymus are recapitulated in the OP9-DLL1 system through the sequential appearance of multi-lineage CD7+CD1a − and T-lineage restricted CD7+CD1a+ T-cell precursors, the latter emerging at day 10–14 of culture.[22,36] Although the earliest fully T-lineage committed cells *in vivo* are CD34+, in the OP9-DLL1 system the induction of CD1a expression and concomitant T-lineage restriction coincide with the loss of

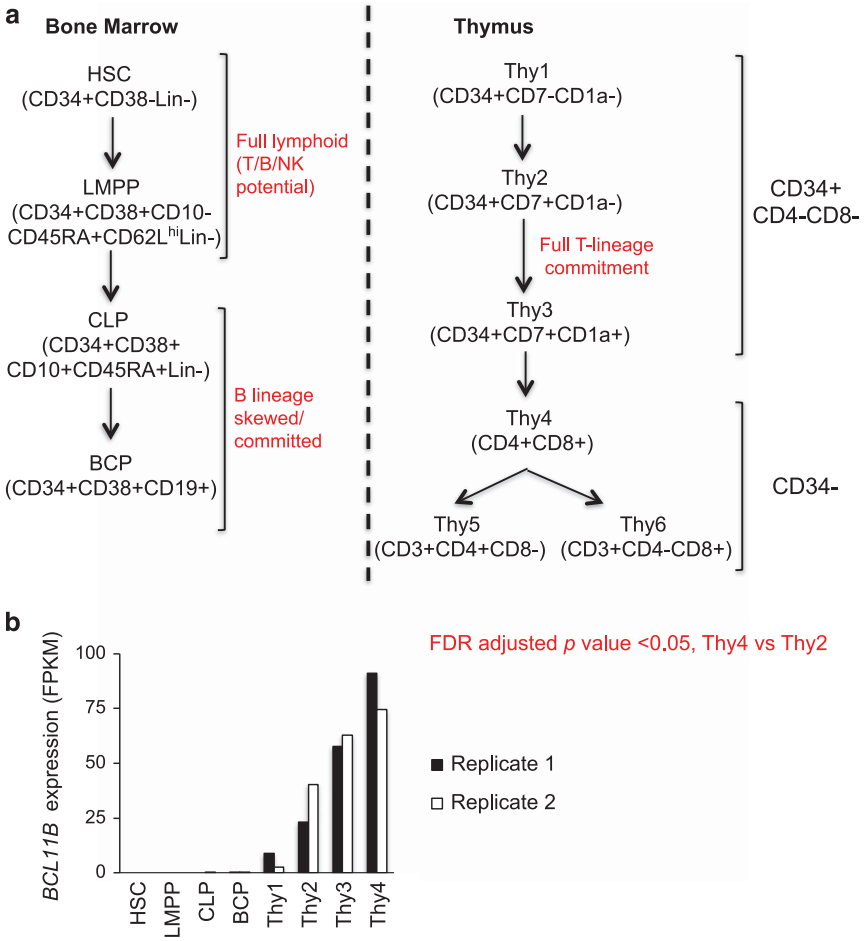


Figure 1. *BCL11B* expression is induced during the initial stages of T-cell differentiation in the human thymus, and upregulation of *BCL11B* is associated with T-lineage commitment and subsequent differentiation. (**a**) Differentiation schema depicting lymphoid progenitors in the human bone marrow (BM) and lymphoid cell types in the human thymus. (**b**) RNA-Seq data for expression of *BCL11B* in hematopoietic stem and lymphoid progenitor cells from human BM and CD34+CD4 – CD8 – progenitor cells (Thy1-3) and double positive (CD4+CD8+, Thy4) cells from the human thymus ($n = 2$ biological replicates per cell type) is depicted. BCP, B-cell committed progenitors; CLP, common lymphoid progenitors; FPKM, fragments per kilobase per million reads; HSC, hematopoietic stem cells; LMPP, lymphoid primed multipotent progenitors.

CD34 expression.[36] Hence, the value of CD34 as a phenotypic marker is limited in the context of dissection of commitment during the initial stages of *in vitro* T-cell differentiation. As our studies were focused on the initial stages of thymopoiesis, we analyzed OP9-DLL1 co-cultures on approximately day 14 of culture to measure differentiation into T-cell precursors, defined as CD7+CD1a – and CD7+CD1a+ cells (Figure 2a).

CB CD34+ cells were transduced with a *BCL11B* KD or scrambled control lentivirus and then cultured on OP9-DLL1 stroma. Four different shRNA sequences (s1, s2, s3, s4) that induced robust KD were used (Supplementary Figures 1b and c). Control and KD co-cultures showed similar percentages of CD7+ cells (Figure 2b). However, KD cells showed significantly reduced differentiation into CD7+CD1a+ T-cell precursors (mean % CD7+CD1a+ cells in control vs KD co-cultures = 30.88 vs 17.85%, $P < 0.05$, Figures 2c and d), indicating a differentiation arrest of KD cells at the CD7+CD1a – stage (Supplementary Figure 2a). Also, KD co-cultures showed higher frequency of CD34+ cells with both KD shRNA sequences in two experiments and one of two shRNA sequences in a third experiment (Supplementary Figure 2b), a finding consistent with an impairment of differentiation in KD co-cultures.

KD of *BCL11B* decreased cell output (Supplementary Figure 2c). To verify that *BCL11B* insufficiency resulted in impaired T-lineage commitment and that the immunophenotypic differences between KD and control co-cultures were not simply due to increased death of CD7+CD1a+ cells in KD co-cultures, we tested whether *BCL11B* insufficiency had an effect on the alternative (non T) lineage potentials of T-cell precursors. We isolated CD7+CD1a – and CD7+CD1a+ cells from KD and control OP9-DLL1 co-cultures and then re-cultured them in conditions supportive of myeloid, erythroid, B and/or NK lineage differentiation (secondary cultures, Figure 3a). Cells sorted from KD OP9-DLL1 co-cultures showed robust *BCL11B* KD (Figure 3b). Consistent with the association of CD1a expression with T-lineage commitment, control CD7+CD1a+ cells generated substantially lower numbers of NK (CD56+ and CD56+CD11c+[37]) and myeloid (CD33+ and CD14/15+) cells relative to those generated by control CD7+CD1a – cells (Figure 3c). These findings recapitulate the loss in alternative lineage potential seen with CD1a expression in primary thymic progenitors (Supplementary Figure 3a), validating the secondary cultures as an assay for investigating T-lineage commitment. KD of BCL11B resulted in the generation of significantly higher numbers of NK and myeloid cells in secondary cultures (Figures 3c and d). Almost no erythroid (CD235+) or B (CD19+) cells were generated in KD or control secondary cultures (data not shown). Of note, the effect of KD on alternative lineage cell outputs significantly varied ($P < 0.05$) with cell type (CD7+CD1a – or CD7+CD1a+). The more pronounced effects of KD in cultures initiated with CD7+CD1a+

