Published in final edited form as:
Cancer Discov. 2021 November ; 11(11): 2846–2867. doi:10.1158/2159-8290.CD-21-0145.

Enhancer hijacking drives oncogenic *BCL11B* expression in lineage ambiguous stem cell leukemia

A full list of authors and affiliations appears at the end of the article.
[#] These authors contributed equally to this work.

Abstract

Lineage ambiguous leukemias are high-risk malignancies of poorly understood genetic basis. Here, we describe a distinct subgroup of acute leukemia with expression of myeloid, T lymphoid and stem cell markers driven by aberrant allele-specific deregulation of *BCL11B*, a master transcription factor responsible for thymic T-lineage commitment and specification. Mechanistically, this deregulation was driven by chromosomal rearrangements that juxtapose *BCL11B* to super-enhancers active in hematopoietic progenitors, or focal amplifications that generate a super-enhancer from a non-coding element distal to *BCL11B*. Chromatin conformation analyses demonstrate long range interactions of rearranged enhancers with the expressed *BCL11B* allele, and association of BCL11B with activated hematopoietic progenitor cell *cis*-regulatory elements, suggesting BCL11B is aberrantly co-opted into a gene regulatory network that drives transformation by maintaining a progenitor state. These data support a role for ectopic *BCL11B* expression in primitive hematopoietic cells mediated by enhancer hijacking as an oncogenic driver of human lineage ambiguous leukemia.

INTRODUCTION

Acute leukemias of ambiguous lineage (ALAL) remain a formidable diagnostic and therapeutic challenge (1,2). Such leukemias either show limited lineage differentiation or exhibit immunophenotypic features of multiple lineages, most commonly myeloid and either T- or B-lymphoid lineage and are often termed mixed phenotype acute leukemia (MPAL). Early T-cell precursor acute lymphoblastic leukemia (ETP-ALL), although often considered a subtype of T-ALL due to expression of cytoplasmic CD3, also exhibits lineage ambiguity with lack of expression of specific conventional T-ALL markers (e.g. CD1a, CD8 and CD5), along with aberrant expression of myeloid or stem cell markers (3,4). Thus, lineage ambiguous leukemias are classified by immunophenotypic features rather than genomic or

[*]Correspondence to: **Jeffery M. Klco, MD, PhD**, Jeffery.klco@stjude.org, Pathology, MS 342, Room D4047B, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105-3678, (901) 595-6807, **Prof. Claudia Haferlach, MD**, Claudia.haferlach@mll.com, MLL Münchner Leukämielabor GmbH, Max-Lebsche-Platz 31, 81377 München, +49 (0)89 99017-400, **Charles G. Mullighan, MBBS(Hons), MSc, MD**, charles.mullighan@stjude.org, Pathology, MS 342, Room D4047D, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105-3678, (901) 595-3387.

Author Contributions
L.E.M., S.B., L.G-P., K.D., I.I., S.A., R.H., P.E.M., C.M.M., T.W., M.B.V., M.R., K.G.R., J.J.Y., M.V.R., W.E.E., A.A.F., W.K., T.H., C.H., C.G.M. generated genomic and experimental data. S.L., M.R.L., J.M.R., M.L.d.B., S.M.K., S.P.H., M.L.L., C-H.P., W.Y., K.R.C, W.S., E.M.P., A.A.F., C.G.M. provided patient samples and data. Z.G., L.E.M., S.B., X.C., P.P., X.M., A.M., A.Z., B.X., Z.C., T-C.C., J.M., A.S., S.K., C.Q., W.Y., K.R.C., K.G.R., J.Y., M.V.R., W.E.E., A.A.F., J.Z., G.W., J.E.D., C.G.M. analyzed genomic data. V.W., J.Y., E.M.P., W.K., T.H., J.M.K., C.H. and C.G.M. analyzed clinicopathologic data. L.E.M., J.M.K., C.H. and C.G.M. designed the study, oversaw experiments, and wrote the manuscript.

biological features which results in a lack of clarity regarding appropriate therapy, which commonly fails. An improved understanding of the biological basis of these leukemias is thus desirable.

Several observations suggested that the ambiguous immunophenotype of such leukemias results from the interaction of leukemia-initiating genetic alterations and the hematopoietic cell in which these alterations arise (4,5). Genomic analyses of ETP-ALL showed that such leukemias commonly harbor mutations in genes regulating myeloid maturation, kinase signaling and chromatin modification which are often observed in myeloid leukemias; furthermore, the transcriptional profile of ETP-ALL is similar to a normal or leukemia myeloid stem cell which suggests a hematopoietic progenitor as the cell of origin (4,6). Similar analyses of MPAL showed distinct patterns of hematopoietic transcription factor alterations, such as *ETV6, WT1* and *RUNX1* in cases with T-lineage and myeloid features (T/myeloid MPAL) and rearrangement of *ZNF384* in B/myeloid MPAL (5). *ZNF384*-rearrangement is also observed in B-ALL cases that lack evidence of myeloid differentiation yet have an otherwise indistinguishable genomic and transcriptomic profile to B/myeloid MPAL (7,8), and such cases may shift phenotype between MPAL and typical ALL or AML during disease progression, suggesting a common etiology (5). Importantly, genotyping and xenotransplantation analyses of purified hematopoietic progenitors from MPAL samples demonstrated that lineage plasticity is independent of genetic variegation and inherent to all leukemic cells in a given sample (5). Thus, it has been proposed that subsets of ALAL, such as T/myeloid MPAL and ETP-ALL, should be considered a distinct entity (9), but the oncogenic drivers and biological relationship of the various types of lineage ambiguous acute leukemia to typical myeloid and lymphoid leukemias remain poorly understood. Moreover, genomic sequencing efforts of leukemia samples may fail to capture alterations that deregulate oncogene expression, such as those driven by non-coding alterations (10–12). Here, using integrated genomic analyses of a large acute leukemia cohort, we identify diverse, predominantly non-coding structural variants driving enhancer hijacking events that deregulate *BCL11B* in hematopoietic progenitor cells as the driver oncogenic events of a distinct subtype of lineage ambiguous leukemia.

RESULTS

To define gene expression-based subtypes of acute leukemia and the driver genomic alterations, we performed gene expression profiling and genetic alteration analysis on transcriptome sequencing (RNA-seq) data from 2573 adult and childhood cases of acute leukemia, including 774 T-lineage ALL, 262 acute myeloid leukemia (AML), 126 MPAL, and 1411 B-ALL cases representative of established subtypes (7) (Supplementary Figs. 1A,B; Supplementary Tables 1,2). As observed in prior analyses restricted to B-ALL (7), subtypes defined by leukemia-initiating somatic chromosomal alterations or sequence mutations exhibited distinct gene expression profiles. In addition, the subset of B/myeloid MPAL that harbored *ZNF384* rearrangements or recurrent alterations observed in typical B-ALL (e.g. high hyperdiploidy, *BCR-ABL1*, low hypodiploidy, or PAX5 P80R/R38C) clustered with respective B-ALL molecular subtypes, indicating that these cases are canonical B-ALL subtypes rather than MPAL cases of distinct genetic basis.

To further explore the genomic basis of lineage ambiguous leukemia, we repeated these analyses after exclusion of canonical B-ALL subtypes to enable better resolution of the transcriptional relationships among samples of T and/or myeloid lineage (Fig. 1A, Supplementary Figs. 1C,D, Supplementary Table 2). T-ALL, AML and T/myeloid MPAL cases predominantly clustered according to shared genomic alterations which allowed discrimination of known T-ALL and AML subtypes, including clear delineation of recently described subtypes such as T-ALL with rearrangement of *SPI1* (13), and identification of a distinct cluster of T-ALL cases with recurrent rearrangements of *LMO2*, primarily to *STAG2*, a member of the cohesin complex, as well as *SIK3* and *FOXJ3.* Notably, we observed poor clustering according to immunophenotype in lineage ambiguous cases (T/ myeloid MPAL, ETP or near-ETP ALL) (Fig. 1A, Supplementary Fig. 1C).

This analysis also identified a distinct cluster of 61 cases that included T/myeloid MPAL (N=23, 37.7%), ETP-ALL (N=21; 34.4%), AML (N=10; 16.4%), acute undifferentiated leukemia (AUL, N=2; 3.3%), and 5 cases described as T-ALL but lacking data to determine ETP-ALL/MPAL status (Fig. 1A, Supplementary Figs. 1C,D and Supplementary Table 3). The immunophenotype of these cases was cytoplasmic CD3+, CD2+, CD34+, CD117+, negative for CD1a, surface CD3, CD5 and CD8, and variable for myeloperoxidase (MPO), a critical marker distinguishing ETP-ALL from T/myeloid MPAL. T cell receptor gene rearrangements were absent in all cases (Supplementary Table 4, Supplementary Fig. 2), suggesting a hematopoietic progenitor cell of origin.

BCL11B rearrangement in lineage ambiguous leukemia

Initial analysis of RNA-seq data of cases in this cluster identified 6 (9.8%) cases with chimeric fusion genes involving the T cell transcription factor gene *BCL11B* (*RUNX1-BCL11B* and *ZEB2-BCL11B*). However, all cases with matched transcriptomic and whole genome sequencing (WGS) data (N=53) exhibited monoallelic expression of *BCL11B*, suggesting *BCL11B* deregulation as the unifying driver alteration (Fig. 1B, Supplementary Figs. 3A–C, Supplementary Table 5). We did not observe monoallelic *BCL11B* expression in canonical T-ALL (N=59 cases examined), with the exception of cases with *TLX1/3* deregulation, to which *BCL11B* is rearranged (Fig. 1B and Supplementary Table 5) (14,15).

Analysis of co-occurring genomic alterations showed that 49 of 61 (80%) *BCL11B* group cases harbored activating internal tandem duplication (ITD) or D835Y mutations in *FLT3,* greatly exceeding the frequency observed in other leukemia subtypes (Fig. 1C, Supplementary Figs. 4A–C, Supplementary Tables 2,3). Thus, of those T/myeloid MPAL samples with *FLT3* alterations, 19 of 20 (95%) belonged to the *BCL11B* group, in contrast to only 4 of 30 (13.3%) which lacked *FLT3* mutations (Supplementary Fig. 4D). Irrespective of the presence of *FLT3* mutations, all *BCL11B* group samples exhibited elevated *FLT3* expression levels as compared to non-*BCL11B* group T/myeloid MPAL, ETP-ALL, AML and T-ALL cases (Fig.1C and Supplementary Figs. 5,6). *BCL11B* expression was similarly high in this group as compared to non-*BCL11B* group MPAL, ETP-ALL and AML (Supplementary Fig. 5). Mutations in *WT1* (19 of 61, 31%), and alterations of *RUNX1* (16 of 61, 26%) were also common (Fig. 1C, Supplementary Tables 6–9, Supplementary Fig. 6).

To determine the genomic basis for *BCL11B* deregulation, we analyzed WGS data which identified structural variations (SVs) involving *BCL11B* in all cases with available data (N=53; Fig. 2A, Supplementary Table 3). All SV breakpoints, except for one corresponding to a *ZEB2*-*BCL11B* fusion, occurred either up- or downstream of the *BCL11B* coding sequence and involved 8 distinct partner loci. The most common was rearrangement of *BCL11B* to the gene desert upstream of *ARID1B* on chromosome 6 (N=23; 37.7%), while 8 (13.1%) cases harbored rearrangement near the ‘blood enhancer cluster’ (BENC) (16) located distal to *MYC* within the *CCDC26* gene on chromosome 8. Other rearrangements mapped to *CDK6* on chromosome 7 (N=3; 4.9%), *ETV6* on chromosome 12 (N=1; 1.6%), and *SATB1* on chromosome 3 (N=1; 1.6%). In addition, 13 (21.3%) cases harbored focal amplification of a 2.5 kb noncoding region 730 kb downstream of *BCL11B* on chromosome 14. These *BCL11B* SVs were otherwise not identified in WGS analysis of 5,550 pediatric and adult hematological malignancies, 344 pediatric brain tumors and 797 pediatric solid tumors (17–19) (Supplementary Table 10).

We verified the presence of this entity in an independent cohort of 91 adults with T-lineage leukemia (70 with RNA-seq), which included 36 ETP-ALL or T/myeloid MPAL cases (Supplementary Table 11); 14/70 (20%) cases were assigned to the *BCL11B* group, which comprised 13 of the 36 (36%) ETP-ALL and T/myeloid MPAL samples (Supplementary Fig. 7). Survival of adult patients with *BCL11B*-rearranged leukemia in this cohort was favorable compared to ETP and non ETP T-ALL, with median relapse free survival of 9.78 years and overall survival 9.9 years (Supplementary Table 12, Supplementary Figs. 8A,B).

BCL11B encodes a C2H2 zinc finger transcription factor of central importance in T cell development (20–23) and its expression coincides with the onset of T lineage commitment in both mouse and human (21,24,25). Continued *BCL11B* expression is required to promote T cell differentiation and repress alternate lineage fate choices, including myeloid and natural killer cell fates which are retained in thymic precursors lacking *BCL11B* (21,22). One mechanism by which BCL11B directs lineage fate is through recruitment of co-repressor complexes to silence genes involved in alternate fate lineages (26), for example by closing chromatin binding sites for the myeloid transcription factor PU.1 (27). Consistent with its role in T cell development, *BCL11B* is commonly targeted by diverse somatic alterations in T-ALL, including deletions and sequence mutations enriched in the DNA-binding zinc finger domains that are postulated to result in loss of function (28–30), as well as rearrangements, particularly a cryptic t(5;14)(q35;q32) translocation with *TLX3*, in which distal *BCL11B* enhancer elements drive overexpression of the *TLX3* oncogene (15,31,32). However, we did not observe these previously described *BCL11B* alterations in this subgroup.

Chromatin state of *BCL11B* rearrangement partner loci

Allele-specific *BCL11B* expression coupled with chromosomal rearrangements suggested that the genetic alterations of *BCL11B* observed in this subtype of leukemia are distinct from the loss-of-function or *TLX3*-deregulating alterations observed in typical T-ALL and may result in oncogenic deregulation of *BCL11B*. To investigate a potential oncogenic role of genomic loci rearranged to *BCL11B*, we examined the chromatin context of genomic breakpoints at the partner loci in multiple hematopoietic cell types. Whether in gene deserts (e.g. *ARID1B*) or genic regions (e.g. *CDK6*), rearrangement breakpoints occurred in proximity to cord blood (cb) CD34+ hematopoietic stem and progenitor cell (HSPC) super-enhancers as assessed by histone 3 lysine 27 acetylation (H3K27ac) ChIP-seq (33,34) (Fig. 2B), which are diminished or absent in more committed T cell progenitors and non-hematopoietic cell types (Fig. 2C, Supplementary Figs. 9A,B). Importantly, these chromosomal rearrangements, which included reciprocal translocations and more complex SVs (Supplementary Figs. 10A,B, 11A–D), always resulted in juxtaposition of an HSPC enhancer locus in *cis* with the upstream or downstream regions of *BCL11B*, often several hundred kilobases from the *BCL11B* promoter (as exemplified by rearrangements near *ARID1B*, Supplementary Fig. 12). Notably, the *CCDC26* rearrangements at chromosome 8q24.21 occurred near two well-characterized *MYC* enhancers, BENC (16) and the NOTCH1-driven *MYC* enhancer (N-Me) (35), that play roles in normal and malignant hematopoiesis (36). Because *BCL11B* is normally repressed in HSPCs (37), a putative cell of origin for these leukemias (5), these observations raise the notion that *BCL11B* SVs deregulate *BCL11B* expression through hijacking of active enhancers in a primitive hematopoietic cell.

Genomic organization of rearranged *BCL11B* loci

To directly demonstrate an enhancer hijacking mechanism for *BCL11B* de-regulation, we used HiChIP (38) to simultaneously profile H3K27ac and chromatin architecture in primary leukemia samples. We first performed H3K27ac HiChIP in normal human cbCD34+ HSPCs and 2 T-ALL cell lines (*TAL1*-deregulated Jurkat (39) and *BCL11B-TLX3* rearranged DND-41 (40)) to assess the stage-specific configuration of the *BCL11B* locus. Consistent with its normally silent state in HSPCs, *BCL11B* was devoid of H3K27ac in cbCD34+ cells (Fig. 3A). In contrast, in both DND-41 and Jurkat T-ALL cells, multiple long-range chromatin interactions were identified between the *BCL11B* promoter and distal enhancer elements located 1 Mb downstream of *BCL11B* (Fig. 3B,C). This enhancer cluster resides near a noncoding RNA termed ThymoD which is required for *BCL11B* activation during thymocyte development in mice (41,42) and is responsible for driving oncogenic *TLX3* expression in T-ALL cases harboring the t(5;14)(q35;q32) translocation. These data highlight the dynamic configuration of the *BCL11B* locus at distinct stages of hematopoietic development and in T lineage leukemias.

We next performed H3K27ac HiChIP in three primary leukemia samples representative of recurrent *BCL11B* SV rearrangements. We first examined an ETP-ALL case with an *ARID1B-BCL11B* rearrangement (Fig. 3D). Similar to normal cbCD34+ cells, the distal *ARID1B* enhancer was also active in the *BCL11B*-rearranged leukemia sample, and HiChIP demonstrated that this enhancer, as well as two additional H3K27ac regions farther upstream of *ARID1B*, formed long-range chromatin contacts with *BCL11B* following rearrangement.

