A unique mutator phenotype reveals complementary oncogenic lesions leading to acute leukemia

Mianmian Yin, …, Paul S. Meltzer, Peter D. Aplan


Graphical abstract

Find the latest version:

http://jci.me/131434/pdf
A unique mutator phenotype reveals complementary oncogenic lesions leading to acute leukemia

Mianmian Yin, Timour Baslan, Robert L. Walker, Yuelin J. Zhu, Amy Freeland, Toshihiro Matsukawa, Sriram Sridharan, Andre Nussenzweig, Steven C. Pruitt, Scott W. Lowe, Paul S. Meltzer, Peter D. Aplan

1Genetics Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; 2Cancer Biology and Genetics Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA; 3Department of Molecular and Cellular Biology, Roswell Park Cancer Institute, Buffalo, NY, USA; 4Laboratory of Genome Integrity, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; 5Howard Hughes Medical Institute, Chevy Chase, MD, USA.

Running title: Detection of lymphoid leukemia tumor suppressor genes.

Keywords: Mcm2, DNA replicative stress, B-cell precursor acute lymphoblastic leukemia, precursor T-cell lymphoblastic leukemia/lymphoma, tumor suppressor genes

* Corresponding author: Peter D. Aplan, Genetics Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Building 37 Room 6002, 37 Convent Drive, Bethesda, MD 20892, USA, PH: 240-760-6889, FAX: 240-541-4477, aplanp@mail.nih.gov
ABSTRACT

Mice homozygous for a hypomorphic allele of DNA replication factor minichromosome maintenance protein 2 (designated Mcm2<sup>cre/cre</sup>) develop precursor T-cell lymphoblastic leukemia/lymphoma (pre-T LBL) with 4-32 small interstitial deletions per tumor. Mice that express a NUP98-HOXD13 (NHD13) transgene develop multiple types of leukemia, including myeloid, T and B lymphocyte. All Mcm2<sup>cre/cre</sup>NHD13<sup>+</sup> mice develop pre-T LBL, and 26% develop an unrelated, concurrent B-cell precursor acute lymphoblastic leukemia (BCP-ALL). Copy Number Alteration (CNA) analysis demonstrated that pre-T LBL were characterized by homozygous deletions of Pten and Tcf3, and partial deletions of Notch1 leading to Notch1 activation. In contrast, BCP-ALL were characterized by recurrent deletions involving Pax5 and Ptpn1, and copy number gain of Abl1 and Nup214 resulting in a Nup214-Abl1 fusion. We present a model in which Mcm2 deficiency leads to replicative stress, DNA double strand breaks, and resultant CNAs due to errors in DNA DSB repair. CNAs which involve critical oncogenic pathways are then selected in vivo as malignant lymphoblasts, due to a fitness advantage. Some CNAs, such as those involving Abl1 and Notch1, represent attractive targets for therapy.
Introduction

DNA replicative stress, typically in response to oncogene-induced hyperproliferation, has been linked to malignant transformation\(^{(1)}\) \((2, 3)\). Although the precise mechanism(s) by which replicative stress might result in cancer remain unknown, it is thought that chronic replicative stress leads to replication fork stalling, collapse, and subsequent DNA double strand breaks (DSB) at or near the site of replication fork collapse \(^{(2)}\). Inefficient repair of these DNA DSB leads to genomic instability reflected by indels, and structural variations involving genes important for malignant transformation \(^{(2)}\).