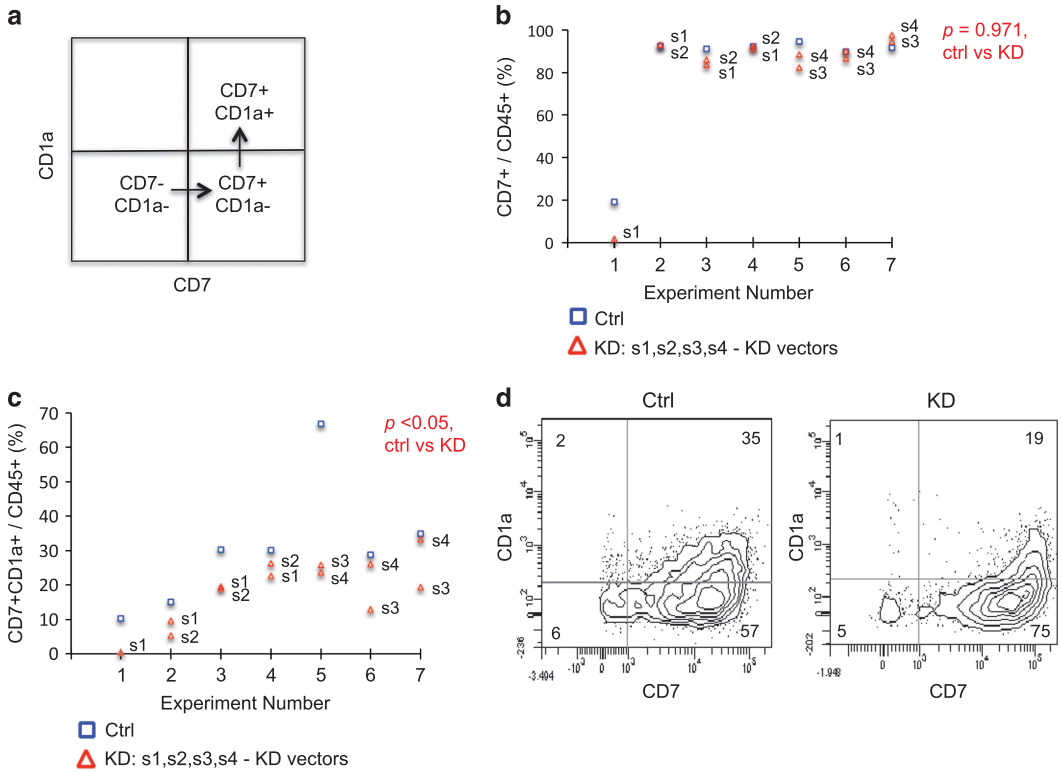


Figure 2. *BCL11B* is required for normal differentiation during the initial stages of human thymopoiesis. CD34+ CB progenitors were transduced with *BCL11B* KD or scramble control (ctrl) shRNA lentivral vectors and then cultured on OP9-DLL1 stroma to induce T-lineage differentiation. (**a**) The plot shows a schema of the initial stages of thymopoiesis that are typically seen in the OP9-DLL1 co-culture system when co-culture is initiated with normal CD34+CB cells. (**b**–**d**) Results from flow cytometry on day 11–17 of culture for % CD7+ cells (**b**) and % CD7+CD1a+ cells (**c**) are depicted ($N = 7$ independent experiments, each experiment done with a different pool of CB donors). (**d**) Immunephenotype analysis results from one representative experiment (results shown in **d** are from KD with shRNA sequence s3). Results shown are from cells pregated on CD45+GFP+ cells for **b**, **c** and **d**. Experiments were done using four *BCL11B* KD shRNA sequences (s1, s2, s3 and s4. s1, s2: polymerase III promoter vectors; U6 vectors were used in experiments 1–3 and H1 vectors were used in experiment 4. s3, s4 – polymerase II promoter vectors). All co-cultures in a given experiment were analyzed at the same time point.

cells than in those initiated with CD7+CD1a − cells (Figure 3c) is consistent with the multi-lineage potential normally seen in thymic progenitors prior to the onset of CD1a expression.

To further investigate the effects of BCL11B on T-lineage commitment, we performed gain of function experiments in uncommitted CD34+ (CD34+CD1a −) progenitors isolated from human thymuses. In agreement with the increased NK and myeloid output seen with *BCL11B* insufficiency in the KD experiments, *BCL11B* overexpression inhibited myeloid and NK potential of thymic CD34+CD1a − progenitors (Supplementary Figure 3b).

Overall, *BCL11B* insufficiency impairs T-lineage commitment, in particular prevents the repression of NK and myeloid lineage potential, and induces a differentiation arrest during the initial stages of human T-cell differentiation.

BCL11B overexpression inhibits growth of T-ALL cells

To study the effects of *BCL11B* in T-ALL cells, we transduced T-ALL cells with *BCL11B*-GFP (overexpressing cells) or control GFP (control cells) lentivirus. T-ALL cells at varying stages of differentiation arrest were used (xenografted cells derived from a CD34+ T-ALL primary sample (primagraft), and CCRF-CEM (pre-T) and MOLT-16 (mature T) CD34 − cell lines,[38] BCL11B protein levels depicted in Supplementary Figure 4a). Both cell lines and primagraft cells harbored *TP53* mutations;[39] a *NOTCH1* mutation was present in primagraft cells (Supplementary Table 4). A low MOI (= 1) was used to reduce effects secondary to supraphysiologic BCL11B levels (Supplementary Figures 4b and c). *BCL11B* overexpression significantly inhibited the growth of T-ALL cells *in vitro* (Figure 4a; Supplementary Figure 5a). The inhibitory effect of BCL11B in leukemic cells was in contrast to its effects in CB CD34+ progenitors, where overexpression did not inhibit cell growth (Supplementary Figure 5b). *BCL11B* overexpression significantly increased apoptosis of primagraft cells (Figure 4b). NSG mice transplanted with overexpressing or control primagraft cells rapidly developed leukemia (median survival time, control vs BCL11B = 40 days vs 41.5 days, $P = 0.74$). However, overexpression decreased the growth of T-ALL cells after transplantation in terms of cell numbers (Figure 4c; Supplementary Figure 5c). Also, human T-ALL cells in mice transplanted with overexpressing cells showed a lower proportion of GFP+ cells than control mice (Supplementary Figure 5d), suggesting intense selection pressure against overexpressing cells. Overall, *BCL11B* overexpression inhibits growth of T-ALL cells, an effect that can at least be seen in TP53 mutant cells.

BCL11B is required for the establishment of a T-lineage commitment transcriptional program

To define the transcriptional effects of *BCL11B* insufficiency during the initial stages of T-cell differentiation, we performed RNA-Seq on CD7+CD1a − and CD7+CD1a+ cells isolated from control and KD OP9-DLL1 co-cultures. KD cells showed upregulation of