To more accurately visualize the structural configuration of the rearranged *ARID1B-BCL11B* locus and identify true *cis* interactions, we re-mapped the HiChIP data using a patient-specific reference genome (see Methods) (Fig. 3D, bottom panel) and used MAPS (43) to call significant chromatin interactions. This revealed that *BCL11B* not only forms long-range contacts with multiple enhancers derived from the *ARID1B* locus, in particular with the more distal enhancer peak, but also demonstrated high levels of H3K27ac throughout the entire *BCL11B* gene body. This chromatin structure is notably different from the DND-41 and Jurkat T-ALL cell lines, where the *BCL11B* promoter primarily interacts with the distal ThymoD element (DND-41) and the *BCL11B* 3’ untranslated region (UTR) (Jurkat) (Figs. 3B,C, Supplementary Fig. 13). Moreover, sequential RNA-DNA fluorescent in situ hybridization (FISH) demonstrated the presence of *BCL11B* RNA colocalized with *ARID1B* enhancer DNA on one allele, whereas the second *BCL11B* allele only exhibited DNA FISH signal, providing orthogonal evidence of allele-specific *BCL11B* expression driven by the hijacked *ARID1B* enhancer (Supplementary Figs. 14A,B).

We observed similar evidence that HSPC super-enhancers drove *BCL11B* expression in two additional leukemia samples: an AUL case with a *CCDC26*/BENC-*BCL11B* rearrangement (Fig. 3E) and a T/myeloid MPAL case with a *CDK6-BCL11B* rearrangement (Fig. 3F). The *CCDC26*/BENC rearrangement results in translocation of the BENC super-enhancer to 380 kb upstream of *BCL11B*, within the *SETD3* gene body. Despite this distance, HiChIP demonstrated direct interactions between the rearranged super-enhancer and *BCL11B* (Fig. 3E), and sequential RNA-DNA FISH demonstrated colocalization of the BENC genomic sequence and *BCL11B* RNA signal only on the rearranged allele (Supplementary Fig. 14C). The *CDK6* SV resulted in rearrangement of two strong enhancer elements downstream of *BCL11B* which formed significant chromatin interactions with the *BCL11B* promoter (Fig. 3F). RNA and DNA FISH also demonstrated colocalization of *CDK6*-derived enhancer RNA and nascent *BCL11B* mRNA at the rearranged locus (Supplementary Fig. 14D). Notably, none of these *BCL11B*-rearranged leukemia cases showed evidence of ThymoD or N-Me activity, the two T cell enhancers clearly present in CD34+ progenitors isolated from human thymus (Figs. 3D–F, Supplementary Figs. 15A,B). The enhancer landscapes (H3K27ac) of all rearrangemed loci in primary leukemia samples, normal cbCD34+ HPSCs, normal thymocytes (27) and T-ALL cell lines are shown in Supplementary Figs.15A–H.

De novo super-enhancer generation through high copy tandem amplification of a noncoding region

One fifth of cases in this subgroup harbored a high-copy amplification of a 2.5 kb, evolutionarily conserved noncoding element 730 kb downstream of *BCL11B* on chromosome 14 (Fig. 4A, Supplementary Figs. 16A–D). Genomic quantitative PCR of 4 cases demonstrated 15–20 copies of this element (Supplementary Fig. 16E), suggesting that amplification above a threshold copy number is important for optimal function of this amplification, which we term BETA (*BCL11B* Enhancer Tandem Amplification). To resolve the structure and genomic location of BETA, we performed long-read DNA sequencing using the PacBio platform for two cases (SJMPAL011911, estimated to contain 14 copies for a final size of 35 kb, and SJTALL005006, estimated to contain 17 copies for a final size of 42.5 kb). The longest subreads from each case confirmed that the 2.5 kb element is present in a tandem array at the endogenous site of the original element (Supplementary Figs. 17A,B, Supplementary Table 13). Reads spanning the entire amplification, including non-amplified flanking DNA sequence, were obtained in SJMPAL011911, demonstrating that the amplification arose in *cis* with *BCL11B*, rather than as an extra-chromosomal DNA fragment or elsewhere in the genome.

In contrast to the chromosomal rearrangement partners of all other *BCL11B* group cases, the chromatin state of this element exhibited weak H3K27 acetylation in normal hematopoietic progenitor cells, including cbCD34+ and CD34+ thymocytes (Figs. 4A,B) (27,34), whereas more committed myeloid and lymphocyte populations surveyed by the Epigenome Roadmap Project (33) lacked active chromatin marks (Supplementary Figs. 18A,B). Thus, despite this weak signal, any regulatory activity encoded by this element likely acts specifically in hematopoietic progenitor cells and not more differentiated cell types. Analysis of transcription factor motifs present in this element identified recurrent ZNF384 and MEIS1/2/3 binding motifs (Supplementary Fig. 18B), and we did not find any evidence of somatically acquired mutations within BETA. We hypothesized that, following high-copy tandem amplification, this element would be transformed into a potent enhancer capable of activating *BCL11B*, similar to the other enhancers involved in *BCL11B* rearrangements. In support of this, all cases with the amplification exhibited evidence of enhancer-derived transcriptional activity on analysis of RNA-seq data, which was absent in samples lacking the amplification (Fig. 4B, Supplementary Fig.16A).

To more conclusively demonstrate that this amplification generates a strong enhancer, we performed H3K27ac HiChIP on the 2 samples analyzed with PacBio sequencing (SJMPAL011911, a T/myeloid MPAL sample and SJTALL005006, an ETP-ALL sample). In both cases, HiChIP demonstrated high levels of H3K27ac signal over the amplified region uniquely in these cases, which was absent in three *BCL11B* group samples lacking BETA, normal cbCD34+ HSPCs, and two T-ALL cell lines (Fig. 4C). BETA also formed multiple long-range chromatin loops with *BCL11B*, supporting its role as a direct regulator of *BCL11B* expression (Fig. 4D). Similar to the other *BCL11B* group cases, the *BCL11B* gene itself was demarcated by H3K27ac along the entire gene body, a pattern not observed in normal thymocyte precursors or the Jurkat or DND-41 T-ALL cell lines (Supplementary Fig. 15A). To determine whether this pattern reflected H3K27ac enrichment along the linear genome, or rather resulted from interactions with the H3K27ac-anchored amplification, we performed H3K27ac ChIP-seq in SJTALL005006 to assess the one-dimensional chromatin state of *BCL11B*. This confirmed that the HiChIP signal does indeed reflect a broad domain of H3K27ac over the entire 100 kb *BCL11B* gene body, independent of chromatin looping, making this one of the largest domains of contiguous H3K27ac in the genome of STJALL005006 (Supplementary Figs. 19A,B).

Examples of oncogenic enhancer amplifications have been reported previously (35,44,45), most notably duplications, however the high copy number of this tandem amplification led us to question the mechanism of its formation. Tandem amplifications have been reported to result from genomic instability resulting from short stretches of sequence homology that lead to DNA secondary structures during replication (46,47). In support of this as a plausible mechanism for BETA formation, two homologous “self-chain” sequences flank the 2.5kb element, and the left and right breakpoints of all 11 samples with WGS data available overlap these flanking homology blocks (Supplementary Fig. 16D, Supplementary Fig. 17A).

Interestingly, we noted that the ThymoD element—normally active in thymocyte progenitors and T-ALL synchronously with the N-Me *MYC* enhancer—was weakly active in the ETP-ALL case which otherwise showed an immature progenitor enhancer landscape as marked by the lack of activation of N-Me (Fig. 4D, Supplementary Fig. 15B). As N-Me and ThymoD are typically activated at the same developmental stage (namely, in CD34+ CD1a− thymocytes, Supplementary Figs. 15A,B), this suggested that BETA may also be able to activate an otherwise dormant T cell enhancer to contribute to *BCL11B* upregulation. Consistent with this, BETA formed significant chromatin interactions with the ThymoD element as well as the *BCL11B* gene (Fig. 4D).

Integrated analysis of gene expression and BCL11B chromatin occupancy in *BCL11B*-rearranged leukemia

Collectively, these data support that *BCL11B* expression is driven by HSPC-derived super-enhancers. Therefore, to investigate whether *BCL11B* group leukemias exhibit an HSPC-like gene expression state, we performed CIBERSORT deconvolution analysis on our cohort of leukemia transcriptomes using signatures derived from normal HSPC and mature populations (Figs. 5A,B). The *BCL11B* group samples showed strongest enrichment for HSPC subtypes, particularly hematopoietic stem cell (HSC) and multi-lymphoid progenitor (MLP) signatures, and this enrichment was significantly higher as compared to non-*BCL11B* group T/myeloid MPAL, ETP-ALL, AML, and T-ALL (Supplementary Fig. 20). We next performed combined single cell ATAC-seq/RNA-seq in two *BCL11B* group samples to functionally connect *BCL11B* expression with these open chromatin signatures in individual cells. Consistent with the bulk gene expression analysis, the scATAC-seq open chromatin profiles of both leukemia samples were significantly enriched for the long-term HSPC (LT-HSPC) and the activated HSPC (Act-HSPC) signatures (Fig. 5C, Supplementary Figs. 21A,B). Importantly, *BCL11B* expression correlated with this enrichment (Fig. 5D), consistent with a role for BCL11B in driving a progenitor cell gene expression program in this leukemia subtype.

To directly identify BCL11B-regulated genes, we performed BCL11B ChIP-seq in four primary *BCL11B* group samples (ETP-ALL cases SJTALL005006 and SJALL068666, T/myeloid MPAL case SJALL067671, and case SJALL067672 lacking MPO data) and the DND-41 T-ALL cell line, and compared these data to publicly available BCL11B binding profiles of normal CD34+ and CD34− thymocytes (21). BCL11B binding was widespread, with 8,437 genes showing evidence of promoter-proximal BCL11B binding in all four *BCL11B* group samples (Supplementary Fig. 22A). These genes, on average, exhibited higher expression in *BCL11B* samples than genes lacking promoter-proximal BCL11B binding (Supplementary Fig. 22B). Due to the high number of BCL11B-bound genes, we examined the subset with promoters bound by BCL11B specifically in the *BCL11B* group leukemias and not in DND-41 or normal proT cells (N=387, 4.58%). These genes were enriched for pathways related to TGF-beta receptor signaling (including *BMP4*, *SMAD3*, *SPP1*, *WNT1*, *SMAD5*, and *ENG*) and hematopoietic stem cell differentiation (including *LYL1*, *CIITA*, *SPI1*, *LMO2*, *NFATC2*, and *GATA2*) (Supplementary Figs. 22C,D). Additionally, 78/387 (20.2%) genes within this subset were upregulated in the *BCL11B* leukemia group compared to all other non-B-ALL leukemia samples. These genes were significantly enriched for biological processes related to lymphocyte proliferation, cytokine-mediated signaling, and regulation of kinase activity, including genes co-expressed with *FLT3* (e.g. *IL3RA*, Supplementary Figs. 22E,F)(48). Thousands of BCL11B binding sites were also identified at promoter-distal (>10kb from a transcription start site), genomic locations in the *BCL11B* group leukemia samples (range 5953–22150, Supplementary Fig. 22A). Motif enrichment analysis of these putative regulatory elements identified RUNX1 as the top-enriched motif in all samples, including T-ALL and normal proT cells, consistent with previously reported findings in mouse thymocytes (26,49) (Supplementary Fig. 22G). *RUNX1* expression was significantly higher in the *BCL11B* group compared to non-*BCL11B* group T/myeloid MPAL (p<0.001), ETP-ALL (p=0.0013) and AML samples (p=0.0043), but not T-ALL (p=NS) (Supplementary Fig. 5), suggesting possible collaboration between these factors in driving lineage-ambiguous leukemia.

We next assessed whether BCL11B occupancy corresponded to the enriched open chromatin HSPC signature identified from single cell analysis. For this analysis, three chromatin signatures had been previously identified from bulk ATAC-seq which span human HSPC populations and show distinct patterns of enrichment among single cells (LT-HSPC, Act-HSPC, and myeloid-erythroid progenitor (MEP)) (Fig. 5E, bottom) (50). We calculated the enrichment of BCL11B binding in each single cell using ChromVAR (51). Enrichment of BCL11B binding with the Act-HSPC open chromatin signature was consistently highest in the *BCL11B* group leukemia samples (Pearson's correlation = 0.64–0.78) compared to DND-41 cells (Pearson's correlation = 0.33), normal CD34+ thymocytes (Pearson's correlation = 0.44) or normal CD34− thymocytes (Pearson's correlation = 0.24) (Fig. 5E). As an example of the putative functional connection between BCL11B and the progenitor cell gene expression program characteristic of this leukemia subtype, increased BCL11B binding can be seen at the *GATA2* locus (Fig. 5F), including occupancy of noncoding regions up- and down-stream of the *GATA2* gene, which is absent in normal thymocytes and DND-41 cells.

Finally, we used single cell RNA-seq to investigate transcriptional heterogeneity within these leukemias. We analyzed 3 *BCL11B*-rearranged leukemia samples (one AML, one ETP-ALL, and one T/myeloid MPAL) and examined the expression patterns of *BCL11B* with the two main lineage-defining markers *MPO* (for myeloid lineage) and *CD3E* (for T lineage) (Supplementary Figs. 23A–C). *MPO* and *CD3E* were expressed in distinct subpopulations of the ETP-ALL and T/myeloid MPAL sample, consistent with bi-lineal potential of the leukemia. In contrast, *BCL11B* was expressed throughout the cell population, present in both *MPO*+ and *CD3E*+ cells, lending support that aberrant *BCL11B* expression drives the lineage ambiguity of these leukemias. Notably, *MPO* expression was highly variable between cells in individual cases, highlighting the limited utility of MPO to classify such cases.

BCL11B drives lineage aberrancy in vitro

These results support that SVs occurring in an uncommitted, extra-thymic hematopoietic progenitor cell result in de-regulated activation of *BCL11B* to initiate development of lineage-ambiguous leukemia. However, whether *BCL11B* expression is sufficient to drive immunophenotypic aberrancy in the absence of normal T cell differentiation inputs (e.g. Notch signaling in the thymic microenvironment (52,53)), is unknown. To address this, we used lentiviral transduction to express *BCL11B* or an empty vector control in cbCD34+ HSPCs isolated from 3 different donors and performed RNA-seq on GFP-sorted cells 96 hours after transduction in order to examine the acute effects of ectopic *BCL11B* expression on transcriptional regulation. Samples clustered predominantly by *BCL11B* overexpression status (Fig. 6A), with 669 genes up-regulated and 564 genes down-regulated compared to empty vector control (fold-change ≥ 2, FDR <0.05, Fig. 6B, Supplementary Table 14). Consistent with its known role as a driver of T lineage specification, upregulated genes were significantly associated with gene sets related to T cell differentiation, whereas downregulated genes were negatively correlated with gene sets related to myeloid differentiation (Fig. 6C, Supplementary Table 15). Exemplars of these trends include components of the T cell co-receptor and critical diagnostic marker CD3 (*CD3D*, *CD3E*, and *CD3G*), and the lymphoid regulator interleukin 7 receptor (*IL7R*), which were completely repressed in control CD34+ HSPCs but were significantly upregulated following *BCL11B* overexpression (Fig. 6D). Conversely, the expression of key myeloid differentiation markers, *MPO* and *LYZ*, as well as the myeloid transcription factor gene *SPI1* (encoding PU.1) were significantly downregulated, consistent with previously reported roles for BCL11B as a repressor of myeloid differentiation (21,22). These results demonstrate that BCL11B is sufficient to drive a T lineage expression program in hematopoietic progenitor cells in the absence of Notch signaling.

To examine the impact of these expression changes on cellular differentiation, and to co-model with the highly recurrent *FLT3*-ITD alteration observed in this subgroup (Fig. 1C), we repeated this experiment and included conditions with or without co-transduction of *FLT3*-ITD. Colony forming assays confirmed the myeloid differentiation block predicted from RNA-seq, where regardless of *FLT3*-ITD expression status, *BCL11B*-transduced cells yielded significantly fewer colonies compared to the empty vector control (Fig. 6E). Hematopoietic differentiation was also assessed in liquid culture media supplemented with either myeloid or lymphoid-promoting cytokines. We observed a distinct population of cells expressing cytoplasmic CD3 (cCD3) uniquely in *BCL11B*-transduced cells, which was most prominent in the lymphoid condition with co-expression of *FLT3*-ITD (Figs. 6F,G). The majority of cCD3+ cells were GFP+ (encoded by the lentiviral vector expressing *BCL11B*), consistent with BCL11B driving the immunophenotype. In contrast, cMPO expression was restricted to the GFP− population of *BCL11B*-transduced cells, which is most apparent in myeloid differentiation media (Fig. 6H,I), whereas in empty vector control cells, a prominent cMPO+ population could be observed in both GFP+ and GFP− cells. Taken together, these results provide evidence that deregulated expression of *BCL11B* in otherwise normal HSPCs can activate genes characteristic of T cell differentiation, block myeloid differentiation, and drive expansion of a subpopulation with a T-lineage immunophenotype (cCD3+).

We next examined the transformation potential of *BCL11B* and/or *FLT3*-ITD in mouse bone marrow HSPC colony forming assays. *BCL11B* overexpression was sufficient to promote replating of up to 4 serial passages, whereas *FLT3*-ITD-transduced cells failed to replate after a single passage (Fig. 6J). However, the combination of *BCL11B* and *FLT3*-ITD led to the strongest replating capacity, which was accompanied by increased cell numbers per colony in each passage (Fig. 6K). In summary, these in vitro results demonstrate that not only can ectopic *BCL11B* expression in HPSCs drive expression of T-lineage genes in the absence of normal Notch signaling, but BCL11B may also synergize with constitutive FLT3 activity to drive transformation of hematopoietic progenitor cells.