Minichromosome maintenance complex component 2 (\textit{Mcm2}) is a core component of the DNA replication-licensing complex required for replication initiation during S-phase. A multi-subunit complex containing \textit{Mcm2-7} is responsible for the initial unwinding of DNA \((4, 5)\). We previously generated a mouse strain containing a CreERT2 casette flanked by an IRES “knocked” into the 3’ UTR of \textit{Mcm2}, creating a \textit{Mcm2}^{IRES-CreERT2} allele (henceforth referred to as \textit{Mcm2}^{IRES-CreERT2}) \(^{(6)}\). Unexpectedly, this knock-in allele resulted in diminished expression of \textit{Mcm2} protein. \textit{Mcm2}^{cre/cre} mouse embryo fibroblasts expressed approximately one-third the amount of \textit{Mcm2} protein compared to wild-type (WT) controls. Remarkably, almost all \textit{Mcm2}^{cre/cre} mice develop a lethal precursor T cell lymphoblastic leukemia/lymphoma (pre-T LBL) within four months of age. Previous studies had identified abnormal karyotypes and increased levels of chromosome breaks in cultured cells with reduced expression of \textit{Mcm} proteins\(^{(7, 8)}\), and array comparative genomic hybridization demonstrated that pre-T LBL in \textit{Mcm2}^{cre/cre} mice had numerous small (average <0.5 Mbp) genomic deletions including several genes known to be relevant for human pre-T LBL, such as \textit{E2a} (\textit{Tcf3}), and \textit{Ptten} \(^{(9)}\), leading to the hypothesis that \textit{Mcm2} deficiency led to deletions of important tumor suppressor genes, resulting in malignant transformation. These findings led us to view \textit{Mcm2}^{cre/cre} cells as a potential tool for inducing mutation, analogous to ionizing radiation\(^{(10)}\), ethyl nitroso-urea\(^{(11)}\) or retroviral insertion\(^{(12)}\). In this view, dysregulation of \textit{Mcm2} protein is not directly oncogenic, instead, \textit{Mcm2} dysregulation leads to widespread genomic deletions, some of which are oncogenic. Cells which undergo a combination of mutations that
dysregulate several complementary, collaborative pathways are selected in vivo due to a fitness advantage, and emerge as a malignancy.

A *NUP98-HOXD13 (NHD13)* fusion gene has been identified in patients with myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) (13). Expression of an *NHD13* fusion in the hematopoietic compartment of mice led to overexpression of *Hoxa* cluster genes and resulted in a highly penetrant MDS phenotype(14). Approximately 80% of *NHD13* mice transform to acute leukemia, most commonly AML, with the remainder dying from complications of MDS, such as severe anemia or infection, without signs of leukemic transformation (14). Leukemic transformation was frequently associated with spontaneous mutations of *Nras, Kras*, or *Cbl*, suggesting that these mutations collaborated with the *NHD13* fusion, and were biologically selected in vivo(15).

In addition to targeted resequencing of candidate genes, retroviral insertional mutagenesis (RIM) identified genes whose overexpression would collaborate with an *NHD13* transgene to induce AML(16). Although RIM can identify gene inactivation events as well as gene activation events, tumor suppressor genes have typically not been identified through RIM screens, perhaps because two alleles need to be targeted for complete inactivation of many tumor suppressor genes, as opposed to a single allele for gene activation events (17). We hypothesized that the unique “deleter” phenotype found in *Mcm2<sup>cre/cre</sup>* mice could potentially be used to identify tumor suppressor genes in the context of AML, by crossing the *NHD13* transgene onto an *Mcm2<sup>cre/cre</sup>* background.
**Results**

*Mcm2*<sup>cre/cre</sup>*NHD13*<sup>+</sup> mice do not develop AML.

To determine whether *Mcm2* deficient mice (*Mcm2*<sup>cre/cre</sup>) can be used to identify tumor suppressor genes important for the development of AML, we generated *Mcm2*<sup>cre/cre</sup>*NHD13*<sup>+</sup> mice. Similar to prior studies, *Mcm2* protein expression in thymus from one month old *Mcm2*<sup>cre/cre</sup> mice was only 28% that of WT mice; addition of the *NHD13* transgene had no effect on the *Mcm2* protein level (Supplemental Figure 1). Mice were euthanized when they presented with signs of leukemia including weight loss, kyphosis, lethargy, visible lymphadenopathy, and dyspnea. *Mcm2*<sup>cre/cre</sup> mice had a markedly decreased survival compared to mice that were heterozygous for this allele (Figure 1A), and adding the NHD13 transgene to the *Mcm2*<sup>cre/cre</sup> mice (*Mcm2*<sup>cre/cre</sup>*NHD13*<sup>+</sup>) led to a modest but significant decrease in survival. Previous reports had demonstrated that *Mcm2*<sup>cre/wt</sup> mice were not predisposed to malignancy (6). To determine whether they had a subtle predisposition toward myeloid malignancy that could be uncovered by the *NHD13* transgene, we compared the median survival of *Mcm2*<sup>cre/wt</sup>*NHD13*<sup>+</sup> to *Mcm2*<sup>wt/wt</sup>*NHD13*<sup>+</sup>; the median survival was not significantly different (259 vs 324 days) (Figure 1A), and was similar to our previous reports with *NHD13*<sup>+</sup> mice (14, 18). In summary, these results showed that the *NHD13* transgene did not accelerate the onset of disease in heterozygous *Mcm2*<sup>cre/wt</sup> mice, and caused a modest disease acceleration in *Mcm2*<sup>cre/cre</sup> mice.