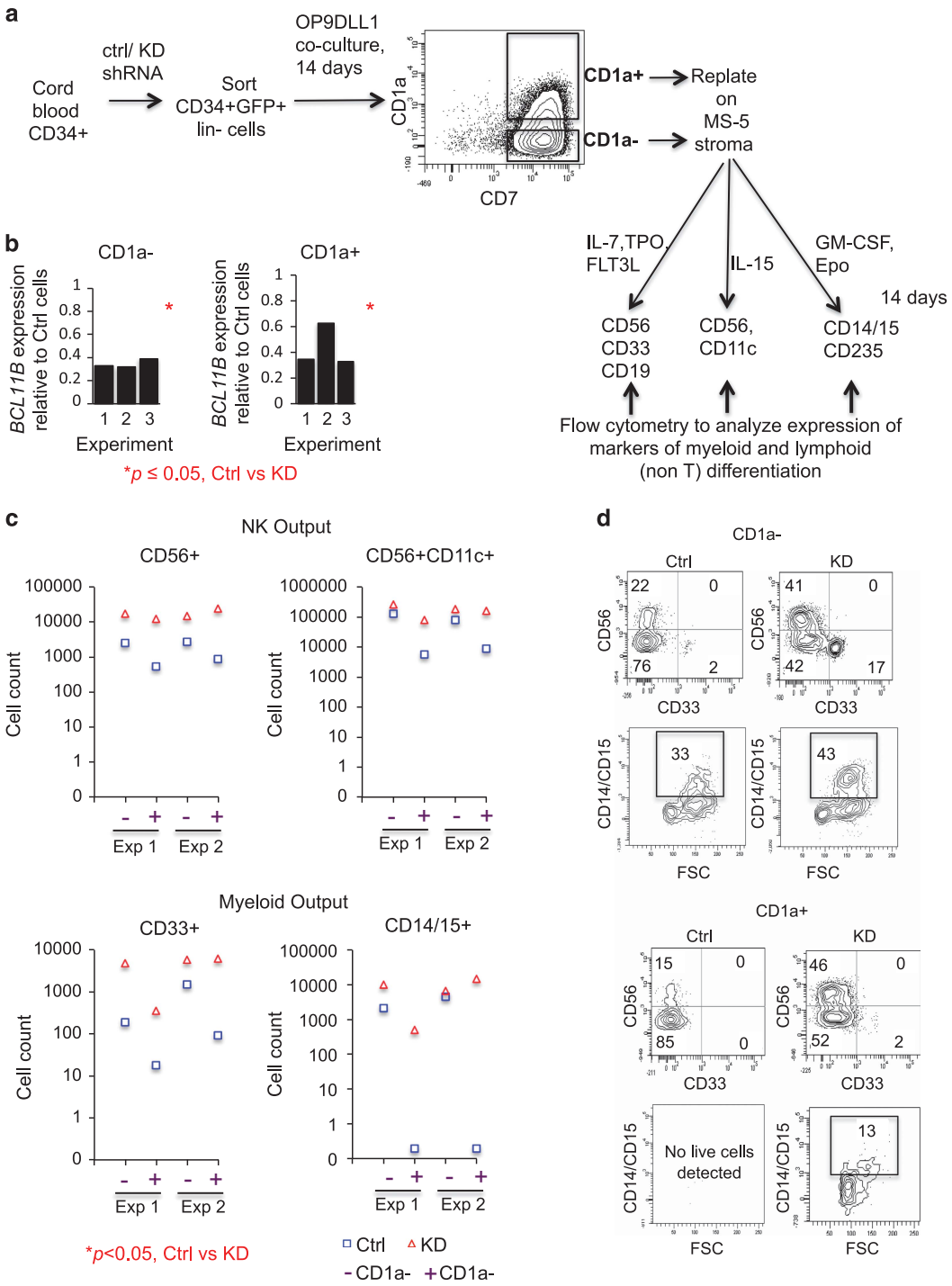


Figure 3. *BCL11B* is required for T-lineage commitment, particularly the repression of NK and myeloid lineage potentials, during the initial stages of human thymopoiesis. (**a**) CB CD34+ progenitors were transduced with control (ctrl) or BCL11B KD shRNA lentivirus and then co-cultured with OP9-DLL1 stroma. CD7+1a− (CD1a−) and CD7+CD1a+ (CD1a+) T-cell precursors generated in these OP9-DLL1 co-cultures were sorted and replated in three different culture conditions that promote alternative (non T, that is, myeloid, erythroid, B and/or NK) lineage differentiation (MS5 stroma co-cultures). (**b**) *BCL11B* gene expression in T-cell precursors sorted from OP9-DLL1 co-cultures ($N=3$ independent experiments, each done with a different pool of CB donors; expression measured by qPCR relative to that of *ACTB*; $P<0.05$ for ctrl CD1a− vs KD CD1a−, $P=0.05$ for ctrl CD1a+ vs KD CD1a+). (**c**) Cell counts for NK (CD56+, CD56+CD11c+) and myeloid (CD33, CD14/15+) lineage cells generated in MS5 co-cultures of control and KD CD1a− and CD1a+ cells (analyzed on day 14 of MS5 co-culture, cell counts are averages of triplicate counts). $N=2$ independent experiments, each done with a different pool of CB donors. (**d**) Immunephenotype analysis results for MS5 co-cultures initiated with CD1a− or CD1a+ cells (data from one representative experiment is depicted). Cells were pregated on CD45+GFP+ cells for the sorting on day 14 of OP9-DLL1 co-culture (**a**, **b**), and the analysis after 14 days of co-culture with MS5 stroma (**c**, **d**). Experiments were done using the KD shRNA sequence s3.

hematopoietic stem and progenitor (*BAALC*, *GFI1B*, *HOXA9, CD34)* and myeloid (*MEF2C*, *MPO*) genes and downregulation of T-lineage genes (*PTCRA*, *RAG1*, *RAG2*, *LCK*, *ZAP70*, *THEMIS*). Also, KD induced upregulation of genes important for NK cell development (*DTX1*)[12] and function (*ZBTB16*),[40] and downregulation of *IL7R*, a gene critical for T-cell differentiation. NOTCH1 signaling is known to be decreased during human T-lineage commitment;[12] KD cells, however, showed upregulation of NOTCH1 target genes (*DTX1*, *NRARP*, *MYC*)[12] (Figures 5a and b). Overall, KD resulted in downregulation of genes that are normally induced with T-lineage commitment, and the aberrantly sustained expression of a transcriptional program characteristic of the earliest stage of human thymopoiesis (Thy1, CD34+CD7 − CD1a − ; Figures 5a and b; Supplementary Figures 6a and b).

We next investigated the transcriptional effects of *BCL11B* through gain of function studies in cells with a T-lineage differentiation arrest, namely T-ALL cells.[4] Overexpressing and control primagraft T-ALL cells were analyzed by RNA-Seq at 72 h post transduction. We chose this early time point to minimize the impact of proliferation, survival or profound differentiation state differences between cells on gene expression. Control T-ALL cells showed high *LYL1*, *LMO2* and *TAL1* expression levels (Supplementary Figure 6c), an expression pattern characteristic of a subgroup of the *LYL1*+ T-ALL subtype.[4] To resolve the