DISCUSSION

Lineage-ambiguous leukemias remain a significant diagnostic and therapeutic challenge due to lack of clarity surrounding their cellular origins and the molecular drivers of their characteristically ambiguous immunophenotype. Here, we identified multiple convergent genomic alterations deregulating *BCL11B* in hematopoietic progenitor cells as the unifying driver event of a subset of lineage ambiguous leukemias, which has major implications for our understanding of the mechanistic role of transcription factor deregulation in leukemogenesis, as well as the taxonomy of leukemia.

BCL11B encodes a T cell transcription factor whose expression is normally regulated by a complex series of molecular events specifically in cells entering T lineage specification (24,25). In uncommitted, extra-thymic HSPCs, the chromatin state of the *BCL11B* promoter is demarcated by high levels of repressive H3K27 trimethylation to maintain the gene in a silent state (37). Although loss-of-function and overexpression studies have elucidated critical roles for *BCL11B* in promoting and maintaining commitment to the T lineage, little is known about the functional consequences of aberrant *BCL11B* expression in a cell type which otherwise actively represses this gene.

Our results highlight multiple roles for *BCL11B* in the pathogenesis of acute leukemia with T-lineage features: as a tumor suppressor in typical T-lineage ALL, with loss of function deletions or mutations; by provision of enhancers that activate non-hematopoietic oncogenes such as *TLX3* in a subset of typical T-ALL; and here, cell stage-specific deregulation by existing or de novo super-enhancers that are maximally or uniquely active in HSPCs. In some respects, these findings recapitulate deregulation of other primitive (*TAL1/LMO2*) or non-hematopoietic (*TLX1/3*) transcription factor genes in T lymphoid leukemogenesis, but several features here are distinctive: hijacking of a regulator of the mature lymphoid series in hematopoietic stem and progenitor cells, and the distinctive and diverse range of normal and neo-enhancers that mediate this activation. These findings also resolve long-standing conceptual speculation in the cell of origin of lineage ambiguous leukemia. Two models of the cellular origins have been promulgated: transformation of a progenitor that retains myeloid and T lymphoid differentiation potential, or trans-/de-differentiation of a transformed T cell to achieve a more stem cell-like state (5,9,54). Here, integration of knowledge of *BCL11B* gene regulation, developmental stage-specific chromatin states, genomic analyses, and functional modeling support the former model: that a hematopoietic stem or early progenitor, rather than a committed T cell, is the cell of origin of *BCL11B*-deregulated leukemia.

Normally, *BCL11B* transcription is activated from both alleles within a few cell divisions of each other at the CD34+ stage of thymocyte development (21,24); thus, the presence of uniformly monoallelic expression of *BCL11B* in this subgroup indicates an aberrant, premature activation mechanism. The observation that all SVs resulted in juxtaposition of the *BCL11B* locus to regions in the genome harboring HSPC super-enhancers provides a plausible mechanism of ectopic *BCL11B* expression. Although these super-enhancers were also identified in CD34+ thymic precursors, two key T cell enhancer elements were highly active in thymic precursors but not in cbCD34+ HSPCs. These enhancers include the ThymoD element, which normally activates *BCL11B* at the earliest stages of T cell differentiation (41), and N-Me, which normally contributes to high *MYC* expression in developing thymocytes (35). Consistent with the cbCD34+ HSPC chromatin state, neither of these elements were active in the 5 cases of *BCL11B*-deregulated leukemias examined by H3K27ac HiChIP, apart from weak ThymoD activity in one BETA-containing case. Moreover, T cell receptor gene rearrangements were not detected in any *BCL11B* group leukemia sample, whereas over half of non-*BCL11B* ETP-ALL and T/myeloid MPAL leukemias harbored such rearrangements, supporting that *BCL11B* group leukemias originate prior to T lineage commitment.

Additionally, we demonstrated that the gene expression pattern of *BCL11B* group leukemias was more similar to normal HSPCs than T lineage cells. Using combined single-cell ATACseq/RNAseq in two primary *BCL11B* group samples we further demonstrated that *BCL11B* mRNA expression levels correlated with enrichment for the HSPC open chromatin signature. BCL11B ChIP-seq in four primary *BCL11B* group leukemias revealed widespread binding of BCL11B at thousands of promoters and putative cis-regulatory elements. The strong enrichment of RUNX1 motifs at these sites suggests collaboration between these factors in driving the progenitor cell expression program. Previous investigation of BCL11B occupancy in various murine T-lineage contexts has illuminated a complex logic to BCL11B gene regulatory function that is highly dependent on cell type, existing chromatin landscape, and availability of other chromatin binding complexes and factors (23,26,49). The fact that BCL11B binds promiscuously in *BCL11B* group leukemia samples and that this binding correlates strongly with the open chromatin signature of normal HSPCs suggests that ectopic BCL11B expression leads to aberrant co-option into a pre-existing stem/progenitor gene expression program reinforced by the myeloid lineage-repressing activities of BCL11B.

Finally, we modeled the initial stages of ectopic *BCL11B* overexpression in human cbCD34+ cells and mouse HSPCs. Acute overexpression of *BCL11B* was sufficient to induce T cell differentiation gene expression programs and repress myeloid programs, which translated to a myeloid differentiation block in vitro and, when combined with *FLT3*-ITD, emergence of a population of cells expressing cCD3. Notch signaling is indispensable for the normal initialization of T cell differentiation (52,53), and it has been suggested that the T-lineage immunophenotype characteristic of ETP-ALL and T/myeloid MPAL might reflect transformation of a thymic progenitor cell (9,55). However, our results demonstrate that HSPCs aberrantly expressing *BCL11B* and high levels of FLT3 activity can acquire a T-lineage immunophenotype in the absence of thymus-dependent Notch signaling. Moreover, *BCL11B* and *FLT3*-ITD overexpression in mouse lineage negative HSPCs conferred self-renewal properties, thereby demonstrating a key read-out of the transformation potential of putative oncogenes. These in vitro results collectively demonstrate oncogenic properties of ectopic *BCL11B* expression in HSPCs: aberrant activation of T-lineage genes coincident with a block in myeloid differentiation in human cells, and enhanced self-renewal properties in a standard mouse hematopoietic transformation assay.

While sharing features of lineage ambiguity and stemness, MPAL and ETP-ALL remain arbitrarily classified by cell surface immunophenotype and are often distinguished by a single, variable marker, myeloperoxidase, rather than leukemia-driving genomic aberrations. This confounds clinical management and the pursuit of more efficacious therapy based on a sound understanding of leukemogenesis. Here we demonstrated that one-third of ETP-ALL and T/myeloid MPAL, together with a small number of AML cases, comprise a distinct group unified by structural alterations deregulating *BCL11B*. Although numbers of cases with outcome data from the ECOG cohorts was small, *BCL11B*-rearranged cases had a superior outcome to non-*BCL11B*-rearranged ETP-ALL. Several prior case reports have described chromosomal rearrangements in myeloid, lymphoid or mixed phenotype leukemia consistent with rearrangement of *BCL11B* (31,56–63). An additional recent series also identified recurrent *BCL11B* rearrangements to *ARID1B*, *CCDC26*, *CDK6* and *ZEB2* in 20 cases of lineage ambiguous leukemia identified by DNA FISH analysis (64). Thus, while several *BCL11B*-rearrangements were identified by conventional diagnostic approaches, this study did not identify the BETA enhancer tandem amplification which occurs in over 20% of *BCL11B* group cases, or the less common rearrangements near *ETV6*, *SATB1*, and *RUNX1*. The ease with which BETA cases can be identified from RNA-seq might increase discovery of these cases in the future.

In summary, our findings show that *BCL11B* deregulation is a subtype-defining event and support the definition of a new entity of lineage-ambiguous leukemia, *BCL11B*-deregulated ALAL, that includes one third of ETP-ALL and T/myeloid MPAL cases as currently defined by the World Health Organization (65). Identification of this group transcends traditional methods of diagnosis to provide much needed clarity to the classification of lineage ambiguous leukemias and demonstrates that a subset of T/myeloid MPAL and ETP-ALL originate in a hematopoietic progenitor cell prior to the initiation of T lineage differentiation. Further studies are needed to examine the potential for inhibition of FLT3 signaling, in view of the frequent mutations in this gene in this subtype, and detailed analysis of the relative role of BCL11B in leukemic transformation according to cell of origin.

METHODS

Cohort compilation

Patients were included with diagnoses of ALL, AML, acute leukemia of ambiguous lineage and ETP-ALL. Patients were treated at St Jude Children's Research Hospital, the Children's Oncology Group, the Tokyo Children's Cancer Study Group (TCCSG), the Japan Association of Childhood Leukemia Study (JACLS) or underwent diagnostic testing by the Munich Leukemia Laboratory (MLL), with available tumor RNA sequencing data (Supplementary Tables 1,2). Patients and/or their guardians provided written informed consent in accordance with the Declaration of Helsinki. The study was approved by the Institutional Review Board of St. Jude Children's Research Hospital, and for cord blood studies, the Duke IRB. *BCL11B* structural variants were examined in all hematological malignancies subjected to whole genome sequencing at the MLL (Supplementary Table 10), and all childhood cancers studied by the Pediatric Cancer Genome Project. An independent cohort of adult T-lineage ALL cases with available immunophenotypic data enrolled on ECOG-ACRIN and CALGB protocols, including cases of T/myeloid MPAL and ETP-ALL immunophenotype, were also studied to determine the prevalence of the *BCL11B* immunophenotype, and the prevalence of the *BCL11B* transcriptomic signature in this cohort (Supplementary Table 11).

Analysis of patient outcome

Cox proportional hazards models were used to compare clinical outcome (relapse free survival (RFS) and overall survival (OS)) distribution of each group to CD1a+ T-ALL in the ECOG-ACRIN and CALGB cohorts.

Cell culture (cell lines and cord blood CD34+ cells)

The DND-41 and Jurkat cell lines were grown in RPMI 1640 supplemented with 10% fetal bovine serum and 50 U/ml penicillin, 50 μg/ml streptomycin and 2 mM L-glutamine. Cells were maintained at a density of 0.3×10^{6}-1.0×10^{6} cells per mL. DND-41 and Jurkat cells were validated by short tandem repeat (STR) analysis and were negative for *Mycoplasma spp.* Mononuclear white blood cells were separated from total umbilical cord blood using Ficoll Accu-Prep (Accurate Chemical AN5511) density gradient centrifugation. CD34+ cells were isolated using the CD34 MicroBead Kit UltraPure (Miltenyi Biotech 130–100-453) according to the manufacturer’s instructions. Cell purity was assessed by flow cytometry using APC-Cy7-CD34 (clone 581, BioLegend 343513). CD34+ cells were maintained in HSC expansion media (66) (StemSpan SFEM II, (StemCell Technologies 09655) supplemented with the following cytokines each at 100 ng/mL: IL6 (Peprotech 200–06), thrombopoietin (Peprotech 300–18), SCF (300–07), and FLT3 ligand (Peprotech 300–19), and 1 μM StemRegenin 1 (Cayman Chemical 10625)).

Gene expression analysis

For all analyses of the pan-leukemia cohort, transcriptome sequencing using Illumina library preparation and sequencers, data processing and gene expression quantitation, fusion gene detection, tSNE analysis and hierarchical clustering were performed as previously described (7). For analysis of *BCL11B* overexpression in cbCD34+ cells, RNA-seq was performed as follows. RNA was quantified using the Quant-iT RiboGreen RNA assay (ThermoFisher) and quality checked by the 2100 Bioanalyzer RNA 6000 Nano assay (Agilent) or 4200 TapeStation High Sensitivity RNA ScreenTape assay (Agilent) prior to library generation. Libraries were prepared from total RNA with the TruSeq Stranded Total RNA Library Prep Kit according to the manufacturer’s instructions (Illumina 20020599). Libraries were analyzed for insert size distribution using the 2100 BioAnalyzer High Sensitivity kit (Agilent), 4200 TapeStation D1000 ScreenTape assay (Agilent), or 5300 Fragment Analyzer NGS fragment kit (Agilent). Libraries were quantified using the Quant-iT PicoGreen ds DNA assay (ThermoFisher) or by low pass sequencing with a MiSeq nano kit (Illumina). Paired end 100 cycle sequencing was performed on a NovaSeq 6000 (Illumina). Total stranded RNA sequencing data were processed by the internal AutoMapper pipeline. Briefly, the raw reads were first trimmed using Trim-Galore v0.60 (67), mapped to the human genome assembly (GRCh38) using STAR v2.7 (68) and then the gene level values were quantified using RSEM v1.31 (69) based on GENCODE v31 annotation. To identify differentially expressed genes, only confidently annotated protein coding transcripts were considered, and low count genes were removed from analysis using a counts per million (CPM) cutoff of 10. Normalization factors were generated using the TMM (70) method. Counts were normalized using voom (71). Voom normalized counts were analyzed using the lmFit and eBayes functions of the limma (72) software package. Genes with an absolute fold change of ≥2.0 and FDR <0.05 were considered differentially expressed.

Gene set enrichment analysis

GSEA (73) was performed using the mSigDB C2 gene sets and an in-house curated list of gene sets.

Whole genome sequencing

For somatic structural variants, five SV callers were implemented in the workflow for SV calling, including Delly (74), Lumpy (75), Manta (76), GRIDSS (77) and novoBreak (78). The SV calls passing the default quality filters of each caller were merged using SURVIVOR (79) and genotyped by SVtyper (80). The intersected call sets were manually reviewed for the supporting soft-clipped and discordant read counts at both ends of a putative SV site using the Integrated Genome Viewer (IGV). Focal inspection of *BCL11B*-involved rearrangements were performed to identify the exact breakpoints with BLAT (81).

Mutation and variant detection

The genetic alterations including sequencing mutations and copy number alterations were collected from previous publications or called specifically for this project using established locally developed pipelines (82,83). All variant calls were further manually reviewed for read depth and to remove artifacts. Published SNVs and indels, SVs and copy number variation results were downloaded for the TARGET-AML (84–86), MPAL (5), and Japan-TALL (13) cohorts. Somatic SNVs and indels were identified from Illumina-based WGS and WES data with Bambino version 1.6 (87). For CGI-based WGS data, SNVs and indels were used from our previous TARGET variant calls (84). For CPML and additional MPAL WES/WGS samples, we used our in-house ensemble approach as described (83) to call SNVs and indel with multiple published tools, including Mutect2 version 4.0.2.1 (88), SomaticSniper version 1.0 (89), VarScan2 version 2.4.3 (90), MuSE version 1.0 (91) and Strelka2 version 2.9.10 (92). The consensus calls by at least two callers were considered as confident mutations. Strelka2 version 2.4.7 (92) and Pindel (93) were used for SNV and indel detection for German-MLL cohorts as previously described (94). Somatic SVs were identified from Illumina-based WGS data using CREST version 1.0 (95). SVs for CGI-based WGS data were obtained from our previous study which included filtering to remove germline SVs (84). Manta version 0.28.0 (76) was used for SV detection and GATK4 version 4.0.2.1 for copy number analysis for the MLL cohort as previously described (94). Preprocessing and copy number analysis of Illumina SNP array data were performed as described (96). Affymetrix SNP array data was analyzed using Rawcopy (97). CONSERTING (98) was used for somatic copy number alternation detection from WGS data.

PacBio sequencing and analysis

High molecular weight DNA was purified from cryopreserved primary leukemia cells using the Gentra DNA Extraction and Purification Kit (Qiagen). Shearing, library prep, and size selection were followed based on the manufacturers protocol (SMRTbell Express Template Prep Kit 2.0). After library preparation, the samples were processed through primer annealing and polymerase binding. These steps were performed using sequencing primer v4 and Sequel II Binding Kit 2.0 following the manufacturer's protocol. The samples were evaluated for fragment sizes using the Fragment Analyzer gDNA kit, samples were sheared using a Megaruptor 3.0, and size selection was performed using the PippinHT, targeting fragments between 18–21 kb. Samples were sequenced on a PacBio Sequel II instrument in CCS mode using 2 SMRTcells per sample, with 70 pM loading concentration, 2 hours of pre-extension time, and a 30-hour movie time. The resulting data from the sequencing indicates that the samples had a mean length of 18–21 kb. The PacBio reads were aligned against the human genome (hg38) using pbmm2 of PacBio SMRTtools v.8.0. Structural variants were identified (pbsv, SMRTtools). Reads aligning to the region of BETA (hg38 chr14:98541905–98544383) were extracted and *de novo* assembled using Canu (99). The assembled contigs were compared with the reference BETA sequence and its flanking sequences using zPicture (100). We also generated circular consensus reads (ccs, SMRTtools) to estimate the copy number of BETA (Supplementary Table 13).

Allele-specific expression

Cis-X (11) (version 1.4.0) was used to discover regulatory noncoding variants that lead to allelic specific expression (ASE) of the corresponding target gene. Recommended parameters (cis-X run … -w 10 -r 10 -f 5) were used to detect ASE genes. To account for intronic SNPs that typically have lower depth in RNA-seq, we lowered the coverage threshold for RNA-seq to 5 to include more markers for the ASE genes. Additionally, ASE runs were called if a minimum of 4 sequential markers showed significant ASE or mono-allelic expression. *BCL11B* was considered have ASE when overlapping with the ASE runs identified.

T cell receptor rearrangement analysis

The fastq files of whole genome sequencing data were aligned to reference V, D, J and C genes of T cell receptors and assembled clonotypes using MiXCR software (101). The TCR reference used in the present study was included in the MiXCR software as default setting; *TRA/TRD*, NG_001332.2; *TRB*, NG_001333.2; *TRG*, NG_001336.2. Called TCR rearrangements were filtered by excluding (i) cloneCount < 5; and (ii) cloneFraction < 0.05.

