ARTICLE

https://doi.org/10.1038/s41467-019-10510-8

OPEN

Synthetic modeling reveals HOXB genes are critical for the initiation and maintenance of human leukemia

Manabu Kusakabe1,5, Ann Chong Sun1,6, Kateryna Tyshchenko1,6, Rachel Wong1, Aastha Nanda1, Claire Shanna1, Samuel Gusscott1, Elizabeth A. Chavez2, Alireza Lorzadeh3, Alice Zhu3, Ainsleigh Hill1, Stacy Hung2, Scott Brown4, Artem Babaian1, Xuehai Wang1, Robert A. Holt4, Christian Steidl2, Aly Karsan4, R. Keith Humphries1, Connie J. Eaves1,1, Martin Hirst3,4 & Andrew P. Weng1

Mechanistic studies in human cancer have relied heavily on cell lines and mouse models, but are limited by in vitro adaptation and species context issues, respectively. More recent efforts have utilized patient-derived xenografts; however, these are hampered by variable genetic background, inability to study early events, and practical issues with availability/reproducibility. We report here an efficient, reproducible model of T-cell leukemia in which lentiviral transduction of normal human cord blood yields aggressive leukemia that appears indistinguishable from natural disease. We utilize this synthetic model to uncover a role for oncogene-induced HOXB activation which is operative in leukemia cells-of-origin and persists in established tumors where it defines a novel subset of patients distinct from other known genetic subtypes and with poor clinical outcome. We show further that anterior HOXB genes are specifically activated in human T-ALL by an epigenetic mechanism and confer growth advantage in both pre-leukemia cells and established clones.

1 Terry Fox Laboratory, BC Cancer Agency, Vancouver V5Z 1L3 BC, Canada. 2 Lymphoid Cancer Research, BC Cancer Agency, Vancouver V5Z 1L3 BC, Canada. 3 Department of Microbiology and Immunology, Michael Smith Laboratories, University of British Columbia, Vancouver V6T 1Z4 BC, Canada. 4 Canada’s Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver V5Z 1L3 BC, Canada. 5 Present address: Department of Hematology, Faculty of Medicine, University of Tsukuba, Tsukuba 305-8575, Japan. 6 These authors contributed equally: Ann Chong Sun, Kateryna Tyshchenko. Correspondence and requests for materials should be addressed to A.P.W. (email: aweng@bccrc.ca)

NATURE COMMUNICATIONS | (2019) 10:2913 | https://doi.org/10.1038/s41467-019-10510-8 | www.nature.com/naturecommunications
Despite the many and important conceptual insights mouse models have brought to our understanding of human cancer, they are, by definition, incapable of revealing mechanisms that are dependent on human-specific elements. It is clear there are important differences between mouse and human cells, particularly with respect to cellular transformation. Particularly noteworthy is the observation that 40–90% of transcription factor binding events are not conserved between mouse and human. Importantly, the translational impact of findings derived from cancer models based on transformation of mouse cells is immediately hampered by the need to verify if and to what extent the identified molecular mechanisms remain similarly operative in human cells. In the case of the hematopoietic system, the complement of cell surface markers used to define hematopoietic stem cells in mouse and human are completely different, thus limiting for instance the ability to translate work on leukemia stem cells identified in mouse models to human disease.

To mitigate species-specific limitations, many studies now incorporate validation of findings using patient-derived xenografts (PDX). Short of in-patient clinical trials, PDX models currently represent the closest we can get to bona fide human disease in terms of a platform for functional studies. Of course, established human cell lines have and will continue to provide valuable insights into molecular mechanisms, but suffer the well-recognized caveat of rigorous selection for growth in vitro that can distort, and thus may not be representative of natural biological processes. The extent of genetic variation present in large PDX collections, however, both in terms of the mutational complement in each tumor and the genetic background of each patient, raises daunting challenges to understanding the mechanistic contribution of individual genetic elements and how they manifest on varied genetic/mutational backgrounds. Finally, neither established cell lines nor PDX models are able to functionally interrogate the earliest of molecular events as oncogenes redirect cells from normal to malignant developmental trajectories.

We thus sought here to take a synthetic approach and create custom-designed tumors using prespecified combinations of genetic elements. We opted for normal human cells as starting material in order to study the process of malignant transformation from beginning to end, and used multipotent hematopoietic progenitor/stem cells from umbilical cord blood (CB) as they are a consistent and renewable resource. We attempted to create synthetic T-cell acute lymphoblastic leukemia (T-ALL) as the genetics have been well described by landscape sequencing, with several oncogenes and tumor suppressors validated by transgenic mouse studies. The major genetic classes of T-ALL involve TLX1/3, TAL1/SCL, LMO1/2, LY11, CALM-AF10, SET-NUP214, and NOTCH1 as defined by chromosomal translocation or over/ contextually inappropriate expression. Importantly, tumor suppressors p16INK4a and p14ARF are deleted/silenced in over 80% of cases and thus represents a near-requisite event for T-ALL establishment. We delivered specified combinations of these various oncogenes into CD34+ CB cells by lentiviral transduction, followed by culture in vitro on OP9-DL1 feeders to examine molecular events occurring in the initial stages of malignant reprogramming of normal T-cell progenitors, and then injection into immunodeficient mice to score for leukemogenesis in vivo.

**Results**

**Transduced oncogenes drive expansion of CB cells in vitro.** We sought here to create human T-ALL de novo from normal CD34+ CB progenitors by lentiviral transduction with a combination of known T-ALL oncogenes. We combined activated NOTCH1 (NOTCH1AE) with LMO2/TAL1, LY11, TLX1, TLX3, HOXA9, MEF2C, and NKK2.1, which were marked with GFP and Cherry fluorescent reporters, respectively (Fig. 1a). We included BMI1 with each of the Cherry vectors on the premise its transcriptional repression/silencing of CDKN2A, which encodes both p16INK4a and p14ARF, would be critical for T-ALL establishment. BMI1 has also been identified as essential for self-renewal of hematopoietic, neural, and intestinal stem cells. Coexpression of multiple genes from a single lentivirus was accomplished by linking cDNAs with picornaviral 2A sequences. Transduced cells were passaged on OP9-DL1 stromal feeders every 4–5 days to study their behavior in vitro. DL1-expressing feeders were utilized so that nontransduced control cells would undergo early T-cell differentiation and thus serve as a close comparator for effects of the delivered oncogenes. Importantly, cells cultured in this manner maintain the ability to engraft live animals and contribute to immune reconstitution. Strikingly, doubly transduced GFP+ Cherry+ (hereafter referred to as G+C+) cells progressively outcompeted singly- and nontransduced populations in vitro for six of seven assessed gene combinations, comprising the majority of cells within 30–50 days (Fig. 1b, c). Of note, G−C− cells expanded for the first few weeks but were outcompeted thereafter by G+C+ cells (Fig. 1c). Further, in cultures transduced with NOTCH1-only, G+ cells expanded initially, but regressed somewhat and never grew to exceed G− cells (Supplementary Fig. 1). In the NOTCH1AE (N) + LMO2/TAL1/BMI1 (LTB) gene combination, G+C+ cells typically attained ~106-fold expansion by day 50 (Fig. 1d) and exhibited an immature CD34+/− CD38+ CD7+ CD1a− CD2− sCD3− T-cell phenotype whereas nontransduced cells in the same conditions differentiated further to a CD34− CD1a+ stage (Fig. 1e, Supplementary Fig. 2).