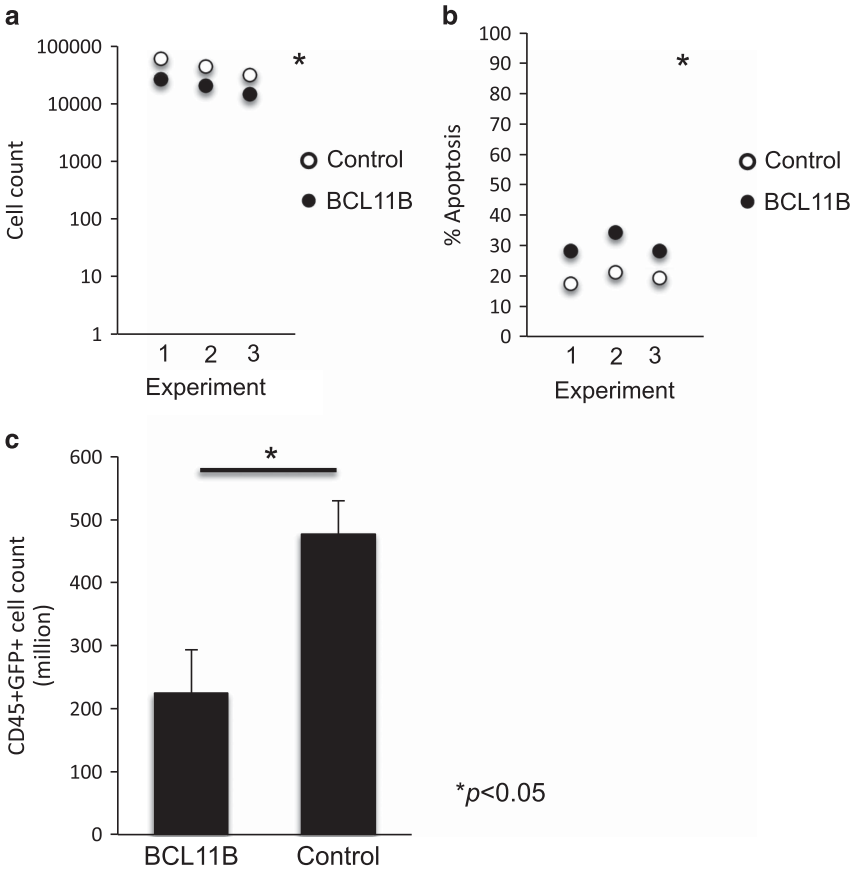
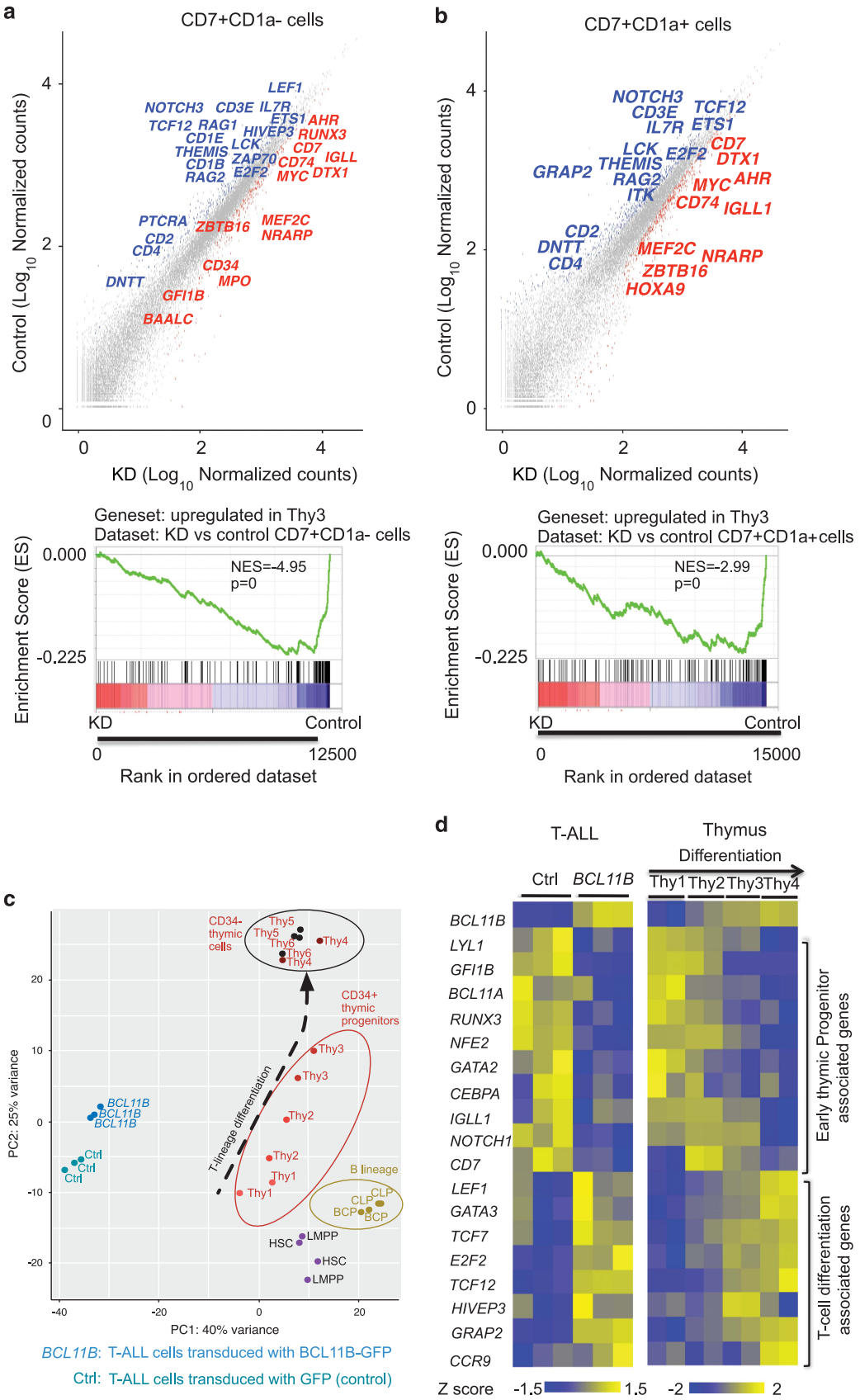


Figure 4. *BCL11B* overexpression inhibits growth of T-ALL cells. Primagraft T-ALL cells were transduced with a BCL11B cDNA-GFP vector (*BCL11B* cells) or a control GFP vector (control cells). GFP+ cells were cultured *in vitro* or transplanted into NSG mice. (**a**) Cell counts on day 7 of culture (starting count = 10 000 cells per arm). (**b**) % apoptotic cells (Annexin V+ cells as measured by flow cytometry) on day 7 of culture of transduced primagraft T-ALL cells. (**c**) Numbers of human CD45+GFP+ cells in spleens of transplanted mice. (**a**, **b**) $N = 3$ experiments, each done in triplicate (cell counts are averages of triplicate counts). (**c**) $N = 2$ experiments, $n = 10$ mice for control cells, $n = 10$ mice for BCL11B cells. Graphs depict means; error bars: s.e.m.

Figure 5. *BCL11B* is required for the establishment of a T-lineage commitment transcriptional program through the repression of stem, NK and myeloid genes, and the upregulation of T-lineage genes. (**a**, **b**) CB CD34+ progenitors were transduced with control or BCL11B KD shRNA lentivirus and then co-cultured with OP9-DLL1 stroma. CD7+CD1a − (**a**) and CD7+CD1a+ (**b**) T-cell precursors generated in these co-cultures were sorted on day 14 and analyzed by RNA-seq ($n = 3$ biological replicates, i.e., each replicate from a different pool of CB donors; experiments done using the KD shRNA sequence s3). Scatter plots show differentially expressed genes (false discovery rate (FDR) < 0.05 for KD vs control, red dots: genes upregulated in KD, blue dots: genes upregulated in control). GSEA: enrichment analyses of genes upregulated in Thy3 cells (CD34+CD7+CD1a+ thymic progenitors) among genes ranked based on KD vs control expression (FDR-adjusted GSEA *P*-values shown, NES: normalized enrichment score). (**c**, **d**) Primagraft T-ALL cells were transduced with BCL11B cDNA-GFP (*BCL11B* cells) or control GFP (control cells) vector ($n = 3$ replicates, each replicate from an independent experiment). GFP+ cells were sorted and analyzed by RNA-Seq. Principal component analysis (**c**) includes T-ALL cells, and populations from normal human BM and thymus ($n = 2$ biological replicates per population) and is based on expression values of genes differentially expressed between overexpressing and control T-ALL cells (FDR < 0.05, $n = 420$ genes). (**d**) Heatmaps: early thymic progenitor and T-cell differentiation-associated genes differentially expressed between control and *BCL11B* T-ALL cells (left), and their expression in populations from normal thymus (right).

differences between overexpressing and control cells with respect to transcriptional changes during normal lymphopoiesis, we performed a principal component analysis of normal hematopoietic cell types[27] and these T-ALL cells, using expression values of genes that were differentially expressed between overexpressing and control T-ALL cells (FDR < 0.05, $n = 420$ genes)



(Figure 5c). Expectedly, the first principal component separated T-ALL cells from normal cells, whereas the second principal component recapitulated the differentiation of early uncommitted BM progenitors (HSC and LMPP) along T-lineage or B-lineage pathways. In particular, the trajectory towards T-lineage differentiation followed increasing values for the second principal component. Strikingly, overexpressing cells and control cells showed a trajectory parallel to that of the early (CD34+) stages of T-lineage differentiation in the thymus, with overexpressing cells being positioned further along the T-lineage pathway than control cells (Figure 5c). Genes induced during normal T-cell differentiation were upregulated and those genes expressed in early thymic progenitors and repressed during normal T-cell differentiation were downregulated, in T-ALL cells that overexpressed *BCL11B* (Figure 5d).