HiChIP library construction

All H3K27ac HiChIP experiments were performed using the Arima-HiC$^+$ kit (Arima Genomics A101020) according precisely to the manufacturer's protocols (Arima-HiC$^+$ document numbers A160168 v00 (HiChIP) and A160169 v00 (library preparation)), with minor modifications during the cell preparation step for cell lines and patient samples as follows: for cell lines and CD34+ cells, freshly cultured cells were counted using trypan blue and a cell suspension volume corresponding to 10 million live cells was washed once with

PBS (-Ca^{2+} -Mg^{2+}) and then crosslinked in a volume of 5 mL PBS with 2% formaldehyde for 10 minutes at room temperature, followed by quenching with Stop Solution 1 according to the Arima protocol. For cryopreserved patient samples, one vial of cells was thawed in a 37°C water bath and the entire contents of the vial were immediately transferred to 5 mL PBS where cells were counted before proceeding with crosslinking as above. The total number of cells used per patient sample ranged from 1.6 million to 8 million cells. The H3K27ac antibody was from Active Motif (am91194). Uniquely barcoded HiChIP libraries were pooled and sequenced to an average depth of 370 million reads on an Illumina NovaSeq instrument.

HiChIP data analysis

HiChIP data were processed and analyzed using MAPS (43) v1.9 as implemented within the Arima Genomics bioinformatics pipeline available at https://github.com/ijuric/MAPS. The following program versions were used: BWA (102) v0.7.12, SAMtools (103) v1.10, and deepTools (104) v3.4.0. For visualization of 2D heatmaps, .hic files were generated with Juicer tools (105) and uploaded to the St. Jude Protein Paint cloud-based server (106). To call significant HiChIP loops, MAPS requires a corresponding ChIP-seq peak set. Because we did not have sufficient patient material to generate ChIP-seq in every sample, we used H3K27ac ChIP-seq data from human CD34+ cells (34) and manually added peak regions corresponding to the *BCL11B* promoter and the region demarcated by BETA breakpoints, which are not marked by H3K27ac in normal CD34+ cells. To overcome the limitation of inter-chromosomal read-pair mapping for read-pairs spanning chromosome 14 rearrangement junctions, we generated patient-specific reference genomes to allow us to call loops and visualize the actual configuration of the *BCL11B* rearrangement in these cases. To generate patient-specific genome files for samples SJMPAL011914, SJAUL068292 and SJALL068279, the breakpoint positions were used to first generate fasta chromosome files corresponding to the reciprocal rearrangement involving chromosome 14 and the partner chromosome. The wild-type chromosome 14 and partner chromosome were then removed from the genome fasta file to prevent HiChIP read pairs mapping to more than 2 locations. Mappability was computed on the patient-specific genome file using GEM (107,108) v2.0.14 and converted to bigwig format; mappability scores, GC percentage and effective fragment lengths for non-overlapping 5 kb bins were then computed using the genomic features generator scripts within the Arima bioinformatics pipeline to generate patient-specific genomic features files. These files were then used to run the standard MAPS pipeline.

ChIP-seq

ChIP-seq was performed using the SimpleChIP kit (Cell Signaling Technologies 56383) according to the manufacturer’s instructions. Briefly, freshly harvested DND-41 or thawed cryopreserved cells from primary patient samples were washed once in PBS and resuspended in PBS at a concentration of 1 million cells/ml. 37% formaldehyde was added to a final concentration of 1% and cells were crosslinked for 10 minutes at room temperature. Crosslinking was quenched with the addition of glycine at a final concentration of 0.1M and incubated at room temperature for 5 minutes. Cells were spun at 4°C, washed once more with ice-cold PBS and flash frozen in aliquots of 4 million cells. For each immunoprecipitation, one aliquot of 4 million crosslinked cells was used. Cells were sheared in Covaris microTUBEs (Covaris 520045) at the equivalent of 2 million cells per tube using a Covaris E220 instrument with the following settings: peak power = 140 watts, duty factor = 5, cycles/burst = 200, time = 9 minutes, temperature = 4°C. The following antibodies were used: 1 μg of H3K27ac (Active Motif am91194) or a cocktail of 4 antibodies for BCL11B, 0.5 μg each (Cell Signaling Technologies 12120, Abcam ab18465, Bethyl Laboratories A300–383A and A300–385A). Immunoprecipitated DNA was prepared for Illumina sequencing using the Kapa Hyper Prep Kit (Kapa KR0961) and 11 (H3K27ac) or 19 (BCL11B) PCR cycles. Libraries were sequenced to an average depth of 387 million reads.

ChIP-seq data analysis

ChIP-seq reads were quality trimmed using Trim_Galore v0.4.4 with the cutadapt program (109) and FastQC was used to filter reads with quality scores >20 for downstream analysis. Reads were mapped to the hg38 reference genome using BWA (102) v0.7.12 and duplicates were marked using biobambam2 (110) v2.0.87. Coverage tracks were generated using deepTools (104) v3.4.0 with RPKM used for normalization and a bin size of 10. ChIP-seq peaks were called using MACS2 (111) v2.1.1 with a minimum false discovery rate (FDR) cutoff of 0.01 using default parameters for both H3K27ac and BCL11B with input DNA used as the control.

Publicly available H3K27ac ChIP-seq data were downloaded from the Gene Expression Omnibus using the following accessions: cbCD34+ H3K27ac accession GSE107147 (34), CD34+ and CD34− thymocytes accession GSE84677 (21). Reads were mapped and processed as above.

Super-enhancer analysis

CD34+ super enhancers were identified using the Ranking of Super Enhancers (ROSE) program (112) with default parameters on H3K27ac ChIP-seq data from cord blood CD34+ cells (34).

Motif enrichment analysis

A 50 bp window centered at the summit of each promoter-distal (10 kb from RefSeq TSSs) ChIP-seq peak was used as input to the findMotifsGenome.pl program (HOMER (113) v4.10) using -size given with default background parameters. JASPAR (114) was used to identify motif instances in BETA. For this, the sequence corresponding to hg38 chr14:98541905–98544383 was used as input and all human CORE motifs were searched with a relative profile score threshold cutoff of 90%.

Genomic PCR validation of *BCL11B* rearrangements

Genomic DNA was extracted from cryopreserved bone marrow aspirates taken at time of diagnosis using the phenol-chloroform organic extraction method with the exception of sample SJMPAL068275 which, due to very low sample amount, was subjected to whole genome amplification using the Qiagen REPLI-g kit (Qiagen 150345). 50 ng were used for each PCR reaction (or 1 μl of a 1:100 dilution for sample SJMPAL068275) with Phusion High-Fidelity DNA Polymerase (New England Biolabs M0530L) with the following primer pairs designed to amplify patient-specific breakpoints on the allele containing the *BCL11B* gene: SJMPAL042793 (For: ACGGTTGATTTCACTGCGAC Rev: CTGTGCCATAACATGCGGAA), SJMPAL068275 (For: GCAGCTATCAAATCCAGAGGC Rev: TAGTACCGGCTGTTGGAGAG), SJMPAL011914 (For: CATTGTCCCTCCAAACCTGC Rev: TTCACCGATGGAAACCTGGA). Cycling conditions were 98°C for 30s followed by 32 cycles of 98°C for 10s, 63°C for 20s, 72°C for 45s, and a final extension of 2 minutes at 72°C. PCR reactions were cleaned with the Qiagen PCR Purification kit and submitted for Sanger sequencing using the primers used in each PCR reaction.

BETA copy number analysis

BETA DNA copy number was quantified using the TaqMan Copy Number Reference Assay with a custom FAM-labeled TaqMan probe targeted to a sequence within a region of the BETA amplification shared by all samples (For: TGAACCGAGCAGAAGTGACAA; Rev: CTGTTAACCTCTCTATTCCTCTTTGTGTT; reporter: 5’- ACCAATCCATCACCCCAGAGCC). A VIC-labeled probe targeting the *RNAseP* gene was used as the internal reference (Thermo Fisher 4403326). Reactions were prepared using the TaqPath ProAmp Master Mix (Thermo Fisher A30866) and performed on samples with available genomic DNA (SJMPAL011911, SJMPAL011912, SJTALL005006, and 3 immunophenotypically defined subpopulations of SJMPAL040459). Genomic DNA from healthy normal cord blood CD34+ cells and a non-BETA sample (SJMPAL011914) were used as additional copy number controls.

DNA and RNA FISH

Cryopreserved leukemia cells were thawed, washed in PBS and applied to slides by cytocentrifugation followed by fixation in 4% PFA containing 0.5% Tween 20 and 0.5% NP−40 for 10 minutes. Following fixation, slides were stored in 70% ethanol at −20°C until ready to proceed with hybridizations. Fixed slides were prepared for RNA hybridization by dehydration in 80% and 100% ethanol for 2 minutes each. 10 µl of denatured probes for both an enhancer RNA (either BENC, *CDK6* or the *ARID1B* enhancer) and nascent RNA for *BCL11B* were combined and applied to dehydrated slides and allowed to hybridize at 37°C overnight. RNA hybridized slides were then washed in 50% formamide and 2X SSC at 37°C for 5 minutes and mounted in Vectashield mounting medium containing DAPI. 3D images were then acquired, coordinates recorded, and slides were then treated with RNaseA in order to erase the RNA signals. Slides were then treated in 0.2N HCl followed by denaturation in 70% formamide and 2X SSC at 80°C for 10 minutes. Denatured slides were then hybridized with an appropriate set of bacterial artificial chromosomal (BAC) DNA probes for detection of DNA for either BENC, *CDK6*, or the *ARID1B* enhancer combined with *BCL11B* and the same microscopic fields as were imaged after RNA FISH were imaged again after the addition of DNA FISH. The two sets of images were then compared to allow for visualization of the locations of enhancer RNA and nascent *BCL11B* RNA and DNA from these same targets. This allows for direct visualization in *cis* for allele-specific activation of *BCL11B* while in contact with specific enhancers. For detection of *BCL11B* expression that occurs as a result of acquisition of a de novo enhancer 730 kb downstream of *BCL11B*, a similar experimental approach was used. First, an RNA FISH experiment designed to detect nascent *BCL11B* expression was performed and imaged followed by DNA FISH for a 2.5kb segment of DNA that is specifically amplified in *cis* when *BCL11B* is expressed (BETA) combined with DNA FISH for *BCL11B*. The same fields as were imaged after RNA FISH were imaged again after DNA FISH and the images were compared. This experiment showed that *BCL11B* expression occurs only on the allele that contains BETA. BAC probes used were: *BCL11B* (RP11–844K17, chr14:99634644–99795981 and RP11–876E22 chr14:99552293–99744370); *BCL11B* flanking (RP11–151N10 chr14:98799667–98991003 and RP11–45Oc22 chr14:99900658–100090060); BETA (WI2–1744B12 chr14:98984071–99024847); *CDK6* (RP11–1102K14 chr7:92302711–92463709 and RP11–809H24 chr7:92117011–92306530); *ARID1B* enhancer (RP11–808B8 chr6:156458627–156620361 and RP11–263N6 chr6:156057468–156215697); BENC (RP11–17E16 chr8:130493286–130632903). All coordinates referenced for FISH are hg19.

Single cell RNA sequencing

Frozen mononuclear cells from diagnosis bone marrow samples from patients SJMPAL068275 (T/myeloid MPAL, *CCDC26*/BENC rearranged), SJAML068287 (AML, *CDK6* rearranged), and SJTALL005006 (ETP-ALL, BETA) were thawed and stained with a cocktail of human-specific antibodies including: BV605 Mouse Anti-CD45 (BD Biosciences 564048), PE mouse anti-CD3 (BD Pharmingen 555340), BUV805 mouse anti-CD13 (BD Biosciences 749264), BV421 mouse anti-CD2 (BD Biosciences 744873), APC mouse anti-CD34 (BD Biosciences 340441) and PerCP/Cyanine5.5 anti- CD38 (Biolegend 356614). For all samples, the following populations were sorted on a BD FACS Aria (BD Biosciences): leukemia blast cells (CD45dim/CD2+/CD3−) and stem/progenitor cells (CD34+CD45dim/neg out all blast negative cells). Sorted cells were washed three times with 1X PBS (calcium and magnesium free) containing 0.04% weight/volume BSA (Thermo Fisher Scientific AM2616) and counted. From each population 5,000 cells were combined and loaded on Chromium Next GEM Chip G (10X Genomics PN-2000177) with Chromium Next GEM Single Cell 3′ GEM Kit v3.1 (10X Genomics PN-1000130) and Chromium Next GEM Single Cell 3′ v3.1 Gel Beads (PN-2000164) according to standard manufacturer’s protocols. Briefly, oil partitions of single-cell and oligo coated gel beads in emulsions (GEMs) were captured and reverse transcription was performed, resulting in cDNA tagged with a cell barcode and unique molecular index (UMI). Next, GEMs were broken and pooled fractions were recovered. Silane magnetic beads were used to purify the first-strand cDNA from the post GEM-RT reaction mixture. Barcoded, full-length cDNA was amplified via PCR and quantified using an Agilent Bioanalyzer High Sensitivity chip (Agilent Technologies 5067–5585 and 5067–5593). Enzymatic fragmentation and size selection were used to optimize the cDNA amplicon size. To construct the final libraries, P5, P7, a sample index, and TruSeq Read 2 (5’GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3’) were added via end repair, A-tailing, adaptor ligation, and PCR (Chromium Next GEM Single Cell 3′ Library Kit v3.1, PN-1000158, 10X Genomics). Final library quality was assessed using an Agilent Bioanalyzer High Sensitivity chip (Agilent Technologies 5067–5584 and 5067–5585). Samples were then sequenced on the Illumina NovaSeq with 28 (barcode+UMI) + 91(read) setting, with a median depth of >50000 reads per cell for majority of the samples.

Single cell data were aligned and quantified using the Cell Ranger (v4.0.0) pipeline (http://www.10xgenomics.com) against the human genome GRCh38 (refdata-gex-GRCh38–2020-A). Cells with fewer than 200 or higher than 6000 features, or mitochondria content higher than 75% were removed. Clusters of cells were identified using Seurat (3.1.0, https://satijalab.org/seurat) using UMAP reduction and characterized based on gene expression of major haemopoietic cell types.

Multimodal single cell ATAC-seq/RNA-seq

Frozen mononuclear cells from diagnosis bone marrow samples from patients SJMPAL011913 (T/myeloid MPAL, *ARID1B-BCL11B* rearranged), and SJTALL005006 (ETP-ALL, BETA) were thawed and enriched for live cells using the Dead Cell Removal Kit (Miltenyi Biotec 130–090-101) according to the manufacturer’s instructions. Live cells were washed twice with cold PBS + 0.04% BSA (Miltenyi Biotec 130–091-376), counted using a Countess II FL Automated Cell Counter (ThermoFisher Scientific A27974) and 0.7 – 0.8 million cells were used for nuclei isolation, according to the Nuclei Isolation for Single Cell Multiome ATAC + Gene Expression Sequencing User Guide (version CG000365 Rev B). Briefly, cells were spun at 300 rcf for 5 minutes at 4°C and resuspended in 100 µL Lysis Buffer (preparing according to 10X Genomics’ instructions and containing 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 0.1% Nonidet P40 Substitute, 0.01% Digitonin, 1% BSA, 1 mM DTT, 1 U/µl Roche RNase inhibitor and water) and incubated on ice for 4 minutes, then 1 mL chilled Wash Buffer (preparing according to 10X Genomics’ instructions) was added before spinning. Cells were washed three times in Wash Buffer and resuspended in chilled Diluted Nuclei Buffer (10X Genomics). To confirm complete lysis and nuclei concentration an aliquot was inspected by Trypan Blue. GEM generation and single cell libraries were prepared according to the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression User Guide (CG000338 Rev B). Briefly, following transposition GEMs were generated by combining barcoded Gel Beads, transposed nuclei, a Master Mix that includes reverse transcription (RT) reagents, and Partitioning Oil on a Chromium Next GEM Chip J (10X Genomics PN-2000264). Incubation of the GEMs in a thermal cycler for 45 minutes at 37°C and for 30 minutes at 25°C generates 10X Barcoded DNA from the transposed DNA (for ATAC) and 10X Barcoded, full-length cDNA from poly-adenylated mRNA (for GEX). This was followed by a quenching step that stopped the reaction. Next, GEMs were broken and pooled fractions were recovered. Silane magnetic beads were used to purify the first-strand cDNA from the post GEM-RT reaction mixture. Barcoded transposed DNA and barcoded full-length cDNA from poly-adenylated mRNA were pre-amplified by PCR and the products were used as input for both ATAC library construction and cDNA amplification for gene expression library construction. Libraries were sequenced on the Illumina NovaSeq according to 10X Genomics setting, with a median depth of >50000 reads per nucleus for majority of the samples.

Comparison of leukemia transcriptional signatures with normal hematopoietic populations

Purified hematopoietic populations from healthy umbilical cord blood were sorted and subjected to RNA-seq as previously described in Xie *et al.* (115). Signature matrix generation was performed through CIBERSORTx (116) using a qvalue cutoff of 0.05. Deconvolution analysis on 2467 pan-leukemia samples was performed through CIBERSORTx with default settings, returning enrichment scores of each hematopoietic score for each patient. Signature enrichment of each purified population was normalized to have a sum of 1 in each sample. Using these normalized scores, a neighborhood graph was generated for the leukemia samples using n = 15 nearest neighbors and UMAP dimensionality reduction was performed using a minimum effective distance of 0.1 and a spread of 1.5. Signature scores were subsequently scaled for visualization.