**Transduced CB cells produce lethal T-cell leukemias in vivo.** To score for leukemia-initiating activity in vivo, transduced CB cells cultured up to 25 days in vitro on OP9-DL1 feeders were injected into NSG mice. In initial protocols, human CD45+ cells were FACS sorted from day 10 cultures and injected intrahepatically into sublethally irradiated neonatal recipients. Of note, the injected hCD45+ cells included a mixture of nontransduced (G−C−), singly transduced (G+C− and G− C+), and doubly transduced (G+C+) populations (Fig. 1c). Subsequent protocols involved sorting of doubly transduced CB cells (hCD45+ G+C+) from day 24–25 cultures and intravenous injection into adult recipients. As our data are most mature for the N+ LTBI gene combination, we will focus here on those results.

We obtained malignant leukemias with T-ALL-like features in 36/43 primary recipients from seven different N+ LTBI transduction experiments with overall median latency of 161 days (range 79–321 days) (Fig. 2a, Supplementary Data 1). Clinically morbid animals typically exhibited hepatosplenomegaly, lymph node and thymic masses, hypercellular bone marrow with extensive infiltration by leukemic blasts, and circulating leukemia cells with immature blast-like cytology (Fig. 2b). Tumors also exhibited clonal TCRG rearrangements as assessed by clinical BIOMED-2 assay (Fig. 2c).

In 22/24 recipients injected with hCD45+ G+C+ cells (FACS sorted from day 24–25 N+ LTBI transduced CB cultures), we obtained G+C+ leukemias of T-cell lineage, typically CD7+ CD2+ sCD3+−/− CD1a+−/− and variable CD4/CD8 pattern including CD4− CD8− (DN), CD4+ CD8+ (DP), and CD4− CD8dim (SP8dim) (Supplementary Fig. 3, Supplementary Data 1). Among 19 recipients of hCD45+ cells (FACS sorted from day 10–11 N+ LTBI-transduced CB cultures), seven mice developed G+C+ leukemias, seven developed G+C− leukemias, and one showed persistent low-level G+C+ engraftment (1–2% in PB) (Supplementary Data 1). Whereas G+C− leukemias demonstrated a spectrum of CD4/CD8 phenotypes, G+C− leukemias were...
mostly DP, but did include one DN case (6 DP + 1 DN in total; Supplementary Fig. 3, Supplementary Data 1). Given the ~6-fold greater abundance of G+C− over G+C+ cells in day 10–11 hCD45+ inocula (Fig. 1c, Supplementary Data 1), these findings suggest there is selection for the full NLTB (G+C+) oncogenic payload in vivo similar to that observed in vitro, although N alone (G+C−) is also capable of producing aggressive leukemia in primary recipients.

CB leukemias with both N and LTB are readily transplantable. We next tested primary leukemias for their ability to produce disease in secondary recipients. We tested five different primary G+C+ leukemias and one low-level engrafted, but nonlethal case, and found all six to produce lethal G+C+ leukemias in secondary recipients (Fig. 2d, Supplementary Data 2). We also tested four different primary G+C− leukemias and sorted G+C− cells from a primary recipient harboring both G+C+ and G+C− subpopulations, and found only two of the five to produce lethal G+C− leukemias in secondary recipients (Fig. 2e, Supplementary Data 2). Thus, whereas the combination of N + LTB consistently yielded fully transformed, serially transplantable leukemias, N alone was less efficient in doing so, yielding lethal, yet non-self-renewing lymphoid expansions in about half of instances (Fig. 2e, Supplementary Data 2). We also performed limiting dilution transplants to compare leukemia-initiating cell (LIC) frequencies in transplantable G+C+ vs. non-transplantable G+C− leukemias. Whereas the G+C+ primary leukemia from mouse #63 exhibited an LIC frequency of less than 1 in 100 cells, the G+C− primary leukemia from mouse #63 showed an LIC frequency of less than 1 in 1,200,000 cells, for a difference of at least ~300-fold (Supplementary Data 2). We also performed limiting dilution transplants to compare leukemia-initiating cell (LIC) frequencies in transplantable G+C+ vs. non-transplantable G+C− leukemias. Whereas the G+C+ primary leukemia from mouse #63 exhibited an LIC frequency of less than 1 in 100 cells, the G+C− primary leukemia from mouse #63 showed an LIC frequency of less than 1 in 1,200,000 cells, for a difference of at least ~300-fold (Supplementary Data 2). These data reveal that while both N + LTB and N-only can produce serially transplantable (i.e., self renewing) leukemias, this process occurs with much lower efficiency when only N is provided, presumably due to the need for stochastic acquisition of additional oncogenic hits.

Transduced CB cells show no evidence of clonality in vitro. To ascertain what stage of leukemogenic transformation the
transduced CB cells had attained in vitro prior to transplantation, we applied the clinical BIOMED-2 assay to detect and track clonal TCRG rearrangements. TCRG profiles of in vitro-expanded cells showed no convincing evidence of a dominant, clonally rearranged population in any of the four different CB trials (Fig. 3a). Further, comparison of TCRG profiles from in vitro-expanded CB cells and the leukemias they produced after transplantation revealed no clear evidence for perdurance of a dominant clone (Fig. 3b, Supplementary Fig. 5). As well, leukemias arising in animals transplanted with the same pool of transduced CB cells exhibited distinct clonal TCRG rearrangements (Figs. 2c, 3b; Supplementary Fig. 5) and also distinct donor

STR patterns (Fig. 3c, Supplementary Fig. 6), suggesting that dominant, clonally rearranged leukemias had not already arisen in vitro prior to transplantation.

We also submitted gDNA for commercial ImmunoSEQ TCRG assay from transduced CB cells that had been expanded in vitro for 14–38 days. Analysis of CDR3 fragment lengths revealed normal Gaussian distributions for both transduced G+C+ cells and G−C− controls (Supplementary Fig. 7). Read counts were about 5–6× higher for G−C− samples, consistent with a higher proportion of cells progressing through early T-cell development. Tracking of individual rearrangements showed no evidence of dominant clones emerging over time in culture among G+C+...
cells, with the clone distribution pattern appearing highly similar to that of control $G−C−$ cells (Fig. 3d). Taken together, these results support the interpretation that NLTB-transduced CB cells do not progress to the point of dominant clonal populations for up to 38 days in culture, and thus do not represent clonal leukemias prior to transplantation.

**CB leukemia can be generated from multiple different donors.** For most of these experiments we used pooled CB cells from hundreds of donors. To determine from how many donors we performed using CB cells from a single donor (Supplementary Data 1). In total, we were able to generate leukemias from at least eight different individual CB donors, supporting that the ability of N + LTB to transform CB cells is not limited to rare individuals in the population.

**CB leukemias show clonal evolution including NRAS mutation.** We performed whole exome sequencing on a set of three lethal $G+C+$ leukemias from secondary recipient mice which had all been injected with cells from a primary recipient that showed persistent low-level engraftment by $G+C+$ cells, but was otherwise healthy. Analysis of single nucleotide variants (SNVs) confirmed the three leukemias had derived from a common ancestral clone as evidenced by 542 common SNVs. Interestingly, 2/3 were highly similar to one another, but the third showed substantial divergence with 3063 private SNVs (Fig. 4a). One of the private SNVs in this third leukemia was an NRAS gain-of-function mutation, which occurs commonly in human T-ALL/7.8.19, and thus these findings would suggest that clonal...
evolution similar to that which occurs in natural disease is operative in this synthetic model.