In summary, *BCL11B* is a key regulator of the T-lineage commitment transcriptional program through the repression of stem, NK and myeloid genes and the upregulation of T-lineage genes.

BCL11B DNA binding during T-cell differentiation

The DNA binding targets of BCL11B in thymic progenitors, and changes in binding during differentiation, remain largely unidentified. To define binding targets, we performed ChIP-Seq for BCL11B on two populations isolated from human thymuses: CD34+ progenitors and the more differentiated CD34 − cells. Consistent with the upregulation of *BCL11B* expression with differentiation (Figure 1b), CD34 − cells showed over twice the binding sites seen in CD34+ cells (Figure 6a; Supplementary Table 5). Almost all sites seen in CD34+ cells were retained in CD34 − cells, and top ranked site motifs were similar in both populations, indicating similar site preferences in both cell types (Supplementary Figure 7a). Notably, the motif (–GGCCGG–) for the highly homologous transcription factor BCL11A[41] was one of the most significantly enriched motifs (Supplementary Figure 7a). Most sites were associated with promoters of protein coding or lncRNA genes (Figures 6b and c), and included the previously reported binding targets, *ID2* and *P57*.[42,43]

To define differentiation stage related differences in binding, we performed differential binding analysis (Supplementary Table 6). High inter-replicate reproducibility and distinct binding profiles for the two cell types were observed (Figure 7a; Supplementary Figure 7b), indicating global changes in binding with differentiation. Most differentially bound genes showed greater binding in CD34 − cells than in CD34+ cells, indicating an overall gain in binding with differentiation (Figure 7b).

Genes with greater binding in CD34 − cells were significantly enriched for T-lineage genes (Supplementary Figure 7c; Supplementary Table 7). To investigate the association between binding and expression of target genes, we performed a combined analysis of the ChIP-Seq data and previously generated RNA-Seq data from CD34+ and CD34 − cell types in the thymus.[26] Thy4 cells, which are fully T-lineage committed, constitute a majority of the thymic CD34 − cells.[35] CD34+ thymic cells are almost entirely made up of uncommitted Thy2 (mean = 64.7% of CD34+ thymocytes, s.e.m. = 8.6%, $n = 8$ thymuses) and fully

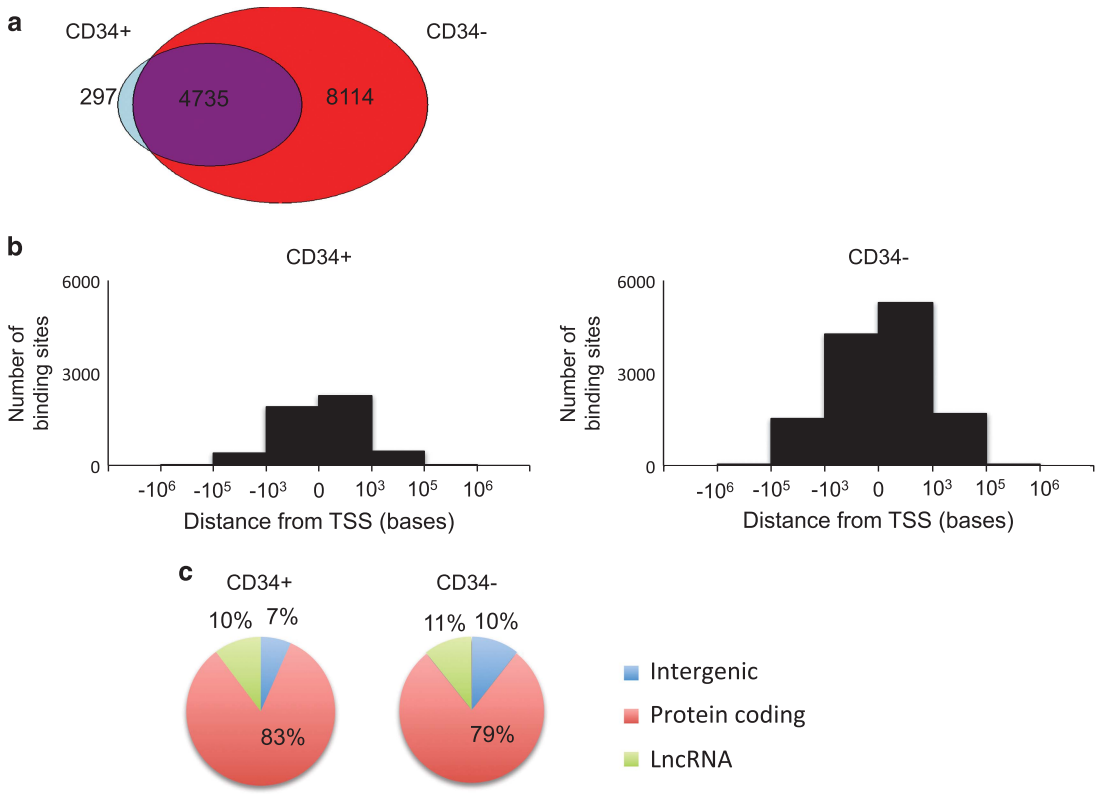


Figure 6. BCL11B predominantly binds to promoters (±1000 bases from transcription start site) of coding and non-coding genes during thymopoiesis, with a gain in binding sites with T-cell differentiation. Progenitors (CD34+) and differentiated cells (CD3 −) were isolated from human thymic tissue ($n = 2$ biological replicates per cell type), and analyzed by ChIP-Seq for BCL11B DNA binding. Binding sites (ChIP-seq peaks) seen in both replicates of at least one cell type were analyzed for: (**a**) numbers of binding sites seen in CD34+ cells only, CD34 − cells only, and both CD34+ and CD34 − cells. (**b**) Distribution of binding sites with respect to distance from transcription start sites of genes. (**c**) Distribution of binding sites with respect to genomic annotation (intergenic, overlapping with or next to a protein coding gene, overlapping with or next to a lncRNA gene. Default HOMER parameters were used to assign genomic annotations).

committed Thy3 cells.[2,3] Given the continuum of transcriptional changes during thymopoiesis, we selected RNA-Seq data from Thy2 and Thy4 cells for the analysis. We used two approaches previously used to study relationships between binding and expression.[34] First, we classified BCL11B bound genes that were differentially expressed into three groups based on binding