Necropsy of both *Mcm2*<sup>cre/cre</sup>*NHD13*<sup>+</sup> and *Mcm2*<sup>cre/cre</sup>*NHD13*<sup>-</sup> mice revealed markedly enlarged thymus, splenomegaly, and hepatomegaly. Complete blood count (CBC) from the *Mcm2*<sup>cre/cre</sup>*NHD13*<sup>-</sup> cohort typically showed leukocytosis but were otherwise normal, whereas CBC from the *Mcm2*<sup>cre/cre</sup>*NHD13*<sup>+</sup> cohort showed both leukocytosis and leukopenia, anemia, and thrombocytopenia, consistent with prior studies on *NHD13* mice (14) (Supplemental Table S1). Leukemic subtype determined by flow cytometry demonstrated all but one of the *Mcm2*<sup>cre/cre</sup> mice (with or without the *NHD13* transgene) had infiltration of thymus, bone marrow and spleen with malignant thymocytes that stained for CD4, CD8, or both. The diagnosis of pre-T LBL was further supported by clonal VDJ or DJ rearrangements of the Tcrb gene.
(Supplemental Table S2), and infiltration of parenchymal organs such as liver, kidney, and lung with CD3+ lymphoblasts (Supplemental Figure 2).

The stage of thymocyte differentiation was influenced by presence of the NHD13 transgene. Whereas most of Mcm2\textsuperscript{cre/cre}NHD13\textsuperscript+ mice displayed a CD4\textsuperscript+CD8\textsuperscript+ (DP) immunophenotype, a wide spectrum of pre-T LBL immunophenotypes was identified in the thymus of Mcm2\textsuperscript{cre/cre}NHD13\textsuperscript+ mice, including DP, CD4\textsuperscript+CD8\textsuperscript{het}, CD8\textsuperscript+CD4\textsuperscript{het}, CD4\textsuperscript+CD8\textsuperscript- and CD4\textsuperscript-CD8\textsuperscript- (Figure 1B and C). Notably, none of the 46 Mcm2\textsuperscript{cre/cre}NHD13\textsuperscript+ mice developed an AML. The Mcm2\textsuperscript{cre/cre}NHD13\textsuperscript+ and Mcm2\textsuperscript{wt/wt}NHD13\textsuperscript+ mice developed disease with similar immunophenotype, primarily AML, and less commonly pre-T LBL or BCP-ALL (Supplemental Table S1). In summary, these results demonstrated that Mcm2\textsuperscript{cre/cre}NHD13\textsuperscript+ mice did not develop AML, nor did Mcm2\textsuperscript{cre/cre}NHD13\textsuperscript+ mice show accelerated myeloid leukemic transformation compared to Mcm2\textsuperscript{wt/wt}NHD13\textsuperscript+ transgenic mice, suggesting that this strategy did not uncover AML tumor suppressor genes.