RNA-seq reveals CB leukemias are highly similar to PDXs. To determine how similar/dissimilar the synthetic leukemias were compared to bona fide human T-ALL, we performed RNA-seq on a set of 17 CB leukemias and compared them to a collection of 22 different PDX T-ALLs. Using the top 1000 variable genes among PDX samples and unsupervised hierarchical clustering (Fig. 4b), we observed that G+C+ leukemias were highly similar to one another, even across different experimental trials and originating from different donors, and that G+C− leukemias clustered separately, indicating that inclusion of LTB alters the transcriptional signature. Importantly, correlation distances between CB leukemias and individual PDXs fell within the range of distances between individual PDXs, revealing that CB leukemias reside within the spectrum of natural T-ALL variation and do not appear instead as distant outliers. We would interpret these data to support that synthetic leukemias are highly reproducible and
represent a reasonable approximation of naturally occurring human T-ALL.

We also performed unsupervised hierarchical clustering using the top variable genes among just NLTB G+C+ leukemias to determine how they differed from one another (Supplementary Fig. 8). The most notable feature was that leukemias clustered according to donor as defined by STR profiling. Since each donor leukemia represents by definition a distinct cell-transduction event, we would conclude that the source of greatest variation among synthetic NLTB leukemias is genetic/epigenetic background of the donor and/or viral integration effects.

We also extracted CDR3/V(D)J junctional reads for each of the TCR loci from RNA-seq data using the MiXCR software package.20 We found CB leukemias to show mostly mono/oligoclonal TCR rearrangements with G+C+ leukemias tending to express rearranged TRCg and TCRD when of CD4− CD8− (DN) phenotype, and rearranged TCRB when of CD4+ CD8+ (DP) phenotype. G+C− leukemias, which were typically DP phenotype, tended to express rearranged TCRB (Fig. 4c). Together with immunophenotyping data, these findings suggest that while NLTB (G+C+) leukemias likely span both pre- and post-β-selection stages of T-cell development, N-only (G+C−) leukemias are most often post-β-selection.21

HOXB genes are upregulated in nascently transduced CB cells. The in vitro component of the CB model allows direct access to transduced cells as they undergo the first very molecular changes as they are redirected from normal to malignant developmental trajectories. Accordingly, we harvested cultures at various time points and FACS-sorted singly and doubly transduced subsets as well as nontransduced control cells for RNA-seq. We focused on genes differentially expressed between NLTB doubly (G+C+) and nontransduced (G−C−) cells from the earliest sets of cultures (days 14 and 24), and identified 468 differentially expressed genes (Fig. 5a, Supplementary Data 3). We performed Reactome pathway analysis22 and found pathways relating to HOX genes in signaling and corroborating our findings that growth in vivo selects for further enhancement of Notch/RAS/RAF/MAPK signaling (Supplementary Table 4), suggesting that cytokines/growth factors may be limiting in vivo, or perhaps that a substantial proportion of cells in vivo reside in less replete microenvironments. Alternatively, fully evolved leukemias may include subsets of cells that are less receptive to signaling agonists.

**NLTB induces altered epigenetic patterning over HOXB genes.** The contiguous distribution of upregulated genes within the anterior (3′) HOXB locus led us to wonder if altered epigenetic patterning might underlie the gene expression changes. Indeed, ChIP-seq analysis from the same set of samples showed significant H3K27me3 loss in G+C+ cells as compared to G−C− control cells at two regions, one near the 3′ end of HOXB2 and the other spanning from upstream of HOXB3/B4 to downstream of HOXB6 (regions 1 and 2, respectively, in Fig. 5c). Region 2 was particularly notable as it showed a corresponding enrichment of H3K27ac marks in G+C+ over G−C− cells (Fig. 5d). Taken together, these results reveal that NLTB-transduced cells exhibit an altered chromatin pattern consistent with gene activation over the HOXB locus (corresponding RNA-seq tracks in Supplementary Fig. 10), and suggest that NLTB may initiate the leukemogenic process by remodeling of chromatin to achieve coordinate regulation of multiple genes required for cellular transformation.

**High HOXB is associated with poor clinical outcome in T-ALL.** To begin to address what role HOXB genes may play in natural/spontaneous human T-ALL, we examined RNA-seq data from a large pediatric cohort comprising 264 diagnostic T-ALL samples (COG TARGET study)12. HOXB2, B3, and B4 were expressed in a subset of cases, and in a notably coordinated fashion (Fig. 6a), corroborating mRNA expression data from transduced CB cells (Fig. 5b). HOXB5 was expressed in only a minority of cases, and was expressed at lower levels as compared to HOXB2/B3/B4 in transduced CB cells. Strikingly, cases with higher levels of HOXB2-B5 mRNA expression exhibited significantly poorer event-free survival (Fig. 6b, Supplementary Table 5). HOXB4 alone was also significant (Fig. 6c), while HOXB2 and HOXB3 were not significant (Fig. 6d, e). These data support that HOXB genes are coordinately expressed in primary human T-ALL and higher levels of expression are associated with more aggressive clinical disease.

**High HOXB defines a distinct disease subgroup in T-ALL.** The poor clinical outcomes of patients with high HOXB gene expression led us to wonder if they might overlap with cases of ETP-ALL23. We found only limited overlap between these two groups by PCA (Fig. 7a) and no positive statistical association, but instead a trend towards negative association (Fig. 7b). We also examined whether high HOXB cases were enriched/depleted within any particular transcription factor/translocation-defined subgroup.24 Interestingly, high HOXB cases were significantly associated with TAL1, NKK2-1, and so-called “unknown” subgroups, while low HOXB cases were significantly associated with TLX1 and TLX3 subgroups (Fig. 7c). Accordingly, HOXB2/B3/B4 gene expression levels were consistently elevated in TAL1+, NKK2−1, and “unknown” subgroups, and consistently decreased in TLX1 and TLX3 subgroups (Supplementary Fig. 11). Of note, cases in the top quintile for expression of TAL1 show higher expression of HOXB2/B3/B4, although this difference was statistically significant only for HOXB4. There was no evidence for increased HOXB gene expression in the top quintile of LMO2 expressers.
Fig. 5 NLTB induces anterior HOXB gene expression with altered epigenetic patterning. a, b RNA-seq analysis of nascently transduced CB cells. CB cells were transduced with N-GFP and LT-B-Cherry lentiviruses, then cultured on OP9-DL1 feeders for 14–24 days. RNA was prepared from FACS-sorted doubly (G+C+), singly (G+C−), and nontransduced (G−C−) subsets. Sorted G−C− cells did not pass QC and are not shown. Samples were collected from three different experimental trials (t28–t30). a mRNA expression heatmap. Differentially expressed genes (G+C+ vs. G−C−; log2FC > 1, p-adj < 0.1) are depicted (468 total, 243 up/225 down). Heatmap is scaled by gene (row) with mean = 0 and SD = 1. b mRNA expression level of HOXB and flanking genes. Rlog values from DESeq2 are plotted. Each datapoint represents an individual sample. Bars indicate mean ± SD for each set of colored dots. Statistical comparisons are for doubly vs. nontransduced cells. *p < 0.05; **p < 0.01; ***p < 0.001; ns not significant (two-tailed t test with Holm–Sidak correction for multiple comparisons). c, d Histone ChIP-seq tracks for c H3K27me3 and d H3K27ac over the HOXB locus in NLTB (G+C+) and nontransduced (G−C−) CB cells cultured for 14, 24, or 47 days on OP9-DL1 feeders. Subsets were FACS sorted prior to fixation for ChIP. Corresponding MACS2 peak calls are shown below each track.