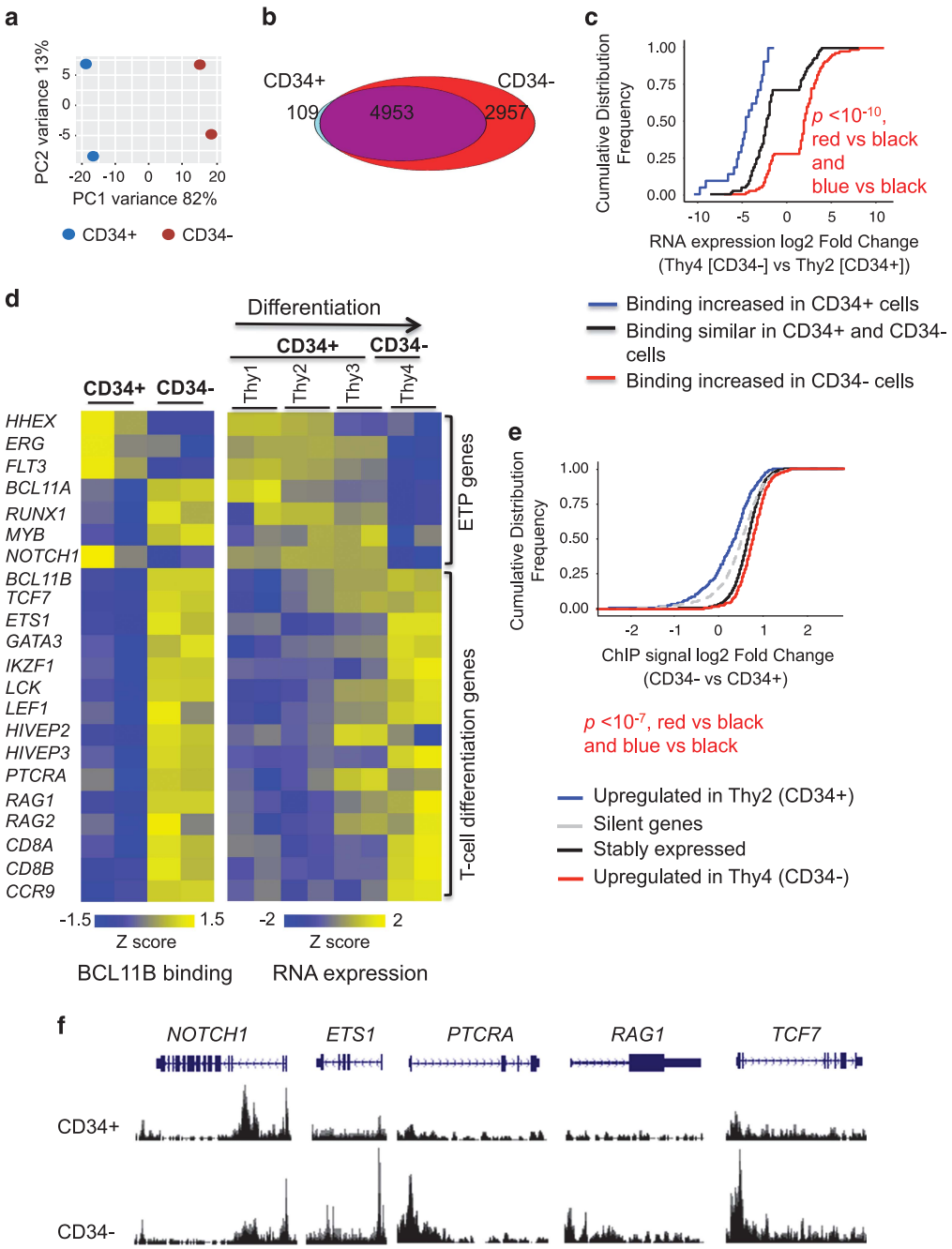


Figure 7. BCL11B shows differentiation stage-specific DNA binding at multiple genes important for T-cell differentiation, and BCL11B binding is associated with upregulation in expression of target genes during thymopoiesis. CD34+ and CD34− cells from human thymuses ($n=2$ biological replicates) were analyzed by ChIP-Seq for BCL11B DNA binding. (**a**) Principal component analysis based on ChIP read counts. (**b**) Numbers of differentially bound genes (FDR < 0.05 for CD34+ vs CD34−) and genes with similar binding in both cell types (FDR ⩾ 0.05 and ChIP-Seq peak in at least one replicate for each cell type). (**c**–**e**) Associations between gene expression and BCL11B DNA binding, based on expression data from early thymic progenitors (CD34+CD7+CD1a−, Thy2) and differentiated thymic cells (CD34−CD4+CD8+, Thy4). (**c**) Differentially expressed (FDR < 0.05 for Thy2 vs Thy4) BCL11B bound genes were classified into three groups based on binding. Cumulative distributions of expression fold changes (fold change > 0 if Thy4 > Thy2) for these groups are depicted. (**d**) Heatmaps depict BCL11B ChIP read counts (left) and gene expression (right) for differentially bound early thymic progenitor associated (ETP) genes and T-cell differentiation-associated genes. (**e**) BCL11B bound genes were classified into four groups based on gene expression, Stable: similarly expressed in Thy2 and Thy4 (FPKM > 1 in both replicates of both cell types). Silent: not expressed in either Thy2 or Thy4 cells. Cumulative distributions of binding fold changes (fold change > 0 if CD34- > CD34+) are depicted for these groups. Kolmogorov–Smirnov test *P*-values are shown (**c**, **e**). (**f**) BCL11B binding at a subset of target genes (UCSC Genome browser tracks).

changes (increased binding in CD34+ cells, increased binding in CD34− cells, or similar binding in both, Supplementary Table 8). Analysis of expression fold change values for these groups showed that overall, increased binding was associated with an upregulation in expression (Figure 7c). For instance, genes with increased binding in CD34+ cells (*NOTCH1*, *FLT3*, *ERG*) were upregulated in Thy2 cells, and genes with increased binding in CD34− cells (*BCL11B*, *MTA3*, *CD8*) were upregulated in Thy4 cells (Figure 7d), a relationship also seen when all bound genes (differentially expressed as well as not differentially expressed) were analyzed (Supplementary Figure 8a). We next performed a reverse approach, grouping genes based on differential expression (Supplementary Table 8), and then comparing binding fold changes at differentially expressed genes with those at transcribed genes (that is, genes with open chromatin) that were not differentially expressed (stably expressed genes). Consistent with results from the first approach, overall, genes upregulated in Thy4 cells showed increased binding in CD34− cells, and genes upregulated in Thy2 cells showed increased binding in CD34+ cells (Figures 7e and f; Supplementary Figures 8b and c). These findings demonstrate a positive association between BCL11B binding and target gene expression during human thymopoiesis. Furthermore, BCL11B KD resulted in downregulation of genes that showed increased binding with differentiation, suggesting a transcriptional activator function for BCL11B at T-lineage genes (Supplementary Figure 9).

Overall, BCL11B has numerous binding targets, many of which are differentiation stage specific. Thymopoiesis is characterized by a gain in BCL11B binding at multiple T-lineage genes with differentiation, and binding is associated with upregulation of these genes, many of which show BCL11B-dependent expression. These data further support a critical role for *BCL11B* in the induction of a T-lineage transcriptional program.

DISCUSSION

Through functional and bioinformatic studies, we here demonstrate *BCL11B* is essential for T-lineage commitment, binds to key T-lineage genes in a stage-specific manner, and has negative and positive regulatory modes of action during the initial stages of human T-cell differentiation, namely the repression of NK and myeloid potentials and the upregulation of genes critical for T-lineage differentiation.