Single nuclei data from the joint scATAC/RNA experiments were aligned and quantified using cellranger-arc (v1.0.1) pipeline (http://www.10xgenomics.com) against the human genome GRCh38 (refdata-gex-GRCh38–2020-A). Signac was used to generate an LSI representation of both scATAC-Seq and scRNA-Seq data and a joint UMAP representation based using both modalities. ChromVAR was used to calculate the enrichment of hematopoietic signatures from Takayma *et al.* (50) within each single cell with the ATAC-Seq data and correlation between hematopoietic signature enrichment and *BCL11B* gene expression computed over all single cells.

Comparison of BCL11B binding with single cell ATAC-seq

As described in Takayama et al. (50), cells from three sorted CD34+CD38-CD45RA- and three CD34+CD38+ populations were processed on the 10X Genomics single cell ATAC-seq platform, and 12,414 single cells were retained based on default cellranger QC criterion and chromVAR depth filtering (51). Cellranger-reanalyze (1.1, 10x Genomics) was subsequently used to map reads over the sites identified in the bulk ATAC-Seq catalogue, and read counts were binarized. chromVAR (with default settings) (51) was used to calculate the enrichment of each of the hematopoietic signatures identified via non-negative matrix factorization (NMF) in each single cell, as well as for BCL11B ChIP-Seq in DND-41 and SJTALL005006 cells. The UMAP package in R (117) was used with default parameters to reduce the dimensionality of the enrichment of the signatures in each cell for visualization, with 21 outlier cells excluded from subsequent analyses.

Lentiviral production

293T cells were cultured in Dulbecco’s modified Eagle’s Medium (DMEM)/10% FBS supplemented with 1X penicillin-streptomycin-glutamine (Invitrogen). To produce concentrated lentivirus, 6 μg CL20c-MSCV lentiviral backbones were co-transfected with packaging vectors (3 μg pCAG-kGF1–1R, 1 μg pCAG-VSVG, and 1 μg pCAG4-RTR2) into 293T cells using PEIpro (Polyplus). Lentiviral vectors were: CL20-MSCV-IRES-GFP (empty vector GFP), CL20-MSCV-IRES-mCherry (empty vector mCherry), CL20-MSCV-BCL11B-IRES-GFP, or CL20-MSCV-FLT3-ITD-IRES-mCherry. Vector supernatants were collected forty-eight hours post-transfection, clarified by centrifugation at 330 × *g* for 5 minutes and 0.22 μM filtered. Lentiviral vector containing supernatants were adjusted to 300 mM NaCl, 50 mM Tris pH 8.0 and loaded onto an Acrodisc Mustang Q membrane (Pall Life

Sciences) according to the manufacturer's instructions using an Akta Avant chromatography system (GE Healthcare Bio-Sciences). After washing the column with 10 column volumes of 300 mM NaCl, 50 mM Tris pH 8.0, viral particles were eluted from the column using 2 M NaCl, Tris pH 8.0 directly onto a PD10 desalting column (GE Healthcare) according to the manufacturer's instructions. Vector containing flow-through was diluted with an equal volume of X-VIVO 10 media (Lonza) or phosphate buffered saline containing 1% human serum albumin (Grifols Biologicals) to achieve an approximate 50-fold concentration from the starting material, 0.22 μM sterile filtered, aliquoted and stored at −80°C.

For lentiviral production to infect mouse lineage-negative cells, 10 μg of each CL20 viral expression vector was mixed with 2 μg of each packaging vector (CAG4-Eco, CAG-KGP1–1R, CAG4-RTR2) and transfected into 293T cells using FuGene HD transfection reagent (Promega E2311). Viral supernatant was collected 48 and 72 hours post-transfection and filtered through a 0.45 μM strainer.

Transduction of cbCD34+ HSPCs

cbCD34+ cells were expanded for 3 days in HSC expansion media (66) (StemSpan SFEM II (StemCell Technologies 09655) supplemented with the following cytokines each at 100 ng/mL: IL6 (Peprotech 200–06), thrombopoietin (Peprotech 300–18), SCF (Peprotech 300–07), and FLT3 ligand (Peprotech 300–19)) and then transduced at a concentration of 4×10^6 cells/mL using 50% volume of concentrated virus with cytokines adjusted accordingly. For transcriptional analysis of *BCL11B*-overexpression, cbCD34+ cells were transduced with CL20-MSCV-GFP (empty vector) or CL20-MSCV-BCL11B-IRES-GFP. For in vitro differentiation (CFU and liquid culture), cbCD34+ cells were transduced with two viruses: GFP and mCherry empty vectors, CL20-MSCV-BCL11B-IRES-GFP + mCherry empty vector, CL20-MSCV-FLT3-ITD-IRES-mCherry + GFP empty vector, or CL20-MSCV-BCL11B-IRES-GFP + CL20-MSCV-FLT3-ITD-IRES-mCherry vectors. Transduced cells were isolated by FACS and then expanded an additional 48 hours in HSC expansion media before cells were harvested for RNA isolation, or plated immediately for in vitro differentiation.

Transduction of mouse lineage-negative HSPCs

Freshly filtered virus was spun onto 6-well non-tissue culture treated plates coated with RetroNectin (Takara T100B) for 90 minutes at 3,000 RPM at 4°C. Mouse lineage-negative HSPCs were obtained from 6–8 week old wild-type C57BL/6 bone marrow using the EasySep Mouse Hematopoietic Progenitor Cell Isolation Kit (StemCell Technologies 19856). Cells were infected on RetroNectin-coated plates for 48 h with CL20-MSCV lentivirus expressing the genes of interest (GFP + mCherry empty vectors, CL20-MSCV-BCL11B-IRES-GFP + mCherry empty vector, CL20-MSCV-FLT3-ITD-IRES-mCherry + GFP empty vector, or CL20-MSCV-BCL11B-IRES-GFP + CL20-MSCV-FLT3-ITD-IRES-mCherry). Cells were maintained in IMDM supplemented with 20% FBS, 50 ng/mL mSCF, 40 ng/mL mFlt3, 30 ng/mL mIL6, 20 ng/mL mIL3, and 10 ng/mL mIL7 (all from Peprotech) with 1X Penicillin/Streptomycin for the duration of the transduction.

In vitro differentiation of human CD34+ HSPCs

For in vitro differentiation, GFP+/mCherry+ cells were isolated by FACS and washed once with PBS. For colony forming assays, 1,500 cells were plated in triplicate in 1.1 mL Methocult supplemented with recombinant human SCF, IL-3, IL-6, EPO, G-CSF and GM-CSF (StemCell Technologies H4435E). Colonies were counted after 14 days. For liquid culture differentiation, cells were plated at a density of 20,000 cells/mL in myeloid differentiation media (StemSpan SFEM II (StemCell Technologies 09655) supplemented with 10 ng/mL GM-CSF, 10 ng/mL G-CSF, 10 ng/mL IL-6, 10 ng/mL IL-3, and 100 ng/mL SCF), or lymphoid differentiation media (StemSpan SFEM II (StemCell Technologies 09655) supplemented with 1000 U/mL IL-2, 5 ng/mL IL-3, 20 ng/mL IL-7, 20 ng/mL SCF, and 10 ng/mL FLT3-L). All recombinant human cytokines were purchased from Peprotech. Half-volume of media was added after 3 days, and cells were analyzed by flow cytometry on day 7. Surface and intracellular markers were stained for flow cytometry analysis using the Fix & Perm Cell Permeabilization Kit (Life Technologies GAS004) according to the manufacturer’s instructions. Briefly, cultured cells ($0.5 -1.0 \times 10^6$) were washed twice with PBS/BSA (0.5% BSA (w/v) in Dulbecco’s Phosphate Buffered Saline, DPBS (1X), Gibco14190–144) and resuspended in 200 µL of PBS/BSA. After adding 5 µL normal rabbit serum (ThermoFisher Scientific 10510), a cocktail of cell surface marker antibodies, containing CD45-Pac Orange (5 µL; Life Technologies MHCD4530), CD235a-PE (5 µL; Beckman Coulter IM2211U), CD34-PerCP-Cy5.5 (20 µL; BD Biosciences 347213), CD33-PE-Cy7 (10 µL; BD Biosciences 333949), CD7-AlexaFluor700 (5 µL; BD Biosciences 561603), and CD135 (Flt-3)-APC (5 µL; BioLegend 313308), was added to each tube and incubated in the dark at room temperature for 10 minutes. Cells were washed twice with PBS/BSA and resuspended in 200 µL FIX & PERM medium A, incubated at room temperature for 15 minutes and washed once with PBS/BSA. Cells were resuspended in 200 µL FIX & PERM medium B and a cocktail of intracellular marker antibodies containing CD3-APC-H7 (5 µL; BD Biosciences 641406) and MPO-eFluor450 (5 µL; Invitrogen 48–1299-42), was added and incubated at room temperature for 15 minutes in the dark. After washing twice with PBS/BSA, the stained cells were resuspended in 0.5 mL of PBS/BSA and at least 50,000 single cell events were acquired on a BD LSRFortessa cytometer. Data were analyzed using DIVA 8 and FlowJo v10.

Colony forming assays in mouse Lin- HSPCs

GFP+/mCherry+ cells were isolated using FACS, cells were washed once with PBS and plated in triplicate at a density of 5,000 cells per dish in 1.1 mL Methocult (StemCell Techologies M3434) with mGM-CSF added at a final concentration of 10 ng/mL. 5,000 cells were replated in triplicate every 7 days for 4 passages.

Data availability

RNA-seq, WGS and PacBio data generated from primary patient samples for this study have been deposited at the European Genome Phenome Archive, accession EGAS00001004810. HiChIP, ChIP-seq and RNA-seq (cbCD34+ only) data have been deposited at the Gene Expression Omnibus under accession GSE165209.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Lindsey E. Montefiori[#1], Sonja Bendig[#2], Zhaohui Gu[#3,4], Xiaolong Chen[#5], Petri Pölönen[1], Xiaotu Ma[5], Alex Murison[6], Andy Zeng[6], Laura Garcia-Prat[6], Kirsten Dickerson[1], Ilaria Iacobucci[1], Sherif Abdelhamed[1], Ryan Hiltenbrand[1], Paul E. Mead[1], Cyrus M. Mehr[1], Beisi Xu[7], Zhongshan Cheng[7], Ti-Cheng Chang[7], Tamara Westover[1], Jing Ma[1], Anna Stengel[2], Shunsuke Kimura[1], Chunxu Qu[1], Marcus B. Valentine[8], Marissa Rashkovan[9], Selina Luger[10], Mark R. Litzow[11], Jacob M. Rowe[12], Monique L. den Boer[13], Victoria Wang[14], Jun Yin[15], Steven M. Kornblau[16], Stephen P. Hunger[17], Mignon L. Loh[18], Ching-Hon Pui[19], Wenjian Yang[20], Kristine R. Crews[20], Kathryn G. Roberts[1], Jun J. Yang[20], Mary V. Relling[20], William E. Evans[20], Wendy Stock[21], Elisabeth M. Paietta[22], Adolfo A. Ferrando[9,23,24,25], Jinghui Zhang[5], Wolfgang Kern[2], Torsten Haferlach[2], Gang Wu[1,7], John E. Dick[6], Jeffery M. Klco[1,*], Claudia Haferlach[2,*], Charles G. Mullighan[1,*]

Affiliations

[1]Department of Pathology, St Jude Children’s Research Hospital Memphis, TN, US

[2]Munich Leukemia Laboratory, Munich, Germany

[3]Department of Computational and Quantitative Medicine, City of Hope Comprehensive Cancer Center, Duarte, CA, US

[4]Department of Systems Biology, City of Hope Comprehensive Cancer Center, Duarte, CA US

[5]Department of Computational Biology, St Jude Children’s Research Hospital, Memphis, TN US

[6]Department of Molecular Genetics, University of Toronto, Toronto Canada

[7]Center for Applied Bioinformatics, St. Jude Children’s Research Hospital, Memphis, TN US

[8]Cytogenetics Core Facility, St. Jude Children’s Research Hospital, Memphis, TN US

[9]Institute for Cancer Genetics, Columbia University, New York, NY US

[10]Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA US

[11]Division of Hematology, Department of Internal Medicine, Mayo Clinic, Rochester, MN US

[12]Department of Hematology, Shaare Zedek Medical Center, Jerusalem, Israel

[13]Princess Máxima Center for Pediatric Oncology, Utrecht, Netherlands

[14]Department of Data Science, Dana Farber Cancer Institute, Boston, MA USA

[15]Division of Clinical Trials and Biostatistics, Mayo Clinic, Rochester, MN US

[16]Department of Leukemia, The University of Texas MD Anderson Cancer Center, Houston, TX US

[17]Department of Pediatrics, Children's Hospital of Philadelphia, and the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA US

[18]Department of Pediatrics, Benioff Children's Hospital and Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, CA US

[19]Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN US

[20]Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN US

[21]University of Chicago Comprehensive Cancer Center, 5841 S. Maryland Ave, M/C2115, Chicago IL US

[22]Department of Oncology, Montefiore Medical Center, Bronx, NY US

[23]Department of Pediatrics, Columbia University, New York, NY US

[24]Department of Pathology and Cell Biology, Columbia University, New York, NY US

[25]Department of Systems Biology, Columbia University, New York, NY US

Funding

This study was supported by grants from the National Institutes of Health R35 CA197695 and P30 CA021765 (to C.G.M.), R01 CA216391 (to J.Z.), R35 CA210065 and P30 CA103696 (to A.A.F.), F32 CA254140 (to L.E.M.), P50 GM115279-01 (to M.V.R and M.L.L), R00 CA241297 (to Z.G.), The St. Jude Children's Research Hospital Chromatin Collaborative, the Henry Schueler 41&9 Foundation (C.G.M.), a St. Baldrick's Foundation Robert H. Arceci Innovation award (C.G.M.), a Burroughs Wellcome Fund Career Award for Medical Scientists (J.M.K.), and the Leukemia & Lymphoma Society's Career Development Program Special Fellow (to Z.G.). This study was conducted in part by the ECOG-ACRIN Cancer Research Group (Peter J. O'Dwyer, MD and Mitchell D. Schnall, MD, PhD, Group Co-Chairs) and supported by the National Cancer Institute of the National Institutes of Health under the following award numbers: U10CA180820, U10CA180794, UG1CA189859, UG1CA232760, UG1CA233234, UG1CA233290, U24CA196171 and U10CA180821 (to E.M.P. and M.R.L.). M.R. is a Damon Runyon-Sohn Pediatric Cancer Fellow supported by the Damon Runyon Cancer Research Foundation (DRSG-2017). This study was supported by funds to J.E.D. from the Princess Margaret Cancer Centre through funding provided by Ontario Ministry of Health, Princess Margaret Cancer Centre Foundation, Ontario Institute for Cancer Research through funding provided by the Government of Ontario, Canadian Institutes for Health Research (RN380110 - 409786), Canadian Cancer Society (grant #703212), Terry Fox New Frontiers Program Project Grant and a Canada Research Chair.

Conflicts of interest

C.G.M. has received research funding from Abbvie and Pfizer; speaking fees from Amgen, advisory board fees from Illumina, and holds stock in Amgen. E.A.P has consulted for Supertechs Inc; S.M.L. received honoraria from Daiichi-Sankyo, Pfizer, Bristol-Meyers Squibb, Acceleron, Agios, Loxo Onxology and has institutional research support from Onconova, Kira, Hoffman La Roche, Ariad and Biosight. A.F. consults for Ayala Pharmaceuticals and SpringWorks Therapeutics received previous research support from Pfizer, Bristol Myers Squib, Merck, Eli Lilly, and receives patent and reagent licensing royalties from Novartis, EMD Millipore and Applied Biological Materials. I. I. received honoraria from Amgen. M.L. receives research support from Abbvie, Astellas, Amgen, Actinium, Pluristem, serves on the Advisory Board of Omeros Corporation and Jazz Pharmaceuticals, and is on the Data Monitoring Committee of Biosight. C-H Pui is on the scientific advisory board of Adaptive Biotechnology Inc., the Data Monitoring Committee of Novartis, and received honorarium from Amgen. M.V.R and St. Jude receive research funding from Servier, and M.V.R's spouse is on the board of BioSkryb, Inc. W.E.E. serves on the board of BioSkryb, Inc. C.H., W.K. and T.H. are part owners of the MLL Munich Leukemia Laboratory, S.S. and A.S. are employed by the MLL Munich Leukemia laboratory.