( Supplementary Fig. 12). Additional studies will be needed to assess what genetic contexts are most permissive to HOXB gene upregulation.

To explore further what features might characterize patient T-ALL cases with high vs. low HOXB gene expression, we performed gene expression enrichment analysis (GSEA) using gene signatures derived from an RNA-seq dataset that included normal hematopoietic progenitors from human bone marrow and thymus24. PCA suggested three groupings of normal cells: (1) normal hematopoietic progenitors from human bone marrow and thymus24. Region 1 and thymus24. Region 2. a mRNA expression heatmap. Differentially expressed genes (G+C+ vs. G−C−; log2FC > 1, p-adj < 0.1) are depicted (468 total, 243 up/225 down). Heatmap is scaled by gene (row) with mean = 0 and SD = 1. b mRNA expression level of HOXB and flanking genes. Rlog values from DESeq2 are plotted. Each datapoint represents an individual sample. Bars indicate mean ± SD for each set of colored dots. Statistical comparisons are for doubly vs. nontransduced cells. *p < 0.05; **p < 0.01; ***p < 0.001; ns not significant (two-tailed t test with Holm–Sidak correction for multiple comparisons). c, d Histone ChIP-seq tracks for c H3K27me3 and d H3K27ac over the HOXB locus in NLTB (G+C+) and nontransduced (G−C−) CB cells cultured for 14, 24, or 47 days on OP9-DL1 feeders. Subsets were FACS sorted prior to fixation for ChIP. Corresponding MACS2 peak calls are shown below each track.

**HOXB3 promotes cell growth in pre- and established leukemia.**

Following on evidence that HOXB gene upregulation occurred as an early event in NLTB-transduced CB cells, we examined RNA-seq data from established CB leukemias and a collection of T-ALL PDXs and found HOXB2/B3/B4 mRNA levels were coordinately elevated in both contexts (Fig. 8a). We also looked at HOXB gene expression in other acute leukemias and found that both T-ALL and B-ALL PDXs were characterized by specific elevation of anterior HOXB genes, while AML PDXs showed a pattern of HOXB gene expression that extended more broadly across the locus (Supplementary Fig. 13). Interestingly, levels of HOXB2/3/4 were significantly higher in T-ALL than in B-ALL, suggesting perhaps a particular relevance of anterior HOXB gene action in T-ALL.

To assess formally whether these HOXB genes play a functional role in established leukemia, we performed a limited scale, pooled shRNA dropout screen using primary CB leukemia cells cultured in vitro. The shRNA pool included a total of 56 shRNAs targeting cases is defined by a biologic process orthologous to conventional transcription factor groupings.
Multiple HOXB (2, 3, 4, 5) and HOXA (5, 7, 9, 10) genes (5–10 shRNAs per gene) plus three non-targeting controls (Supplementary Table 6 and Supplementary Fig. 14). We included four HOXA genes in the screen as they were significantly downregulated in G+C+ vs. G−C− cells (Supplementary Data 3), raising the possibility they may antagonize T-ALL growth. We found significant depletion of four different shRNAs against HOXB3 and two against HOXB5 (Fig. 8b; Supplementary Fig. 15), suggesting their knockdown had resulted in growth disadvantage. We also performed the same screen using established T-ALL cell lines HSB2 and PEER and found consistent depletion of shRNAs against HOXB3, but less consistent depletion of shRNAs against HOXB5 (Supplementary Fig. 16).

We focused further efforts on HOXB3 since HOXB5 expression was detectable in only a small number of primary T-ALLs (Fig. 6a). We confirmed accelerated depletion of three different HOXB3 shRNAs as compared to nonsilencing controls in singleton transduction experiments by flow cytometry using NGFR-tagged viral shRNA constructs in the same primary NLTB leukemia as used for the 59-plex shRNA screen, and also in a second, independent primary NLTB leukemia (Fig. 8c). Similar results were obtained using three different established human T-ALL cell lines (Fig. 8d) selected for study as they expressed relatively high levels of HOXB3 in combination with TAL1 and/or LMO2 (Supplementary Fig. 17). These findings are consistent with the notion that HOXB3 contributes to maintenance of established leukemias.

To address whether HOXB3 also played a role in the early stages of leukemia initiation, we went back to normal CB cells nascently transduced with NLTB to test the effect of HOXB3 shRNAs on clonogenic expansion. We employed a variation on conventional methylcellulose colony forming cell (CFC) assays, but compatible with OP9-DL1 co-cultures in which transduced CB cells were sorted at limiting dilution into individual wells of a 96-well plate containing OP9-DL1 feeders and then assayed ~3 weeks later by flow cytometry. Net yields per well of viable, triply transduced hCD45+ cells (Supplementary Data 3), shHOXB3_644 down to ten input cells per well and for shHOXB3_643 down to 50 input cells per well as compared to shScr control (Fig. 8e). If we analyze the data according to a single-hit model/Poisson distribution where a positive well is defined as containing at least 500 triply transduced hCD45+ cells, we are able to calculate the frequency of well-initiating cells, or WIC. By this approach, we find the WIC frequency for control shRNA-transduced cells to be ~1/41 cells (95% CI 1/30–56 cells), while shHOXB3-transduced cells showed statistically significant, 6- to 12-fold lower WIC frequencies of ~1/240 cells (95% CI 1/140–430) for shHOXB3_643 and ~1/510 cells (95% CI 1/230–1100 cells) for shHOXB3_644 (p = 6.9e-10 and 1.6e-13, respectively; chi-square test) (Fig. 8f). Similar, statistically significant differences were also observed using alternate cell yield cutoff values as low as 200. Importantly, the goodness of fit test for shScr cells supports that the null hypothesis/single-hit model is not rejected (p = 0.78; chi-square test). Taken together, these results support the notion that NLTB-induced HOXB3 gene expression is required for clonogenic expansion of T-progenitor
cells undergoing the earliest stages of malignant transformation, and that this dependency persists in established leukemia cells.

**Discussion**

The synthetic leukemia model described here provides an efficient and reproducible means for generating human T-ALL that appears indistinguishable from spontaneously arising patient tumors. In addition to overcoming limitations associated with models involving transformation of mouse cells, it allows custom design/specification of the tumor’s genetic composition, thus facilitating creation of near isogenic sets of tumor samples necessary to deconvolute the contributions of individual genetic elements to tumor phenotypes. This approach holds several advantages over deploying large collections of PDXs which contain a wide assortment of genetic variants that can obscure discovery/validation of bona fide genetic associations, and are logistically difficult to generate and share. As well, all tumors continue to evolve during serial propagation in vivo and obtaining early passage vials from PDX lines can be difficult if not impossible. In contrast, synthetic leukemias can be generated again and again from normal cells, thereby resetting the cancer evolutionary clock.