The requirement of BCL11B for human T-lineage commitment is consistent with findings from murine studies.[13,44] However, *Bcl11b* knockout murine pre-commitment progenitors upregulated T-lineage genes normally, suggesting abnormally sustained expression of stem cell genes rather than failure of induction of T-lineage genes accounted for their differentiation arrest.[9] Furthermore, murine knockout cells showed abnormally upregulated *Il7r* expression.[13] In contrast, human KD cells showed downregulation of key T-lineage genes including *IL7R*, indicating important species related differences in regulatory effects of *BCL11B*. These findings add to previously reported species-specific regulatory aspects of thymopoiesis such as differences in the effects of NOTCH1 signaling on human and murine T-lineage commitment.[11,12] Although the function of BCL11B as a repressor of NK potential is conserved, differences were observed regarding the specific NK associated genes upregulated in human and murine BCL11B-deficient cells isolated from OP9-DLL1 co-cultures. *ZBTB16*, a gene important for human NK cell function, was upregulated in both species. However, human cells did not show upregulation of *TBX21*, *ID2*, *ZNF35* or *IL2RB*, genes whose markedly elevated expression is a hallmark of the NK like signature characteristic of murine cells even in conditions not very permissive of NK differentiation (that is, NOTCH signaling OP9-DLL1 co-cultures without IL-15).[12,45] Differences in experimental design (KD vs knockout) represent a caveat to the interpretation of these incongruities. Nevertheless, the exquisite sensitivity of these genes to BCL11B insufficiency in mice raises the possibility of species-specific regulatory effects of *NOTCH1* or *BCL11B* on these genes.

Knowledge of BCL11B binding targets has been limited to studies of a few gene promoters, which suggested both transcriptional repressor and activator roles for BCL11B.[42,46,47] Binding and expression analyses in primary thymocyte subsets and KD experiments suggested an activator role for BCL11B on the expression of T-lineage genes. Thus, our studies revealed a BCL11B transcriptional program previously undescribed in murine or human thymopoiesis, emphasizing the importance of investigating binding targets in fractionated populations. Further studies are needed to investigate the mechanisms underlying the negative regulatory effects of BCL11B on stem, NK and myeloid genes.

ETP-ALL, a T-ALL subtype reported to have poor outcomes, is characterized by a differentiation arrest at an initial stage of thymopoiesis, upregulation of stem and myeloid genes, and downregulation of T-lineage genes.[20] *BCL11B* is repressed in ETP-ALL.[7,20] *BCL11B* mutations (predicted to be hypofunctional) are seen across T-ALL subtypes with varying degrees of differentiation arrest.[18] Our data on the effects of BCL11B insufficiency on the initial stages of human thymopoiesis strongly support the notion that *BCL11B* repression at least partly contributes to the aberrant differentiation in ETP-ALL. Furthermore, the stage-specific binding of BCL11B to critical T-lineage genes during thymopoiesis and its role as an inducer of T-lineage genes provide a molecular basis for interpreting dysregulations in T-ALL with *BCL11B* aberrations. In our studies, BCL11B overexpression in T-ALL cells induced expression of T-lineage genes and inhibited proliferation but did not impact survival of mice transplanted with primary relapsed T-ALL cells. The use of relapsed T-ALL cells with multiple oncogenic aberrations (*TP53 and NOTCH1* mutations) that rapidly generated clinical leukemia in recipients, along with the outgrowth of GFP− cells in mice transplanted with overexpressing cells, might account for the lack of effect on survival.

Our studies delineate the regulatory actions of *BCL11B* critical for the establishment of the T-lineage program during the initial stages of human T-cell differentiation, providing a framework for understanding normal thymopoiesis and aberrations seen in T-ALL and immunodeficiencies.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank A George and Jonathan Kirzner for technical assistance; Dr Sara Twogood and Dr Wittenberg (Hollywood Presbyterian Hospital) for collecting CB; the Cardiothoracic Surgery Department at the Children's Hospital Los Angeles (CHLA) for collecting thymuses; the CHLA Next generation sequencing core; CHLA and the Broad Stem Cell Research Center flow cytometry facilities; the Center for AIDS Research Virology Core Lab (supported by the US National Institutes of Health (AI-28697), the UCLA AIDS Institute and the UCLA Council of Bioscience Resources) for CB and reagents; Dr D Kohn for the MND-PGK-GFP plasmid; Dr D Trono for psPAX2, pMD2.G and pLVTHM plasmids; Dr D Sabatini for the scramble shRNA plasmid; and Dr Zuniga-Pflucker for the OP9-DLL1 cell line. This work was supported by the St Baldrick's Foundation (CP), the US National Institutes of Health (K12 HD052954, K12 HD034610 and P30CA014089 to CP; P01 HL073104 to GMC; R01 CA172896 to YMK), the Nautica Foundation (CP), the Tower Cancer Research Foundation (CP), the Couples Against Leukemia Foundation (CP), the Joseph Drown Foundation (CP) and the UCLA Broad Stem Cell Research Center (GMC and DC).

AUTHOR CONTRIBUTIONS

VLH: collection and assembly of data (performance of experiments), data analysis and interpretation; AL: collection and assembly of data (performance of experiments), data analysis and interpretation; FL: bioinformatic analysis of RNA-Seq and ChIP-Seq data; DC: bioinformatic analysis of RNA-Seq data from normal bone marrow and thymic progenitors and assembly of annotated lncRNA database; JM: statistical analysis; YMK: generation of T-ALL xenograft; RB: conception and design, data analysis and interpretation; GMC: conception and design, data analysis and interpretation, manuscript writing; CP: conception and design, collection and assembly of data (performance of experiments), data analysis and interpretation, manuscript writing and final approval of manuscript.