REFERENCES

1. Khan M, Siddiqi R, Naqvi K. An update on classification, genetics, and clinical approach to mixed phenotype acute leukemia (MPAL). Ann Hematol 2018;97(6):945–53. [PubMed: 29546454]
2. Charles NJ, Boyer DF. Mixed-Phenotype Acute Leukemia: Diagnostic Criteria and Pitfalls. Archives of pathology & laboratory medicine 2017;141(11):1462–8. [PubMed: 29072953]
3. 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(2):147–56. [PubMed: 19147408]
4. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature 2012;481(7380):157–63. [PubMed: 22237106]
5. Alexander TB, Gu Z, Iacobucci I, Dickerson K, Choi JK, Xu B, et al. The genetic basis and cell of origin of mixed phenotype acute leukaemia. Nature 2018;562(7727):373–9. [PubMed: 30209392]
6. Neumann M, Greif PA, Baldus CD. Mutational landscape of adult ETP-ALL. Oncotarget 2013;4(7):952–3. [PubMed: 23807748]
7. Gu Z, Churchman ML, Roberts KG, Moore I, Zhou X, Nakitandwe J, et al. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. Nat Genet 2019;51(2):296–307. [PubMed: 30643249]
8. Hirabayashi S, Ohki K, Nakabayashi K, Ichikawa H, Momozawa Y, Okamura K, et al. ZNF384-related fusion genes define a subgroup of childhood B-cell precursor acute lymphoblastic leukemia with a characteristic immunotype. Haematologica 2017;102(1):118–29. [PubMed: 27634205]
9. Gutierrez A, Kentsis A. Acute myeloid/T-lymphoblastic leukaemia (AMTL): a distinct category of acute leukaemias with common pathogenesis in need of improved therapy. Br J Haematol 2018;180(6):919–24. [PubMed: 29441563]
10. He B, Gao P, Ding YY, Chen CH, Chen G, Chen C, et al. Diverse noncoding mutations contribute to deregulation of cis-regulatory landscape in pediatric cancers. Sci Adv 2020;6(30):eaba3064. [PubMed: 32832663]
11. Liu Y, Li C, Shen S, Chen X, Szlachta K, Edmonson MN, et al. Discovery of regulatory noncoding variants in individual cancer genomes by using cis-X. Nat Genet 2020;52(8):811–8. [PubMed: 32632335]
12. Rheinbay E, Nielsen MM, Abascal F, Wala JA, Shapira O, Tiao G, et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. Nature 2020;578(7793):102–11. [PubMed: 32025015]
13. Seki M, Kimura S, Isobe T, Yoshida K, Ueno H, Nakajima-Takagi Y, et al. Recurrent SPI1 (PU.1) fusions in high-risk pediatric T cell acute lymphoblastic leukemia. Nat Genet 2017;49(8):1274–81. [PubMed: 28671687]
14. Gianni F, Belver L, Ferrando A. The Genetics and Mechanisms of T-Cell Acute Lymphoblastic Leukemia. Cold Spring Harbor perspectives in medicine 2020;10(3).
15. Nagel S, Scherr M, Kel A, Hornischer K, Crawford GE, Kaufmann M, et al. Activation of TLX3 and NKX2–5 in t(5;14)(q35;q32) T-cell acute lymphoblastic leukemia by remote 3’-BCL11B enhancers and coregulation by PU.1 and HMGA1. Cancer Res 2007;67(4):1461–71. [PubMed: 17308084]
16. Bahr C, von Paleske L, Uslu VV, Remeseiro S, Takayama N, Ng SW, et al. A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. Nature 2018;553(7689):515–20. [PubMed: 29342133]
17. Downing JR, Wilson RK, Zhang J, Mardis ER, Pui CH, Ding L, et al. The pediatric cancer genome project. Nat Genet 2012;44(6):619–22. [PubMed: 22641210]
18. Palomo L, Meggendorfer M, Hutter S, Twardziok S, Adema V, Fuhrmann I, et al. Molecular landscape and clonal architecture of adult myelodysplastic/myeloproliferative neoplasms. Blood 2020;136(16):1851–62. [PubMed: 32573691]
19. McLeod C, Gout AM, Zhou X, Thrasher A, Rahbarinia D, Brady SW, et al. St. Jude Cloud: A Pediatric Cancer Genomic Data-Sharing Ecosystem. Cancer discovery 2021;11(5):1082–99. [PubMed: 33408242]

20. 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(5987):93–6. [PubMed: 20595615]

21. Ha VL, Luong A, Li F, Casero D, Malvar J, Kim YM, et al. The T-ALL related gene BCL11B regulates the initial stages of human T-cell differentiation. Leukemia 2017;31(11):2503–14. [PubMed: 28232744]

22. Li L, Leid M, Rothenberg EV. An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. Science 2010;329(5987):89–93. [PubMed: 20595614]

23. Sidwell T, Rothenberg EV. Epigenetic Dynamics in the Function of T-Lineage Regulatory Factor Bcl11b. Frontiers in immunology 2021;12:669498. [PubMed: 33936112]

24. Kueh HY, Yui MA, Ng KK, Pease SS, Zhang JA, Damle SS, et al. Asynchronous combinatorial action of four regulatory factors activates Bcl11b for T cell commitment. Nat Immunol 2016;17(8):956–65. [PubMed: 27376470]

25. Tydell CC, David-Fung ES, Moore JE, Rowen L, Taghon T, Rothenberg EV. Molecular dissection of prethymic progenitor entry into the T lymphocyte developmental pathway. J Immunol 2007;179(1):421–38. [PubMed: 17579063]

26. Hosokawa H, Romero-Wolf M, Yui MA, Ungerback J, Quiloan MLG, Matsumoto M, et al. Bcl11b sets pro-T cell fate by site-specific cofactor recruitment and by repressing Id2 and Zbtb16. Nat Immunol 2018;19(12):1427–40. [PubMed: 30374131]

27. Roels J, Kuchmiy A, De Decker M, Strubbe S, Lavaert M, Liang KL, et al. Distinct and temporary-restricted epigenetic mechanisms regulate human alphabeta and gammadelta T cell development. Nat Immunol 2020;21(10):1280–92. [PubMed: 32719521]

28. 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(15):4169–73. [PubMed: 21878675]

29. Liu Y, Easton J, Shao Y, Maciaszek J, Wang Z, Wilkinson MR, et al. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. Nat Genet 2017;49(8):1211–8. [PubMed: 28671688]

30. 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(11):1321–7. [PubMed: 20972433]

31. Bernard OA, Busson-LeConiat M, Ballerini P, Mauchauffé M, Della Valle V, Monni R, et al. A new recurrent and specific cryptic translocation, t(5;14)(q35;q32), is associated with expression of the Hox11L2 gene in T acute lymphoblastic leukemia. Leukemia 2001;15(10):1495–504. [PubMed: 11587205]

32. Van Vlierberghe P, Homminga I, Zuurbier L, Gladdines-Buijs J, van Wering ER, Horstmann M, et al. Cooperative genetic defects in TLX3 rearranged pediatric T-ALL. Leukemia 2008;22(4):762–70. [PubMed: 18185524]

33. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518(7539):317–30. [PubMed: 25693563]

34. Zhang X, Jeong M, Huang X, Wang XQ, Wang X, Zhou W, et al. Large DNA Methylation Nadirs Anchor Chromatin Loops Maintaining Hematopoietic Stem Cell Identity. Mol Cell 2020;78(3):506–21. [PubMed: 32386543]

35. Herranz D, Ambesi-Impiombato A, Palomero T, Schnell SA, Belver L, Wendorff AA, et al. A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. Nat Med 2014;20(10):1130–7. [PubMed: 25194570]

36. Lancho O, Herranz D. The MYC Enhancer-ome: Long-Range Transcriptional Regulation of MYC in Cancer. Trends Cancer 2018;4(12):810–22. [PubMed: 30470303]

37. 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(2):467–82. [PubMed: 22500808]

38. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat Methods 2016;13(11):919–22. [PubMed: 27643841]

39. Schneider U, Schwenk HU, Bornkamm G. Characterization of EBV-genome negative “null” and “T” cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma. Int J Cancer 1977;19(5):621–6. [PubMed: 68013]

40. Drexler HG, Gaedicke G, Minowada J. Isoenzyme studies in human leukemia-lymphoma cell lines−-1. Carboxylic esterase. Leuk Res 1985;9(2):209–29. [PubMed: 2985879]

41. Isoda T, Moore AJ, He Z, Chandra V, Aida M, Denholtz M, et al. Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate. Cell 2017;171(1):103–19. [PubMed: 28938112]

42. Li L, Zhang JA, Dose M, Kueh HY, Mosadeghi R, Gounari F, et al. A far downstream enhancer for murine Bcl11b controls its T-cell specific expression. Blood 2013;122(6):902–11. [PubMed: 23741008]

43. Juric I, Yu M, Abnousi A, Raviram R, Fang R, Zhao Y, et al. MAPS: Model-based analysis of long-range chromatin interactions from PLAC-seq and HiChIP experiments. PLoS Comput Biol 2019;15(4):e1006982. [PubMed: 30986246]

44. Northcott PA, Lee C, Zichner T, Stutz AM, Erkek S, Kawauchi D, et al. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. Nature 2014;511(7510):428–34. [PubMed: 25043047]

45. Radtke I, Mullighan CG, Ishii M, Su X, Cheng J, Ma J, et al. Genomic analysis reveals few genetic alterations in pediatric acute myeloid leukemia. Proc Natl Acad Sci U S A 2009;106(31):12944–9. [PubMed: 19651601]

46. Zhou W, Zhang F, Chen X, Shen Y, Lupski JR, Jin L. Increased genome instability in human DNA segments with self-chains: homology-induced structural variations via replicative mechanisms. Hum Mol Genet 2013;22(13):2642–51. [PubMed: 23474816]

47. Lam KW, Jeffreys AJ. Processes of de novo duplication of human alpha-globin genes. Proc Natl Acad Sci U S A 2007;104(26):10950–5. [PubMed: 17573529]

48. Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, et al. Massive mining of publicly available RNA-seq data from human and mouse. Nature communications 2018;9(1):1366.

49. Kojo S, Tanaka H, Endo TA, Muroi S, Liu Y, Seo W, et al. Priming of lineage-specifying genes by Bcl11b is required for lineage choice in post-selection thymocytes. Nature communications 2017;8(1):702.

50. Takayama N, Murison A, Takayanagi SI, Arlidge C, Zhou S, Garcia-Prat L, et al. The Transition from Quiescent to Activated States in Human Hematopoietic Stem Cells Is Governed by Dynamic 3D Genome Reorganization. Cell Stem Cell 2021;28(3):488–501 e10. [PubMed: 33242413]

51. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods 2017;14(10):975–8. [PubMed: 28825706]

52. Radtke F, Wilson A, Stark G, Bauer M, van Meerwijk J, MacDonald HR, et al. Deficient T cell fate specification in mice with an induced inactivation of Notch1. Immunity 1999;10(5):547–58. [PubMed: 10367900]

53. Schmitt TM, Ciofani M, Petrie HT, Zúñiga-Pflücker JC. Maintenance of T cell specification and differentiation requires recurrent notch receptor-ligand interactions. J Exp Med 2004;200(4):469–79. [PubMed: 15314075]

54. Berquam-Vrieze KE, Nannapaneni K, Brett BT, Holmfeldt L, Ma J, Zagorodna O, et al. Cell of origin strongly influences genetic selection in a mouse model of T-ALL. Blood 2011;118(17):4646–56. [PubMed: 21828136]

55. Riemke P, Czeh M, Fischer J, Walter C, Ghani S, Zepper M, et al. Myeloid leukemia with transdifferentiation plasticity developing from T-cell progenitors. Embo j 2016;35(22):2399–416. [PubMed: 27572462]

56. Padella A, Simonetti G, Paciello G, Giotopoulos G, Baldazzi C, Righi S, et al. Novel and Rare Fusion Transcripts Involving Transcription Factors and Tumor Suppressor Genes in Acute Myeloid Leukemia. Cancers (Basel) 2019;11(12).

57. Abbas S, Sanders MA, Zeilemaker A, Geertsma-Kleinekoort WM, Koenders JE, Kavelaars FG, et al. Integrated genome-wide genotyping and gene expression profiling reveals BCL11B as a putative oncogene in acute myeloid leukemia with 14q32 aberrations. Haematologica 2014;99(5):848–57. [PubMed: 24441149]

58. Bezrookove V, van Zelderen-Bhola SL, Brink A, Szuhai K, Raap AK, Barge R, et al. A novel t(6;14)(q25-q27;q32) in acute myelocytic leukemia involves the BCL11B gene. Cancer Genet Cytogenet 2004;149(1):72–6. [PubMed: 15104287]

59. Georgy M, Yonescu R, Griffin CA, Batista DA. Acute mixed lineage leukemia and a t(6;14)(q25;q32) in two adults. Cancer Genet Cytogenet 2008;185(1):28–31. [PubMed: 18656690]

60. Hayashi Y, Pui CH, Behm FG, Fuchs AH, Raimondi SC, Kitchingman GR, et al. 14q32 translocations are associated with mixed-lineage expression in childhood acute leukemia. Blood 1990;76(1):150–6. [PubMed: 2364166]

61. Pallavajjala A, Kim D, Li T, Ghiaur G, Jones RJ, Burns KH, et al. Genomic characterization of chromosome translocations in patients with T/myeloid mixed-phenotype acute leukemia. Leuk Lymphoma 2018;59(5):1231–8. [PubMed: 28882084]

62. Wang W, Beird H, Kroll CJ, Hu S, Bueso-Ramos CE, Fang H, et al. T(6;14)(q25;q32) involves BCL11B and is highly associated with mixed-phenotype acute leukemia, T/myeloid. Leukemia 2020;34(9):2509–12. [PubMed: 32099038]

63. Wu SQ, Kuo J, Chen XR, Chen SA, Quinn JJ. Translocation (6;14) in childhood acute mixed lineage leukemia. Cancer Genet Cytogenet 2003;141(2):178–9. [PubMed: 12606141]

64. Di Giacomo D, La Starza R, Gorello P, Pellanera F, Kalender Atak Z, De Keersmaecker K, et al. 14q32 rearrangements deregulating BCL11B mark a distinct subgroup of T and myeloid immature acute leukemia. Blood 2021. doi 10.1182/blood.2020010510

65. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 2016;127(20):2391–405. [PubMed: 27069254]

66. Boitano AE, Wang J, Romeo R, Bouchez LC, Parker AE, Sutton SE, et al. Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. Science 2010;329(5997):1345–8. [PubMed: 20688981]

67. Kechin A, Boyarskikh U, Kel A, Filipenko M. cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. Journal of computational biology : a journal of computational molecular cell biology 2017;24(11):1138–43. [PubMed: 28715235]

68. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29(1):15–21. [PubMed: 23104886]

69. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011;12:323. [PubMed: 21816040]

70. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 2010;11(3):R25. [PubMed: 20196867]

71. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol 2014;15(2):R29. [PubMed: 24485249]

72. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43(7):e47. [PubMed: 25605792]

73. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102(43):15545–50. [PubMed: 16199517]

74. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 2012;28(18):i333–i9. [PubMed: 22962449]

75. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol 2014;15(6):R84. [PubMed: 24970577]

76. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Kallberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 2016;32(8):1220–2. [PubMed: 26647377]

77. Cameron DL, Schroder J, Penington JS, Do H, Molania R, Dobrovic A, et al. GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. Genome Res 2017;27(12):2050–60. [PubMed: 29097403]

78. Chong Z, Ruan J, Gao M, Zhou W, Chen T, Fan X, et al. novoBreak: local assembly for breakpoint detection in cancer genomes. Nat Methods 2017;14(1):65–7. [PubMed: 27892959]

79. Sedlazeck FJ, Dhroso A, Bodian DL, Paschall J, Hermes F, Zook JM. Tools for annotation and comparison of structural variation. F1000Res 2017;6:1795. [PubMed: 29123647]

80. Ebler J, Schonhuth A, Marschall T. Genotyping inversions and tandem duplications. Bioinformatics 2017;33(24):4015–23. [PubMed: 28169394]

81. Kent WJ. BLAT--the BLAST-like alignment tool. Genome Res 2002;12(4):656–64. [PubMed: 11932250]

82. Rusch M, Nakitandwe J, Shurtleff S, Newman S, Zhang Z, Edmonson MN, et al. Clinical cancer genomic profiling by three-platform sequencing of whole genome, whole exome and transcriptome. Nature communications 2018;9(1):3962.

83. Zhang J, Benavente CA, McEvoy J, Flores-Otero J, Ding L, Chen X, et al. A novel retinoblastoma therapy from genomic and epigenetic analyses. Nature 2012;481(7381):329–34. [PubMed: 22237022]

84. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature 2018;555(7696):371–6. [PubMed: 29489755]

85. McNeer NA, Philip J, Geiger H, Ries RE, Lavallée VP, Walsh M, et al. Genetic mechanisms of primary chemotherapy resistance in pediatric acute myeloid leukemia. Leukemia 2019;33(8):1934–43. [PubMed: 30760869]

86. Bolouri H, Farrar JE, Triche T Jr., Ries RE, Lim EL, Alonzo TA, et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. Nat Med 2018;24(1):103–12. [PubMed: 29227476]

87. Edmonson MN, Zhang J, Yan C, Finney RP, Meerzaman DM, Buetow KH. Bambino: a variant detector and alignment viewer for next-generation sequencing data in the SAM/BAM format. Bioinformatics 2011;27(6):865–6. [PubMed: 21278191]

88. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013;31(3):213–9. [PubMed: 23396013]

89. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics 2012;28(3):311–7. [PubMed: 22155872]

90. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012;22(3):568–76. [PubMed: 22300766]

91. Fan Y, Xi L, Hughes DS, Zhang J, Zhang J, Futreal PA, et al. MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. Genome Biol 2016;17(1):178. [PubMed: 27557938]

92. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods 2018;15(8):591–4. [PubMed: 30013048]

93. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 2009;25(21):2865–71. [PubMed: 19561018]

94. Höllein A, Twardziok SO, Walter W, Hutter S, Baer C, Hernandez-Sanchez JM, et al. The combination of WGS and RNA-Seq is superior to conventional diagnostic tests in multiple myeloma: Ready for prime time? Cancer genetics 2020;242:15–24. [PubMed: 31980417]

95. Wang J, Mullighan CG, Easton J, Roberts S, Heatley SL, Ma J, et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution. Nat Methods 2011;8(8):652–4. [PubMed: 21666668]

96. Roberts KG, Gu Z, Payne-Turner D, McCastlain K, Harvey RC, Chen IM, et al. High Frequency and Poor Outcome of Philadelphia Chromosome-Like Acute Lymphoblastic Leukemia in Adults. J Clin Oncol 2017;35(4):394–401. [PubMed: 27870571]

97. Mayrhofer M, Viklund B, Isaksson A. Rawcopy: Improved copy number analysis with Affymetrix arrays. Scientific reports 2016;6:36158. [PubMed: 27796336]

98. Chen X, Gupta P, Wang J, Nakitandwe J, Roberts K, Dalton JD, et al. CONSERTING: integrating copy-number analysis with structural-variation detection. Nat Methods 2015;12(6):527–30. [PubMed: 25938371]

99. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 2017;27(5):722–36. [PubMed: 28298431]

100. Ovcharenko I, Loots GG, Hardison RC, Miller W, Stubbs L. zPicture: dynamic alignment and visualization tool for analyzing conservation profiles. Genome Res 2004;14(3):472–7. [PubMed: 14993211]

101. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. Nat Methods 2015;12(5):380–1. [PubMed: 25924071]

102. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754–60. [PubMed: 19451168]

103. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25(16):2078–9. [PubMed: 19505943]

104. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 2016;44(W1):W160–5. [PubMed: 27079975]

105. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell systems 2016;3(1):95–8. [PubMed: 27467249]

106. Zhou X, Edmonson MN, Wilkinson MR, Patel A, Wu G, Liu Y, et al. Exploring genomic alteration in pediatric cancer using ProteinPaint. Nat Genet 2016;48(1):4–6. [PubMed: 26711108]

107. Marco-Sola S, Sammeth M, Guigo R, Ribeca P. The GEM mapper: fast, accurate and versatile alignment by filtration. Nat Methods 2012;9(12):1185–8. [PubMed: 23103880]

108. Derrien T, Estelle J, Marco Sola S, Knowles DG, Raineri E, Guigo R, et al. Fast computation and applications of genome mappability. PLoS One 2012;7(1):e30377. [PubMed: 22276185]

109. Martin MC. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 2011;17(1):10–2.

110. Tischler G, Leonard S. biobambam: tools for read pair collation based algorithms on BAM files. Source Code Biol Med 2014;9.