The ability to achieve multilog expansion of transduced CB cells in vitro afforded us the unique opportunity to explore the earliest of biochemical events that occur as cells were redirected from normal cells, thereby resetting the cancer evolutionary clock.

The synthetic leukemia model described here provides an efficient and reproducible means for generating human T-ALL that appears indistinguishable from spontaneously arising patient tumors. In addition to overcoming limitations associated with models involving transformation of mouse cells, it allows custom design/specification of the tumor’s genetic composition, thus facilitating creation of near isogenic sets of tumor samples necessary to deconvolute the contributions of individual genetic elements to tumor phenotypes. This approach holds several advantages over deploying large collections of PDXs which contain a wide assortment of genetic variants that can obscure discovery/validation of bona fide genetic associations, and are logistically difficult to generate and share. As well, all tumors continue to evolve during serial propagation in vivo and obtaining early passage vials from PDX lines can be difficult if not impossible. In contrast, synthetic leukemias can be generated again and again from normal cells, thereby resetting the cancer evolutionary clock.

The ability to achieve multilog expansion of transduced CB cells in vitro afforded us the unique opportunity to explore the earliest of biochemical events that occur as cells were redirected from normal cells, thereby resetting the cancer evolutionary clock.
We had initially thought HOXB high cases might exhibit an HSC-like signature; however, our results suggest instead that these cases are more similar to late thymic progenitors. In combination with our functional data showing that HOXB3 and perhaps also HOXB5 support net cell growth in this context, we would posit that HOXB genes may be more accurately characterized as promoting expansion of lineage-restricted but immature progenitors, rather than the more prevalent view that they act to expand existing multipotent stem cells\textsuperscript{33,34}, an idea that has been proposed recently by others as well\textsuperscript{35}.

Exploration of the earliest events in cellular transformation may also favor identification of mechanisms operative in leukemia stem cells or that are shared by all cells in the tumor as opposed to later events that variably accumulate within individual subclones. Others have used variant allele frequencies to provide insight into the subclonal structure of tumors; however, this approach cannot inform on the cellular contexts in which variants occur or their associated phenotypes in isolation from subsequent mutational events. The synthetic approach affords an experimentally tractable way to model both the complement of genetic alterations and the cellular contexts in which they may occur.

Synthetic modeling offers a straightforward approach to deconvoluting the contributions of individual genetic
components in the leukemogenic process. For instance, we delivered NOTCH1ΔE and LTB on two separate viruses and thus could readily discern from in vitro cultures that NOTCH1ΔE provided an initial growth advantage, whereas LTB caused developmental arrest at an early stage with no apparent growth advantage. Further, while both NLTB and N-only produced aggressive leukemia in primary recipients, NLTB leukemias were consistently transplantable whereas most N-only leukemias were not transplantable. We interpret this to indicate that NOTCH1ΔE provides for prodigious expansion of immature T cells (enough even to overwhelm the primary recipient mouse), but in itself does not confer self-renewal properties required for serial transplantability. It is notable that in transplants using unsorted cells, G−C−/N-only cells always outnumbered G−C+/NLTB cells by ~6-fold on average, and yet G−C+/NLTB cells prevailed in over half of resulting leukemias. This supports the notion that LTB contributes positively to the leukemogenic process, perhaps even specifically to self-renewal. As further studies will be needed to identify what additional hits may be consistently present, we regard N-only not as sufficient for leukemogenesis, but it is sufficiently enabling to yield transplantable disease with high penetrance and reasonable latency. Of note, a similar CB T-ALL syndrome is to be examined. Observations made in vitro will of course be translated into in vivo and in vivo assays that are not feasible otherwise, and thus represents a crucial advantage. Further, while both NLTB and N-only produced an initial growth advantage, whereas LTB caused a marked organ toxicity. Another potentially limiting aspect is the in vitro culture phase prior to transplantation which may select for features dissimilar to propagation in vivo; however, the high level of similarity by RNA-seq between PDX samples and synthetic CB leukemias generated by transplantation after 10–24 days on OP9-DL1 feeders (Fig. 4b) suggests that untoward effects of culture in vitro may be limited. In counterbalance, the in vitro culture phase enables access to cells for biochemical and phenotypic assays that are not feasible otherwise, and thus represents a necessary evil if the earliest of events in cellular transformation are to be examined. Observations made in vitro will of course need to be verified using exclusively in vivo models and in primary patient samples.

Methods

Isolation of human hematopoietic stem/progenitor cells. Anonymized normal human cord blood (CB) samples were obtained with informed consent from women undergoing caesarean deliveries of full-term births according to protocols approved by the Research Ethics Board of the University of British Columbia and Children’s & Women’s Hospital of BC. CD34+ CB cells were obtained at >95% purity from pooled collections using a two-step Rosette-Sep/EasySep human CD34-positive selection kit (StemCell Technologies) according to the manufacturer’s protocols and/or FACS sorting. The purity of FACS-sorted cells was >99% as determined by post-sort analysis. CD34+ cells were seeded into 96-well round bottom plates and prestimulated in StemSpan SFEM II (StemCell Technologies) with 100 ng/ml − human SCF, 100 ng/ml − human FLT3L, 50 ng/ml − human TPO, and 20 μg/ml − human LDL, or 10 ng/ml − human SCF, 20 ng/ml − human TPO, 20 ng/ml − human IGF2, and 10 ng/ml − human FGFα (Peprotech) for 16 h.

Lentiviral constructs and transduction. Human NOTCH1 (ΔE allele), TAL1, BMI1, and mouse LMO2 cDNAs obtained from Dr. J Auster (Boston), Harvard Plasmid, and Dr. E. Lavlor (UCLA). NOTCH1ΔE and GFP cDNAs were connected with equine rhinitis A virus 2A (E2A) peptide14, LMO2, TAL1, BMI1 and Cherry cDNAs were similarly connected with those asa virus 2A (T2A), foot-and-mouth disease virus 2A (F2A), and E2A peptides, respectively. These polycistronic cDNAs were cloned into pRRL-α CTS-MNDU3-PGK-GFP-WPRE immediately downstream of the MNDU3 promoter. All constructs were verified by sequencing. Additional vector construction details are available upon request.

Higher tier viral transfetions were produced by transient transfection of 293T cells using polyethylenimine HCl MAX (Polyscience) with second-generation packaging/envelope vectors pCMV dR8.74 (Addgene #22036), pRSV-Rev (Addgene #112533), and pCMV VSV-G (Addgene #8454), followed by ultracentrifuged concentration (25,000 rpm for 90 min at 4 °C) Beckman SW32Ti rotor.

CB cells were transduced in 96-well plates coated with 5 μg/ml fibronectin (StemCell Technologies) by direct addition of concentrated viral supernatants and transferred to OP9-DL1 co-cultures 6 h later. CB transduction efficiencies are shown in Fig. 1b, c, and Supplementary Data 1. CB leukemia and cell lines were transduced by spinoculation (1800 x for 2 h at 37 °C) with viral supernatants in 4 μg/ml − polybrene. Primary CB leukemia transduction efficiencies were 15–30% for 105_LEP and 10–15% for m160. Cell line transduction efficiencies were 20–35% for HS2, 20–30% for PEER, and 55–60% for PF382.