REFERENCES

1 Plum J, De Smedt M, Leclercq G, Taghon T, Kerre T, Vandekerckhove B. Human intrathymic development: a selective approach. *Semin Immunopathol* 2008; **30**: 411–423.
2 Hao QL, George AA, Zhu J, Barsky L, Zielinska E, Wang X *et al.* Human intrathymic lineage commitment is marked by differential CD7 expression: identification of CD7− lympho-myeloid thymic progenitors. *Blood* 2008; **111**: 1318–1326.
3 Weerkamp F, Baert MR, Brugman MH, Dik WA, de Haas EF, Visser TP *et al.* Human thymus contains multipotent progenitors with T/B lymphoid, myeloid, and erythroid lineage potential. *Blood* 2006; **107**: 3131–3137.
4 Ferrando AA, Neuberg DS, Staunton J, Loh ML, Huard C, Raimondi SC *et al.* Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* 2002; **1**: 75–87.
5 Van Vlierberghe P, Ferrando A. The molecular basis of T cell acute lymphoblastic leukemia. *J Clin Invest* 2012; **122**: 3398–3406.
6 Durinck K, Van Loocke W, Van der Meulen J, Van de Walle I, Ongenaert M, Rondou P *et al.* Characterization of the genome-wide TLX1 binding profile in T-cell acute lymphoblastic leukemia. *Leukemia* 2015; **29**: 2317–2327.
7 Treanor LM, Zhou S, Janke L, Churchman ML, Ma Z, Lu T *et al.* Interleukin-7 receptor mutants initiate early T cell precursor leukemia in murine thymocyte progenitors with multipotent potential. *J Exp Med* 2014; **211**: 701–713.
8 Wallaert A, Durinck K, Van Loocke W, Van de Walle I, Matthijssens F, Volders PJ *et al.* Long noncoding RNA signatures define oncogenic subtypes in T-cell acute lymphoblastic leukemia. *Leukemia* 2016; **30**: 1927–1930.
9 Yui MA, Rothenberg EV. Developmental gene networks: a triathlon on the course to T cell identity. *Nat Rev Immunol* 2014; **14**: 529–545.
10 Parekh C, Crooks GM. Critical differences in hematopoiesis and lymphoid development between humans and mice. *J Clin Immunol* 2013; **33**: 711–715.
11 Van de Walle I, De Smet G, De Smedt M, Vandekerckhove B, Leclercq G, Plum J *et al.* An early decrease in Notch activation is required for human TCR-alphabeta lineage differentiation at the expense of TCR-gammadelta T cells. *Blood* 2009; **113**: 2988–2998.
12 Van de Walle I, Dolens AC, Durinck K, De Mulder K, Van Loocke W, Damle S *et al.* GATA3 induces human T-cell commitment by restraining Notch activity and repressing NK-cell fate. *Nat Commun* 2016; **7**: 11171.
13 Li L, Leid M, Rothenberg EV. An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. *Science* 2010; **329**: 89–93.
14 Califano D, Cho JJ, Uddin MN, Lorentsen KJ, Yang Q, Bhandoola A *et al.* Transcription factor Bcl11b controls identity and function of mature type 2 innate lymphoid cells. *Immunity* 2015; **43**: 354–368.
15 Albu DI, Feng D, Bhattacharya D, Jenkins NA, Copeland NG, Liu P *et al.* BCL11B is required for positive selection and survival of double-positive thymocytes. *J Exp Med* 2007; **204**: 3003–3015.
16 Punwani D, Zhang Y, Yu J, Cowan MJ, Rana S, Kwan A *et al.* Multisystem anomalies in severe combined immunodeficiency with mutant BCL11B. *N Engl J Med* 2016; **375**: 2165–2176.
17 De Keersmaecker K, Real PJ, Gatta GD, Palomero T, Sulis ML, Tosello V *et al.* The TLX1 oncogene drives aneuploidy in T cell transformation. *Nat Med* 2010; **16**: 1321–1327.
18 Gutierrez A, Kentsis A, Sanda T, Holmfeldt L, Chen SC, Zhang J *et al.* The BCL11B tumor suppressor is mutated across the major molecular subtypes of T-cell acute lymphoblastic leukemia. *Blood* 2011; **118**: 4169–4173.
19 Nagamachi A, Yamasaki N, Miyazaki K, Oda H, Miyazaki M, Honda Z *et al.* Haploinsufficiency and acquired loss of Bcl11b and H2AX induces blast crisis of chronic myelogenous leukemia in a transgenic mouse model. *Cancer Sci* 2009; **100**: 1219–1226.
20 Coustan-Smith E, Mullighan CG, Onciu M, Behm FG, Raimondi SC, Pei D *et al.* Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. *Lancet Oncol* 2009; **10**: 147–156.
21 Park E, Gang EJ, Hsieh YT, Schaefer P, Chae S, Klemm L *et al.* Targeting survivin overcomes drug resistance in acute lymphoblastic leukemia. *Blood* 2011; **118**: 2191–2199.
22 La Motte-Mohs RN, Herer E, Zuniga-Pflucker JC. Induction of T-cell development from human cord blood hematopoietic stem cells by Delta-like 1 *in vitro*. *Blood* 2005; **105**: 1431–1439.
23 Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; **25**: 1105–1111.
24 Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015; **31**: 166–169.
25 Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010; **11**: R106.
26 Subramanian ATP, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005; **102**: 15545–15550.
27 Casero D, Sandoval S, Seet CS, Scholes J, Zhu Y, Ha VL *et al.* Long non-coding RNA profiling of human lymphoid progenitor cells reveals transcriptional divergence of B cell and T cell lineages. *Nat Immunol* 2015; **16**: 1282–1291.
28 Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009; **10**: R25.
29 Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008; **9**: R137.
30 Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 2011; **27**: 1696–1697.
31 Ma W, Noble WS, Bailey TL. Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nat Protoc* 2014; **9**: 1428–1450.
32 Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 2012; **481**: 389–393.
33 McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010; **28**: 495–501.
34 Zhang JA, Mortazavi A, Williams BA, Wold BJ, Rothenberg EV. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell* 2012; **149**: 467–482.
35 Lanier LL, Allison JP, Phillips JH. Correlation of cell surface antigen expression on human thymocytes by multi-color flow cytometric analysis: implications for differentiation. *J Immunol* 1986; **137**: 2501–2507.
36 Awong G, Herer E, Surh CD, Dick JE, La Motte-Mohs RN, Zuniga-Pflucker JC. Characterization *in vitro* and engraftment potential *in vivo* of human progenitor T cells generated from hematopoietic stem cells. *Blood* 2009; **114**: 972–982.
37 Li W, Okuda A, Yamamoto H, Yamanishi K, Terada N, Yamanishi H *et al.* Regulation of development of CD56 bright CD11c+NK-like cells with helper function by IL-18. *PLoS One* 2013; **8**: e82586.
38 Burger R, Hansen-Hagge TE, Drexler HG, Gramatzki M. Heterogeneity of T-acute lymphoblastic leukemia (T-ALL) cell lines: suggestion for classification by immunophenotype and T-cell receptor studies. *Leuk Res* 1999; **23**: 19–27.
39 Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P *et al.* Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum Mutat* 2007; **28**: 622–629.
40 Eidson M, Wahlstrom J, Beaulieu AM, Zaidi B, Carsons SE, Crow PK *et al.* Altered development of NKT cells, gammadelta T cells, CD8 T cells and NK cells in a PLZF deficient patient. *PLoS One* 2011; **6**: e24441.
41 Avram D, Fields A, Senawong T, Topark-Ngarm A, Leid M. COUP-TF (chicken ovalbumin upstream promoter transcription factor)-interacting protein 1 (CTIP1) is a sequence-specific DNA binding protein. *Biochem J* 2002; **368**: 555–563.
42 Zhang LJ, Vogel WK, Liu X, Topark-Ngarm A, Arbogast BL, Maier CS *et al.* Coordinated regulation of transcription factor Bcl11b activity in thymocytes by the mitogen-activated protein kinase (MAPK) pathways and protein sumoylation. *J Biol Chem* 2012; **287**: 26971–26988.
43 Topark-Ngarm A, Golonzhka O, Peterson VJ, Barrett Jr B, Martinez B, Crofoot K *et al.* CTIP2 associates with the NuRD complex on the promoter of p57KIP2, a newly identified CTIP2 target gene. *J Biol Chem* 2006; **281**: 32272–32283.
44 Ikawa T, Hirose S, Masuda K, Kakugawa K, Satoh R, Shibano-Satoh A *et al.* An essential developmental checkpoint for production of the T cell lineage. *Science* 2010; **329**: 93–96.
45 Rothenberg EV, Zhang J, Li L. Multilayered specification of the T-cell lineage fate. *Immunol Rev* 2010; **238**: 150–168.

46 Cismasiu VB, Ghanta S, Duque J, Albu DI, Chen HM, Kasturi R *et al.* BCL11B participates in the activation of IL2 gene expression in CD4+ T lymphocytes. *Blood* 2006; **108**: 2695–2702.

47 Cismasiu VB, Adamo K, Gecewicz J, Duque J, Lin Q, Avram D. BCL11B functionally associates with the NuRD complex in T lymphocytes to repress targeted promoter. *Oncogene* 2005; **24**: 6753–6764.

Supplementary Information accompanies this paper on the Leukemia website (http://www.nature.com/leu)