111. 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(9):R137. [PubMed: 18798982]

112. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 2013;153(2):307–19. [PubMed: 23582322]

113. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 2010;38(4):576–89. [PubMed: 20513432]

114. Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res 2020;48(D1):D87–d92. [PubMed: 31701148]

115. Xie X, Liu M, Zhang Y, Wang B, Zhu C, Wang C, et al. Single-cell transcriptomic landscape of human blood cells. National Science Review 2021;8(3).

116. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 2019;37(7):773–82. [PubMed: 31061481]

117. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol 2019;37(1):38–44.

Statement of significance:

Lineage ambiguous leukemias pose significant diagnostic and therapeutic challenges due to a poorly understood molecular and cellular basis. We identify oncogenic deregulation of *BCL11B* driven by diverse structural alterations, including de novo super-enhancer generation, as the driving feature of a subset of lineage ambiguous leukemias that transcend current diagnostic boundaries.

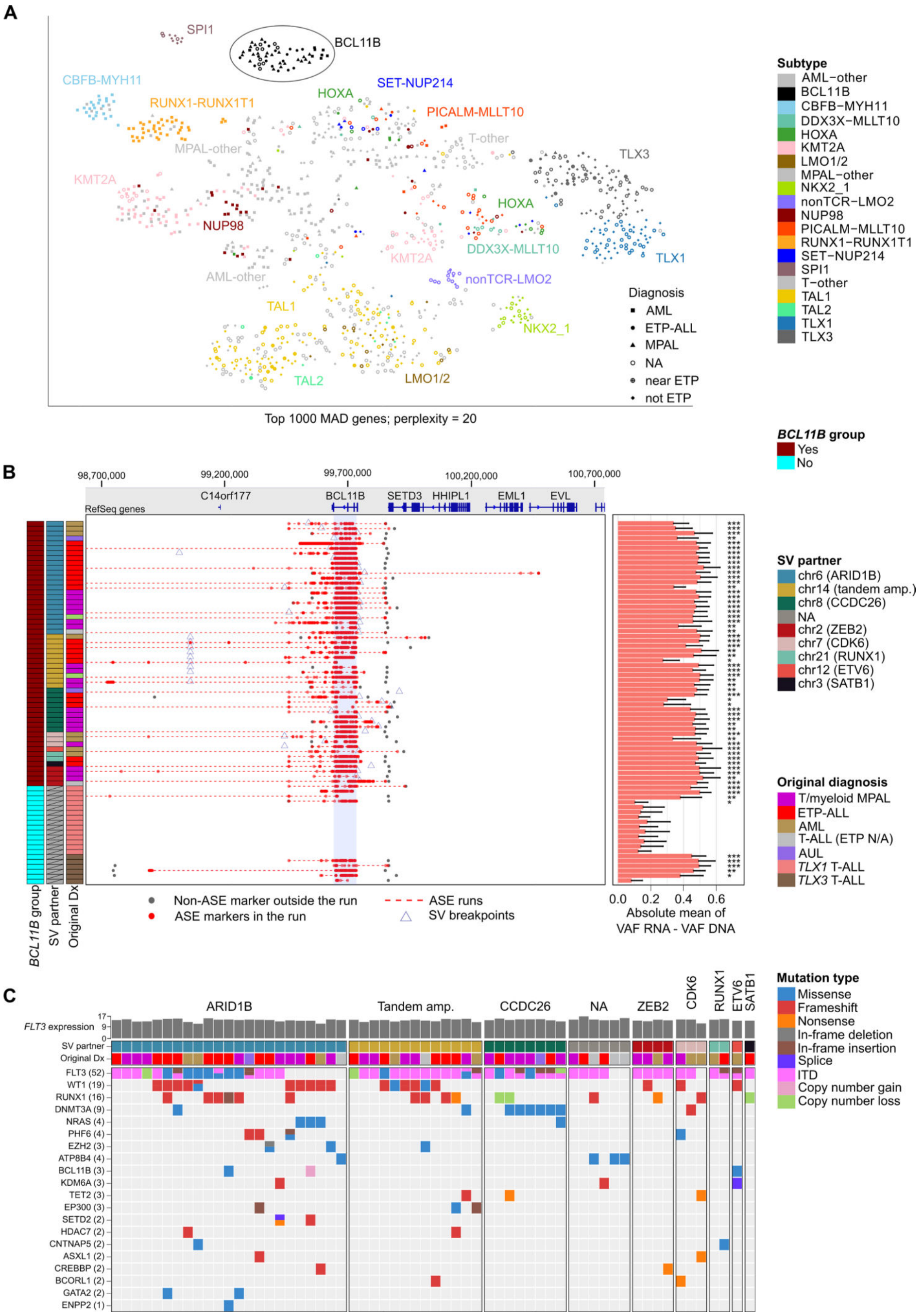

Fig 1. A new subtype of leukemia defined by a distinct gene expression profile and allele-specific *BCL11B* expression.

(**A**) tSNE projection analysis of 1,114 leukemia transcriptomes (excluding B-ALL and MPAL cases clustered with B-ALL). Samples are colored by driver genomic alterations and shaped corresponding to original diagnosis. The top 1000 most variable genes (based on absolute median deviation) were used in the tSNE analysis with the perplexity of 20. Samples belonging to the *BCL11B* group are circled. (**B**) Allele-specific expression (ASE) analysis of the *BCL11B* group samples and representative T-ALL samples. Each row represents a primary leukemia sample with available matched WGS and RNA-seq required to discern allele frequencies at heterozygous SNPs (see Methods and Supplementary Table 5). Red dots indicate positions of significant allelic imbalance in the RNA-seq data and dashed red lines indicate continuous runs such SNPs. Right panel shows the absolute mean difference between variant allele frequencies (VAF) in RNA-seq vs. WGS data. Significant detection of *BCL11B* ASE is indicated with asterisks. (**C**) Oncoprint of the *BCL11B* group showing the most recurrently mutated genes. Normalized *FLT3* expression levels (variance stabilizing transformation) are shown above. Samples are grouped according to the *BCL11B* SV rearrangement partner.

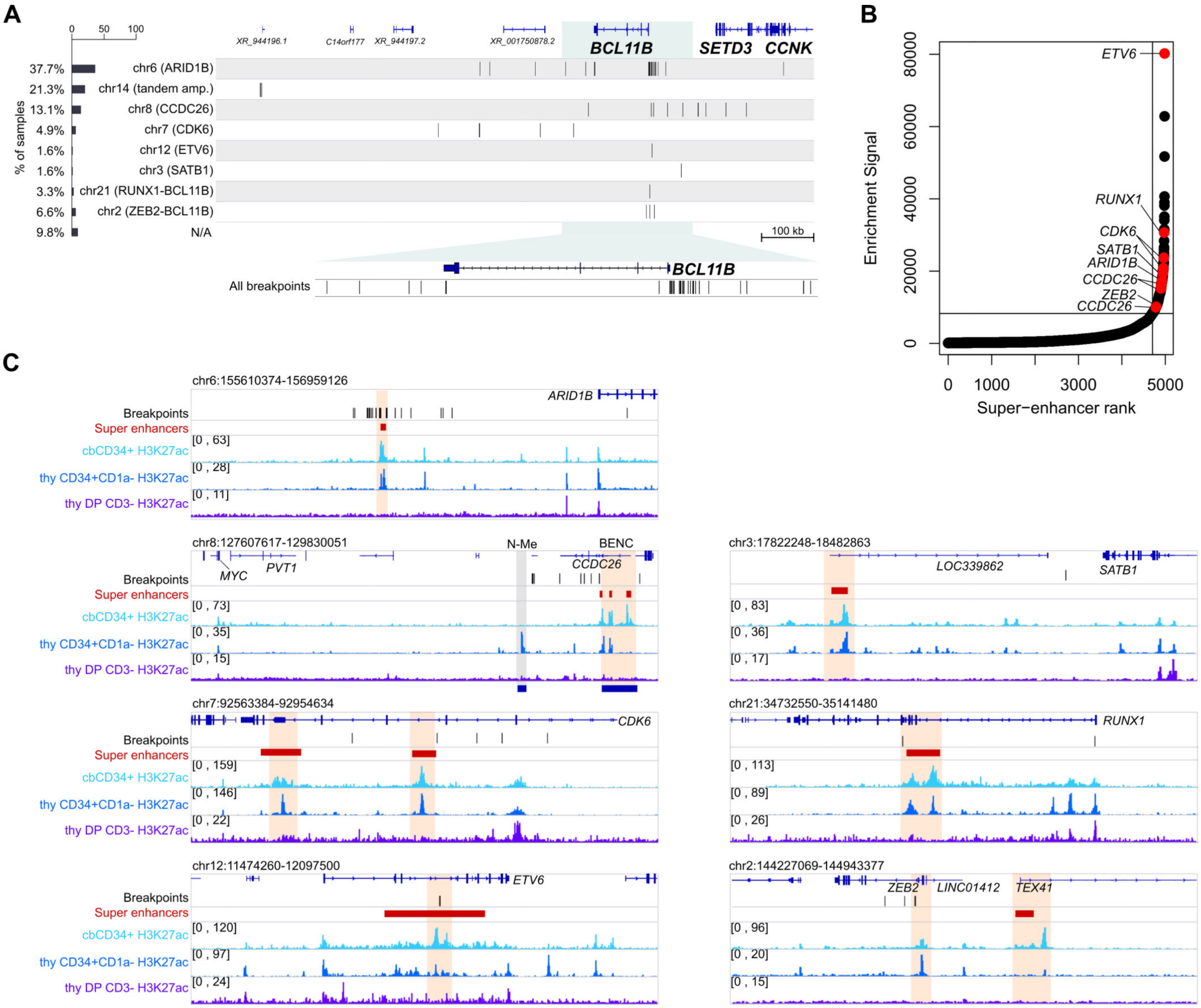

Fig 2. *BCL11B* SVs occur near cbCD34+ HSPC super enhancers.

(**A**) *BCL11B* breakpoint positions on chromosome 14, grouped by rearrangement partner. Below, zoomed in on the *BCL11B* gene to show breakpoints occur up- or down-stream of *BCL11B*. (**B**) Super-enhancer analysis of H3K27ac ChIP-seq data from cbCD34+ cells (34). Loci harboring *BCL11B* rearrangements are shown in red. Horizontal line indicates enrichment cutoff, vertical line indicates super-enhancer cut-off. (**C**) Genome browser tracks of *BCL11B* rearrangement partner loci. Breakpoint positions are indicated with black tick marks along with super-enhancer calls (red bars) from cbCD34+ data (light blue track). H3K27ac ChIP-seq coverage tracks are also shown for purified human thymocytes (CD34+CD1a− and double positive (DP) CD3− progenitors) (27). Red highlighted area corresponds to predicted hijacked HSPC enhancers. Grey bar highlights the N-Me T cell progenitor-specific enhancer.

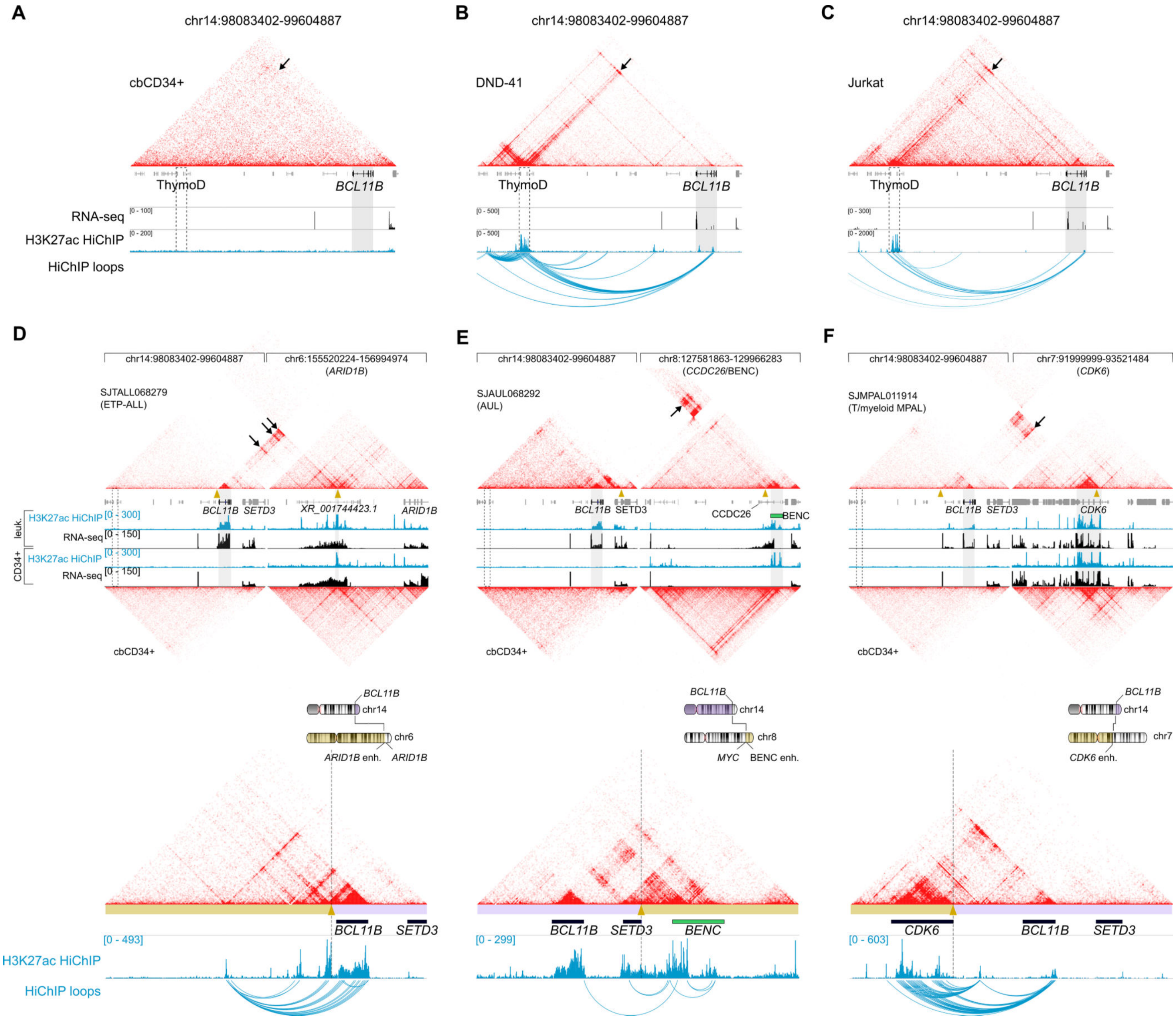

Fig 3. *BCL11B* rearrangements rewire existing CD34+ HSPC super-enhancers.

(**A-C**) H3K27ac HiChIP in human cbCD34+ HSPCs (**A**), DND-41 (*BCL11B-TLX3*) T-ALL cells (**B**) and Jurkat (*TAL1* deregulated) T-ALL cells (**C**). Raw interaction maps are displayed as a heatmap and HiChIP coverage tracks are shown below. Significant H3K27ac-anchored interactions (FDR <0.01) are shown as arcs. Dotted grey box indicates the position of the ThymoD enhancer, shaded grey box indicates the *BCL11B* gene, and black arrows point to the region of the heatmap corresponding to ThymoD-*BCL11B* interactions. (**D-F**) Each panel shows H3K27ac HiChIP data from a different patient sample, with comparison to the same genomic region in healthy normal cbCD34+ HSPCs. Raw interaction maps are displayed at 5 kb resolution and breakpoint positions are shown as orange arrowheads. HiChIP and RNA-seq coverage are shown below for both leukemia and cbCD34+ samples. Bottom panel shows interaction maps generated using patient-specific genomes containing the rearranged chromosome sequences. Purple bars indicate the chromosome 14 (*BCL11B* containing) derived region and gold bars indicate the SV partner-derived region. Insets show schematic representations of each rearrangement. (**D**) A patient sample harboring an *ARID1B-BCL11B* rearrangement. (**E**) Same as in (**D**), for a patient sample harboring a *CCDC26*/BENC-*BCL11B* rearrangement and (**C**) for a patient sample with a *CDK6-BCL11B* rearrangement. All coordinates are hg38.