Cell culture. Transduced CB cells and explanted primary CB leukemia cells were cultured on top of confluent monolayers of OP9-DL1 cells in αMEM media supplemented with 20% FBS (Invitrogen), 15% human SCF, 5 μg/ml − 1 IL-7 and 3 ng/ml − 1 IL-7 (Peprotech). Human T-ALL cell lines were cultured in RPMI 1640 media supplemented with 10% FCS, with 1 μM sodium pyruvate, 2 μM Gluta-MAX ( Gibco), and antibiotics.

Cell lines. OP9-DL1 cells were obtained from J.C. Zuniga-Pflucker (University of Toronto), 293T, HS2B, PEER, and PF382 cells were obtained from J. Aster (Brigham & Women’s Hospital, Boston). Cell line authentication by PowerPlex 16HS multiplex STR DNA profiling (Promega) was performed by Genetica DNA Laboratories (Burlington, NC). Cultures were confirmed mycoplasma-free and regular surveillance testing performed using the MycoAlert mycoplasma detection kit (Lonza).

Patient-derived xenograft samples. PDX samples from the ProXe repository6 are identified by TA prefix (n = 17). PDX samples established in the Weng lab include H3255-4, M22-7, M669-26, M71-1-2, and M30-2v2, and have been reported previously9,10.

Mice. NSG (NOD.Cg-Pkdcreid122gm1m19/Tszd) mice were bred and housed in a specific pathogen-free animal facility at the British Columbia Cancer Research Centre. All experimental procedures were approved by the University of British Columbia Animal Care Committee.

Transplantation by intrahepatic/intravenous injection. Neonatal NSG mice (4–10 days of age) were sublethally irradiated (100 cGy X-ray at 150 cGy min −1), then injected intrahepatically with ~0.1 to 1.0 × 106 sorted CB cells (hCD45+ or hCD45− GEP + Cherry+) mixed with rhIL-7 (0.5 μg per mouse; Peprotech) and anti-IL-7 mAb (2.5 μg per mouse; clone M25; Bio X Cell, West Lebanon, NH) in PBS (total volume 30 μL). Mice were boosted with IL-7/−7−7 mAb cocktail by IP injection every 4–5 days for the first 28 days post transplantation. Adult NSG recipients were sublethally irradiated (200 cGy X-ray at 150 cGy min −1) prior to intrahepatic injection of transduced CB cells (primary recipients) or CB leukemias (secondary recipients). Transplanted animals were monitored by monthly peripheral blood (PB) sampling and sacrificed at predefined, humane clinical morbidity endpoints.

Histology. Mouse tissues (spleen, thymus, lymph node, sternum) were fixed in 10% neutral-buffered formalin for 48 h, then stored in 70% ethanol before paraffin embedding. Hematoxylin and eosin staining was performed on 4 μm paraffin sections.

Flow cytometry. Absolute cell counts were obtained in flow data using AccuCount counting beads (Invitrogen). Live/dead cell gating was performed by staining with propidium iodide or DAPI (Invitrogen). We used anti-hCD271 (Miltenyi Biotec, Biologend) to detect the lentiviral NGFR marker. We performed flow cytometric analyses on FACSCalibur and LSRFortessa instruments and sorting on FACSAria3 and Fusion instruments (BD Biosciences). We analyzed flow cytometry data using FlowJo software (TreeStar). Example gating strategies are provided in Supplementary Figs. 18 and 19.

Western blot. Whole-cell lysates were separated by SDS-PAGE, transferred to Hybond-ECL membranes (Amersham) and blocked with 5% nonfat dry milk. Membranes were probed with primary antibodies the FLAG epitope (M2 clone; Sigma) or β-actin (AC-15; Sigma Aldrich), then with HRP-conjugated secondary antibodies (Jackson ImmunoLabs) and detected with enhanced chemiluminescence (ECL; Pierce). Band intensities were quantified with Image Studio Lite (LI-COR) software.
BIOMED-2 TCRG assay. The BIOMED-2 TGR gene cloning assay for ABI fluorescence detection (InVitroscribe) was performed by the BC Cancer Agency Cancer Genetics Lab.

ImmunoSEQ assay. Genomic DNA was prepared from FACS-selected cells using Qiagen AllPrep DNA/RNA micro kit according to the manufacturer's instructions. Samples were quantified using Nanodrop, diluted for library preparation in buffer EB, and submitted to Adaptive Biotechnologies (Seattle, WA) for immunoSEQ TCRG assay (survey resolution). Briefly, somatically rearranged human TCRG CDR3 sequences were amplified from genomic DNA using a two-step, amplification bias-controlled multiplex PCR approach41,42. The first PCR consists of forward and reverse amplification primers specific for every V and J gene segment, and amplifies the hypervariable complementarity-determining region 3 (CDR3) of the immune receptor locus. The second PCR adds a proprietary barcode sequence and Illumina® adapter sequences43. CDR3 libraries were sequenced on an Illumina instrument according to the manufacturer's instructions. Raw sequence reads were demultiplexed according to Adaptive's proprietary barcode sequences. Demultiplexed reads were then further processed to: remove adapter and primer sequences; identify and correct for technical errors introduced through PCR and sequencing; and remove primer dimers and other contaminating sequence. The data are filtered and clustered using both the relative frequency ratio between similar clones and a modified nearest-neighbor algorithm, to merge closely related sequences. The resulting sequences were sufficient to allow annotation of the (V(N) D(N) J(N)) genes constituting each unique CDR3 and the translation of the encoded CDR3 amino acid sequence. V, D and J gene designations were based on annotation in accordance with the IMGT database (www.imgt.org). The set of observed biological TCRG CDR3 sequences were normalized to correct for residual multiplex PCR amplification bias and quantified against a set of synthetic TCRG CDR3 sequence analogs43. Data were analyzed using the immunoSEQ Analyzer toolset.

Whole exome sequencing. Genomic DNA was prepared from FACS-selected cells using Qiagen AllPrep DNA/RNA mini kit according to the manufacturer's instructions. The SureSelectXT Human All Exon V5 predesigned capture library (Agilent #5190-6208) was used along with the SureSelect XT Library Prep Kit ILM (Agilent #5500-0132) in order to generate human exome libraries from gDNA as per the manufacturer's instructions. Libraries were paired-end 125 bp sequenced on an Illumina HiSeq 2500 at seven samples per lane. Reads were aligned to the human reference genome (hg19) using bwa-mem version 0.7.5a with optical and PCR duplicates removed using the Picard tool (http://broadinstitute.github.io/picard/). Somatic SNV/indel variants were identified by VarScan45 and were filtered for a minimum allele frequency of 1% and ten variant reads. Putative germline variants were removed based on a GMAF > 1%. All variants were annotated using SnPeff (version 4.2)46 and filtered for effects predicted to have a moderate impact or high probability at the protein level.

RNA-seq. Total RNA was isolated from live cells with TRIzol reagent followed by purification over PureLink RNA Mini Kits (Invitrogen). RNA-seq was performed using a polyA-enriched or ribosomal RNA-depleted NEBNext rRNA Depletion Kit (New England Biolabs, cat# E6310) strand-specific library construction protocol and paired-end 125 bp or 75 bp sequencing on an Illumina HiSeq 2500 instrument at eight samples per lane. Pair-end reads were trimmed from the 3′ end based on quality score (end min quality level (Phred) = 20) using Partek FastQ (version 6.0.17.8503; Partek Inc, St Louis, MO). Samples were then demultiplexed and human genome reference assembly GRCh37/hg19 using STAR 2.5.2b in Partek Flow software. All further analysis steps were done in RStudio 0.99.903 (R version 3.3.1). Raw gene expression counts were calculated using featureCounts in Rsubread v1.24.2 with GRCh37 Ensemble release 75 as annotation47. Genes with summated counts of 1 or less across all included samples were filtered out. mRNA expression values were then derived after normalization with rlog function in DESeq2 v1.18.1, which log2 transforms and normalizes the data for library size, as well as minimizes the effect of low-expression genes47. Differential gene expression analyses were performed using filtered raw counts with DESeq2 v1.18.1.

To remove batch effects between PDX and synthetic CB leukemia samples, we performed a principal component analysis using Partek Flow software based on multiple genomic and methylation annotations and principal biological covariates50,51. To determine relationships between PDX and synthetic CB leukemia samples, we performed and plotted with pheatmap 1.0.8 (https://CRAN.R-project.org/package=pheatmap). Multiscale bootstrap resampling was performed using pvclust in R (http://stat.sys.kyoto-u.ac.jp/prog/pvclust/).

Native histone ChIP-seq. We performed native chromatin immunoprecipitation (ChIP) using validated antibodies against H3K27me3 and H3K4me3 and constructed ChIP-seq libraries according to established protocols52,53. Libraries were sequenced on an Illumina HiSeq 2500. Raw sequence reads were inspected for quality, sample swap, and reagent contamination using custom in-house scripts. Paired-end 125 bp reads were aligned to human genome build GRCh37-lite using BWA v0.5.5v5.35 and the alignment files were converted to bam format through SAMtools56. Wig files were subsequently converted into big wig format for visualization in UCSC genome browser. Peak calling was performed using MACS2 (Model-Based Analysis for ChIP-seq, v2.1.1)57 by passing the IP bam and the corresponding input bam, with a value cutoff = 0.01 for narrow histone mark (H3K27ac), and 0.1 (in broad mode) for broad histone mark (H3K27me3).

shRNA library construction and transduction. Separate 5 mL bacterial overnight cultures were grown for each of 56 different lentiviral shRNA constructs targeting a total of 47 genes (average seven shRNAs per gene (Supplementary Table 6), which were then pooled for a single plasmid DNA prep (Qiagen). The final plasmid prep was sequenced on an Illumina MiSeq to assess representation of each shRNA species. All input shRNA clones were detected in the final plasmid prep, with ~55% within 4-fold and ~75% within 10-fold of the mean read counts (Supplementary Fig. 14). Lentiviral preps for the 56-plex pooled shRNA library and three negative control shRNAs (shScramble, shLuc, empty vector) were done separately and then mixed prior to transducing cells. Cells were transduced at a target multiplicity of infection of 0.3 to favor single lentiviral integration in the majority of transduced cells.

shRNA knockdown growth screen. The lentiviral shRNA vector, plKO contains a puromycin selection cassette; however, we did not apply puromycin selection after transduction in order to minimize the time duration between initial shRNA transduction and collection of the t0 time point sample. Since cultures thus contained both transduced and nontransduced cells, an excess of cells was carried in culture for the duration of the screen in order to maintain at least ~60,000 transduced cells, corresponding to ~1000-fold representation of each shRNA clone. Cultures were passaged at regular intervals as needed to maintain logarithmic growth phase for a total of ~5 population doublings.

shRNA enumeration by NGS. shRNA hairpins were PCR amplified from DNAzol extracted genomic DNA (Thermo Fisher) according to protocols developed by the RNAi Consortium (GGP Web Portal). Briefly, we used PCR primers mapping to a region of the U6 promoter directly upstream of the shRNA sequence and that included P5 and P7 attachment sequences for binding to the Illumina flow cell, Illumina primer binding sites, a 6-nucleotide barcode sequence for library multiplexing, and a staggered region to create sequence diversity during cluster identification (Supplementary Table 7). In order to maintain library representation throughout the PCR amplification process, we processed all genomic DNA from each sample. No more than 825 ng of total DNA was included per reaction58, corresponding to at least 500 templates per shRNA species (59-plex library, assuming 25% transduction efficiency). We performed a total of 30 cycles of amplification with 5% High-Fidelity 2X Master Mix (New England Biolabs). Amplicons of the predicted size (1500 bp) were gel purified from precast E-Gel EX 2% agarose gels (Invitrogen) using the QIAquick gel extraction kit (Qiagen). The size and concentration of the each of the final libraries was verified by Agilent Bioanalyzer prior to pooling for sequencing on an Illumina MiSeq instrument. Paired-end FASTQ files from t0 and t1 were aligned to a reference list of shRNA hairpin sequences with 0 mismatches allowed. Differential shRNA representation was quantified using the edgeR package in R (Bioconductor)59. Results were plotted using the ggplot2 package in R. PCR primers and amplification conditions are available upon request.

Statistics. Quantitative data were analyzed using GraphPad Prism 8.0.1 software and various R packages. Well-initiating cell (WIC) frequencies were calculated from limiting dilution culture results using the online ELDA tool available at http://biostat.wi.edu/software/elda/60.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. RNA-seq data from 17 of 22 T-ALL, 24 B-ALL, and 24 AML PDX samples referenced during the study are available in the NCBI SRA database under the accession code SRP105099. RNA-seq data and associated clinical annotations for samples from the COG T-cell ALL study referenced during the study are available in the database of Genotypes and Phenotypes (dbGaP) under the accession code phs000218/000464. RNA-seq data for normal hematopoietic progenitors referenced during the study are available in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) database under the accession code GSE69239. RNA-seq data for T-ALL cell lines referenced during the study are available in the European Genome-phenome Archive (EGA) database under the accession code EGAS0000100536. Whole exome sequencing (WES), RNA-seq, and ChIP-seq data generated during the current study excluding that in Supplementary Fig. 9 have been deposited in the EGA database under accession code EGAS0000103637. ChIP-seq peak call (BED) files have been deposited in the GEO.
References