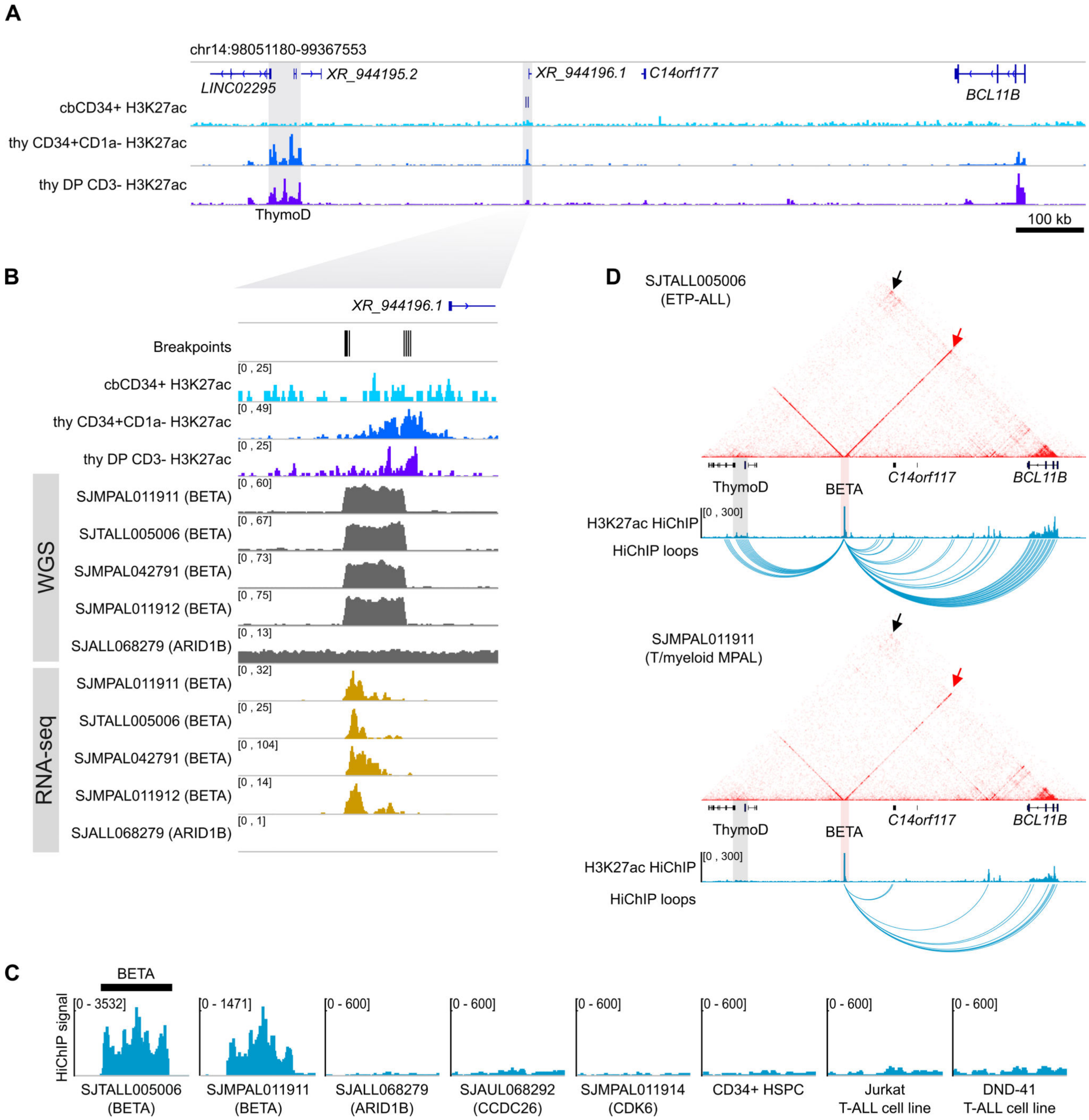

Fig 4. Tandem amplification of a non-coding region generates a de novo super enhancer.

(**A**) Genome browser snapshot showing the *BCL11B* locus and ~1Mb downstream gene desert. The positions of the T cell enhancer, ThymoD, and the de novo amplified genomic region found in 20% of cases (BETA) are highlighted in grey. H3K27ac ChIP-seq coverage tracks are shown for cbCD34+ HSPCs as well as thymic CD34+CD1a− progenitors and committed DP CD3- thymocytes (27). (**B**) Genomic region centered on BETA. WGS and RNA-seq coverage are shown for 4 representative BETA cases along with a non-BETA case (SJALL068279, *ARID1B-BCL11B*) for comparison. (**C**) H3K27ac HiChIP coverage centered on BETA in all cell types analyzed. (**D**) H3K27ac HiChIP data in 2 BETA cases. Heatmaps show the raw pair-wise interaction frequencies and significant chromatin interactions are shown as arcs. Black arrows point to the location of ThymoD-*BCL11B* chromatin interactions and red arrows point to the BETA-*BCL11B* interaction. Coordinates shown are chr14:98051180–99367553 (hg38).

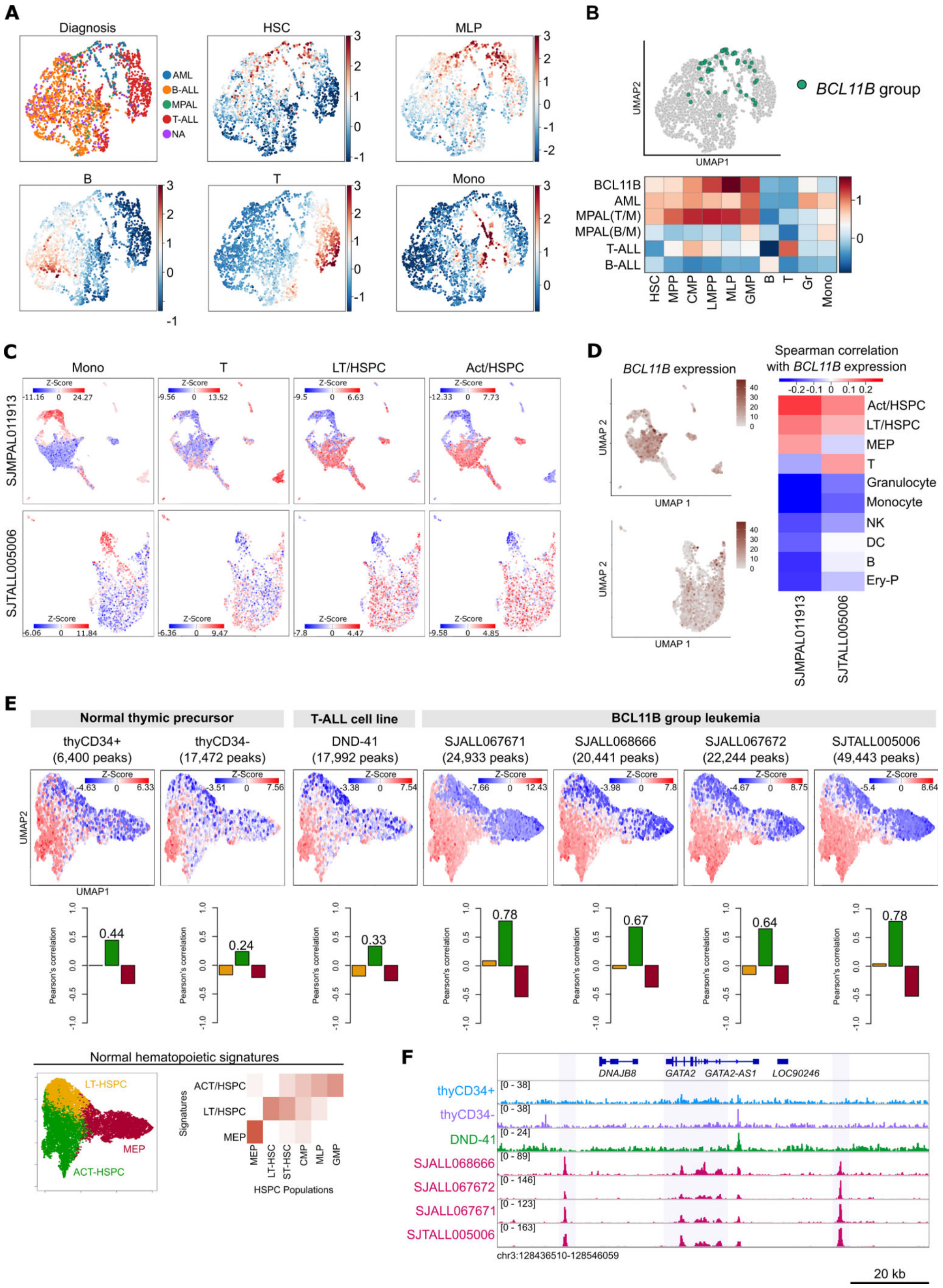

Fig 5. BCL11B binding correlates with an HSPC gene expression signature in *BCL11B*-deregulated leukemia.

(**A-B**) CIBERSORT deconvolution of 2467 leukemia transcriptomes with signatures from purified HSPC and mature populations. (**A**) Patient samples are colored based on their diagnosis and enrichment for each gene expression signature. (**B**) *BCL11B* group leukemia samples were projected onto the HSPC 2D UMAP showing that they cluster with the hematopoietic stem cell (HSC)/lymphoid-myeloid primed progenitor (LMPP) populations (top) and the summary of enrichment z-scores for all leukemia-normal comparisons is shown below. (**C**) Combined single cell ATAC-seq/RNA-seq on two *BCL11B* group samples. UMAP clustering was performed after combining both modalities. Cells are colored based on their enrichment for each normal hematopoietic open chromatin signature (using the scATAC-seq data). See Supplementary Fig. 21 for all signatures analyzed. (**D**) *BCL11B* expression from each sample is plotted in the respective UMAP (left) and the correlation of *BCL11B* expression with each normal hematopoietic signature is shown at right. Heatmap shows the Spearman's correlation z-score. (**E**) Three HSPC-spanning chromatin accessibility signatures representing normal hematopoietic cells identified in Takayama et al. (47) are shown at the bottom. Columns in the heatmap correspond to six HSPC cell populations reflected by these signature groupings (rows), ranked from lowest (white) to highest (red) values. (top) UMAP 2D projection of normal HSPC populations colored based on their enrichment for BCL11B ChIP-seq peaks in the indicated sample, with barplots showing the correlation between the single-cell enrichment of BCL11B ChIP-Seq peaks and the three HSPC-spanning chromatin accessibility signatures. (**F**) BCL11B occupancy at the *GATA2* locus in normal thymocytes, DND-41 cells, and four *BCL11B*-group leukemia samples. Shaded areas highlight regions of BCL11B occupancy specifically in the *BCL11B* group samples. Coordinates are hg38. HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; LMPP, lymphoid-primed multipotent progenitor; MLP, multi-lymphoid progenitor; GMP, granulocyte-macrophage progenitor; B, B-cell (CD19+); T, T-cell (CD3+); Gr, Granulocyte; Mono, monocyte.

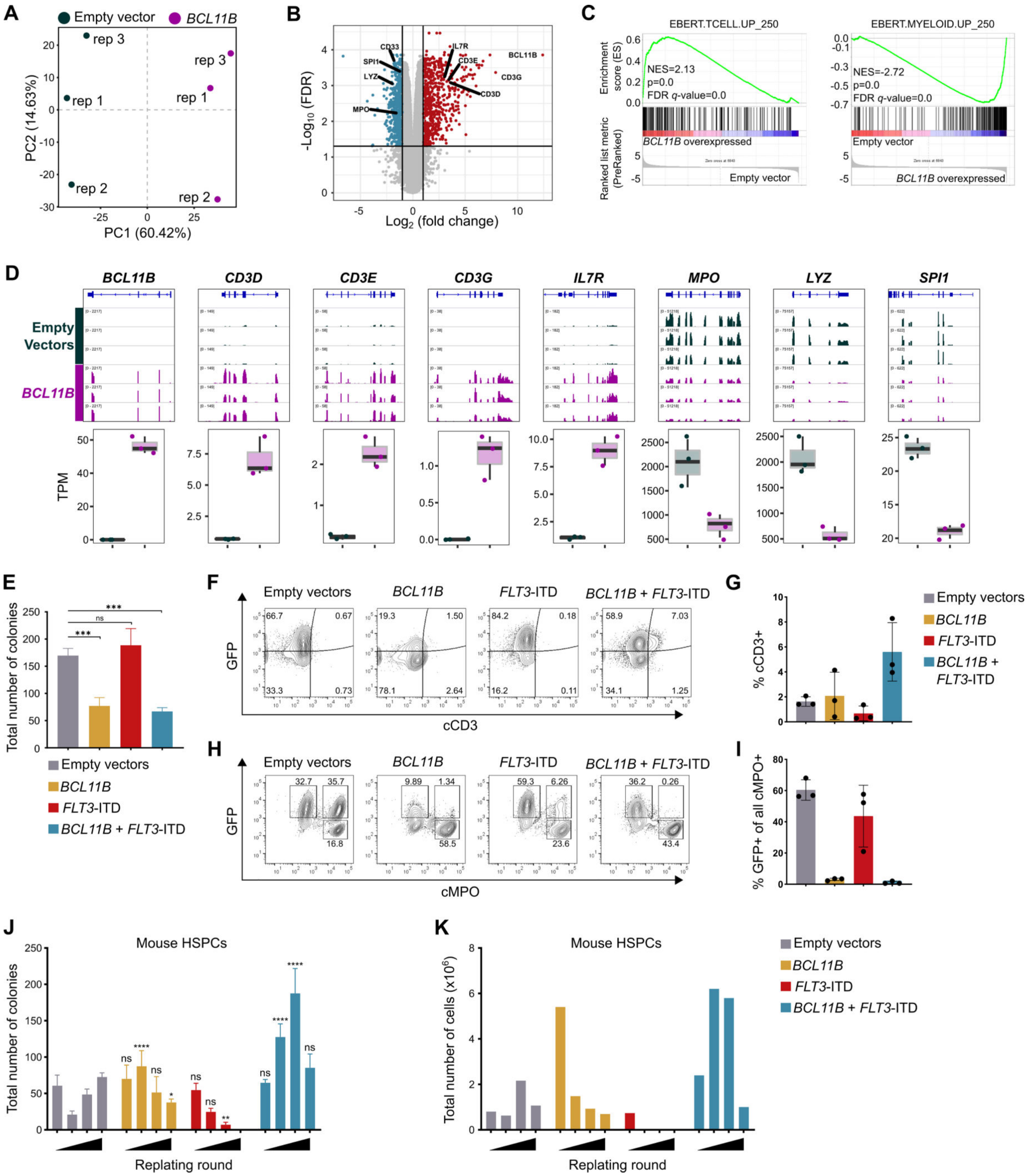

Fig 6. *BCL11B* and *FLT3*-ITD overexpression of in human and mouse HSPCs.

(A) Principle components analysis of RNA-seq data from 3 biological replicates of *BCL11B*- or empty vector-transduced cbCD34+ HSPCs using the top 3000 most variable genes. **(B)** Volcano plot showing genes up- (red) or down-regulated (blue) in *BCL11B*-overexpressing cells compared to control. Vertical lines indicate the fold change cutoff of 2; horizontal line indicates the FDR cutoff of 0.05. **(C)** GSEA of *BCL11B*-transduced versus empty vector control cbCD34+ HSPCs. Genes upregulated following *BCL11B* overexpression are positively enriched for T cell differentiation genes, whereas genes downregulated are negatively enriched for genes related to myeloid differentiation (a full list of GSEA results can be found in Supplementary Table 15). **(D)** Upper panel: coverage tracks of RNA-seq reads at selected genes (*CD3D*, *CD3E*, *CD3G*, *IL7R* are T-lineage genes upregulated following *BCL11B* overexpression; *MPO*, *LYZ*, *SPI1* are myeloid lineage genes downregulated following *BCL11B* overexpression). Lower panel: boxplots displaying TPM values for each replicate. Box shows the first and third quartiles; whiskers show data range. **(E)** Results from colony forming assays in GFP+/mCherry+ sorted human cbCD34+ cells. Total number of colonies across 3 technical replicates per condition (1,500 cells plated per dish) is shown. Data are representative of 3 biological replicates. ***p < 0.001 (one-way ANOVA with all samples compared to the empty vector control). **(F-I)** Flow cytometry analysis of transduced cbCD34+ cells grown for 7 days in lymphoid differentiation media **(F,G)** or myeloid differentiation media **(H,I)**. Cells shown were gated on CD45+ singlets. **(J,K)** Analysis of serial replating of GFP+/mCherry+ sorted mouse lineage-negative HSPCs. (**J**) Total number of colonies across 3 technical replicates per condition (5,000 cells plated per dish per round) is shown. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001, n.s., not significant (one-way ANOVA with all samples compared to the empty vector control). (**K**) Total number of cells generated in each round.