1. Rampasaran, A. & Weinberg, R. A. Comparative biology of mouse versus human cells: modelling human cancer in mice. Nat. Rev. Cancer 3, 952–959 (2003).
2. Odom, D. T. et al. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. Nat. Genet. 39, 730–732 (2007).
3. Medyouf, H. et al. High-level IGF1R expression is required for leukemia-initiating cell activity in T-ALL and is supported by Notch signaling. J. Exp. Med. 208, 1809–1822 (2011).
4. Parekh, C. & Crooks, G. M. Critical differences in hematopoiesis and lymphoid development between humans and mice. J. Clin. Immunol. 33, 711–715 (2013).
5. Lai, Y. et al. Current status and perspectives of patient-derived xenograft models in cancer research. J. Hematol. Oncol. 10, 106 (2017).
6. Townsend Elizabeth, C. et al. The public repository of xenografts enables discovery and randomized phase II-like trials in mice. Cancer Cell 29, 574–586 (2016).
7. Zhang, J. et al. The genetic basis of early T-cell precursor acute lymphoblastic leukemia. Nature 481, 157–163 (2012).
8. Liu, Y. et al. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. Nat. Genet. 49, 1211–1218 (2017).
9. Jacoby, E., Chien, C. D. & Fry, T. J. Murine models of acute leukemia: important tools in current pediatric. Leuk. Res. Front. Oncol. 4, 95 (2014).
10. Belver, L. & Ferrando, A. The genetics and mechanisms of T cell acute lymphoblastic leukemia. Nat. Rev. Cancer 16, 494–507 (2016).
11. Sulong, S. et al. A comprehensive analysis of the CDKN2A gene in childhood acute lymphoblastic leukemia reveals genomic deletion, copy number neutral loss of heterozygosity, and association with specific cytogenetic subgroups. Blood 113, 100–107 (2009).
12. Jacobs, J. J. L., Kieboom, K., Marino, S., DePinho, R. A. & van Lohuizen, M. The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the inka locus. Nature 397, 164–168 (1999).
13. Park, I.-K., Morrison, S. J. & Clarke, M. F. Bmi1, stem cells, and senescence regulation. J. Clin. Investig. 113, 175–179 (2004).
14. Szymczak, A. L. et al. Correction of multi-gene deficiency in vivo using a single ‘self-crediting’ 2A peptide-based retroviral vector. Nat. Biotech. 22, 589–594 (2004).
15. Awong, G. et al. Human pro-T-cells generated in vitro facilitate hematopoietic stem cell-derived T-lymphopoesis in vivo and restore thymic architecture. Blood 122, 4210–4219 (2013).
16. Blom, B. & Spits, H. Development of human lymphoid cells. Annu. Rev. Immunol. 24, 287–320 (2006).
17. Awong, G. et al. Characterization in vitro and engraftment potential in vivo of human progenitor T cells generated from hematopoietic stem cells. Blood 114, 972–982 (2009).
18. van Dongen, J. J. et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene rearrangements in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-C798-9306. Leukemia 17, 2257–2317 (2003).
19. Perentes, J. P. et al. RAS oncogene mutations and outcome of therapy for human acute T-cell leukemia (T-ALL). Blood 106, 274–286 (2005).
20. Spleman, F. et al. A new recurrent inversion, inv(7)(p15q34), leads to transcriptional activation of HOXA10 and HOXAI1 in a subset of T-cell acute lymphoblastic leukemias. Leukemia 19, 358–366 (2005).
21. Alharbi, R. A., Pettengell, R., Pandha, H. S. & Morgan, R. The role of HOX genes in normal hematopoiesis and acute leukemia. Leukemia 27, 1000–1008 (2013).
22. Uren, A. G., Kool, J., Berns, A. & van Lohuizen, M. Retroviral insertion mutagenesis: past, present and future. Oncogene 24, 7656–7672 (2005).
23. Futreal, P. A. et al. A census of human cancer genes. Nat. Rev. Cancer 4, 177–183 (2004).
24. Hnisz, D. et al. Activation of proto-oncogenes by disruption of chromosome neighborhood. Science 351, 1454–1458 (2016).
25. Antonchuk, J., Sauvageau, G. & Humphries, R. K. HOXB4-induced expansion of adult hematopoietic stem cells ex vivo. Cell 109, 39–45 (2002).
26. Björnsson, J. M. et al. Reduced proliferative capacity of hematopoietic stem cell deficient in Hoxb3 and Hoxb4. Mol. Cell. Biol. 23, 3872–3883 (2003).
27. Lim, D. H. et al. Committed hemopoietic progenitors, not stem cells, are the principal responders to Hox gene transduction. bioRxiv https://doi.org/10.1101/174490 (2017).
28. García-Peydró, M. et al. The NOTCH1/CD44 axis drives pathogenesis in a T cell acute lymphoblastic leukemia model. J. Clin. Invest. 128, 2802–2818 (2018).
29. Gundry Michael, C. et al. Highly efficient genome editing of murine and human hematopoietic progenitor cells by CRISPR/Cas9. Cell Rep. 17, 1453–1461 (2016).
30. Giambra, V. et al. NOTCH1 promotes T cell leukemia-initiating activity by RUNX-mediated regulation of PKC-θ and reactive oxygen species. Nat. Med. 18, 1693–1698 (2012).
31. Medyouf, H. et al. Acute T-cell leukemias remain dependent on Notch signaling despite PTEN and INK4A/ARF loss. Blood 115, 1175–1184 (2010).
32. Yost, A. J. et al. Defined, serum-free conditions for in vitro culture of primary human T-ALL blasts. Leukemia 27, 1437–1440 (2013).
33. Carlson, C. S. et al. Using synthetic templates to design an unbiased multiplex PCR assay. Nat. Commun. 4, 2680 (2013).
34. Robins, H. S. et al. Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. Blood 114, 4099–4107 (2009).
35. Robins, H. et al. Ultra-sensitive detection of rare T cell clones. J. Immunol. Methods 375, 14–19 (2012).
36. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997 [q-bio.GN] (2013).
37. Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22, 568–576 (2012).
38. Cangolli, P. A. et al. Program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6, 80–92 (2012).
39. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
40. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930 (2014).
41. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 530 (2014).
42. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 28, 882–883 (2012).
43. Leek, J. et al. sva: surrogate variable analysis. R package version 3.2.2 (2016).
44. Gentleman, R., Carey, V., Huber, W. & Irizarry, R. A. An R and S-PLUS package for differential expression in multifactor experiments. Bioinformatics 19, 2050-2051 (2003).
45. Gentleman, R. A. et al. Nucleosome density ChIP-Seq identifies distinct chromatin modification signatures associated with M Nas accessiblity. Cell Rep. 17, 2112–2124 (2016).
46. Lorzadeh, A. et al. Nucleosome density ChIP-Seq identifies distinct chromatin modification signatures associated with M Nas accessiblity. Cell Rep. 17, 2112–2124 (2016).
47. Lorzadeh, A., Lopez, Gutierrez R., Jackson, L., Moks, M. & Hirst, M. Generation of native chromatin immunoprecipitation sequencing libraries for nucleosome density analysis. J. Vis. Exp. 56085 https://doi.org/10.3791/56085 (2017).
48. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26, 589–595 (2010).
49. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
50. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R17 (2008).
58. Strezoska, Ž. et al. Optimized PCR conditions and increased shRNA fold representation improve reproducibility of pooled shRNA screens. PLoS ONE 7, e42341 (2012).
59. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2010).
60. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. J. Immunol. Methods 340, 70–78 (2009).

Acknowledgements
This work was supported by Program Project Grant funding from Terry Fox Research Institute. M.K. received postdoctoral support from Japan Society for the Promotion of Science. A.C.S. received a Canada Graduate Scholarship-Master’s studentship from Canadian Institutes of Health Research.

Author contributions
M.K. and A.P.W. designed experiments, interpreted results, and wrote the manuscript; M.K., A.C.S., K.T., R.W., A.N., C. Shanna, S.G., E.A.C. and A.L. generated data; K.T., A.Z., A.H., S.B., R.A.H., S.H. and A.B. performed informatics analyses; X.W., C. Steidl, A.K., R.K.H., C.J.E. and M.H. provided advice and discussion.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-10510-8.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Peer review information: Nature Communications thanks Iva Nikolic, Rachel Rau and María Toribio for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access
This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019