**Recurrent deletions in pre-T LBL.**

Based on prior studies (9), we suspected that the pre-T LBL identified in Mcm2\textsuperscript{cre/cre}NHD13\textsuperscript+ mice were initiated, at least in part, by ~0.5 Mb genomic deletions. In addition, although there was no acceleration of AML in Mcm2\textsuperscript{cre/cre}NHD13\textsuperscript+ mice compared to Mcm2\textsuperscript{wt/wt}NHD13\textsuperscript+ mice, we wished to determine if Mcm2\textsuperscript{cre/cre}NHD13\textsuperscript+ mice had recurrent genomic deletions. Copy number alteration (CNA) analysis for pre-T LBL and AML samples were determined by sparse whole genome sequence (WGS; see methods), an approach that can accurately identify copy number losses (deletions) and gains at a resolution of roughly 125 kb, and map alteration breakpoints to a resolution of approximately 25 kb (19). As anticipated, we detected 4-32 small (100-1000 kb) deletions and gains in pre-T LBL from Mcm2\textsuperscript{cre/cre} samples (Figure 2A and Supplemental Figure 3, A-C) with focal deletions more common than gains (Supplemental Figure 3D).
Inspection of Figure 2A shows recurrent interstitial deletions of chromosome 5, 6, 7, 9, 10, 12, 14, 17 and 19; Mcm2^{cre/cre}NHD13^{+} pre-T LBL showed a deletion pattern similar to Mcm2^{cre/cre}NHD13^{+} pre-T LBL. We saw no recurrent acquired CNA in the Mcm2^{cre/wt}NHD13^{+}AML samples, indicating that the heterozygous Mcm2^{cre} allele did not predispose to recurrent deletions, even in the presence of an NHD13 sensitizing transgene. A higher magnification view (Figure 2, B-C) demonstrates that 1) deletions can be more (e.g., Pten) or less (e.g., Bcl7a) tightly clustered, 2) homozygous or heterozygous, and 3) no deletions are identical between different samples.

There were 15 common deleted regions (defined as deleted or gained in at least 15% of samples) in the 29 pre-T LBL primary tumors (11 Mcm2^{cre/cre}NHD13^{+} and 18 Mcm2^{cre/cre}NHD13^{+})(Supplemental Table S3); nine regions were deleted in >60% of the samples. These included expected regions, such as Tcra and Tcrb, as well as previously reported regions that encompass known tumor suppressor genes, such as Pten(20), Tcf3(21), Cdkn1a(22), Bcl11b(23) (Supplemental Figure 4), and Zfp36l2(24), and genes involved in normal T cell differentiation, including Tcf12, Bcl7a, and Bcl7c. A single region from chromosome 8 was recurrently amplified, specifically in Mcm2^{cre/cre}NHD13^{+} mice; there was no obvious candidate oncogene in this region.

To validate the small (125-1000 kb) deletions detected by sparse WGS, we performed conventional WGS, with ~28X coverage, for three Mcm2^{cre/cre}NHD13^{+} pre-T LBL, using non-malignant tail DNA as a germline control. WGS identified a total of 93 genomic deletions in 3 Mcm2^{cre/cre}NHD13^{+} pre-T LBL (Supplemental Table S4). 85% (57/67) of the CNAs identified by sparse WGS were verified (Supplemental Figure 5A). An additional 36 deletions were identified by WGS; most of these were smaller than 125 kb (Supplemental Table S4), consistent with the anticipated resolution of sparse WGS method. When only deletions greater than >125 kb are considered, samples 2739 and 2854 were 100% concordant, whereas sample 2883 was only 68% concordant. One potential reason for the poor correlation in sample 2883 is the observation that this sample had substantial contamination with WT DP thymocytes.
(flow cytometry showed 60.5 % CD4⁺CD8⁺het and 35.9% WT CD4⁺CD8⁺ thymocytes) (Supplemental Figure 5, B-D).

Fifteen of the deleted regions involved T-cell receptor alpha, beta, gamma, or delta. Since these regions undergo programmed deletions in the course of normal development due to VDJ recombination, we focused our analysis on the remaining 78 deletions that were linked to the Mcm2 deficiency. We identified a total of 11 homozygous deletions in the three samples, these homozygous deletions included Pten, Dnmt3a, Tcf3, and Bcl7a (Supplemental Table S5). An additional 9 deletions were identified that involved adjacent, but not overlapping regions (Supplemental Table S5). Sample 2754 had one large Ikzf1 deletion, encompassing the entire gene, and a focal 9177 bp deletion, which deleted exons 4 and 5. RT-PCR revealed several splice forms of Ikzf1 not seen in WT BM (Supplemental Figure 5, E-G), all missing exons 4 and 5. The interstitial deletion of Ikzf1 and aberrant splice forms was reminiscent of IKZF1 deletions and aberrant isoforms associated with human lymphoid leukemia(25). Detailed analysis of the sequence at the deletion junction of the 78 deleted regions in pre-T LBL samples revealed that 51 of the 78 junctions had 1-5 bp of microhomology (Supplemental Table S4). 25 junctions contained non-templated insertions at the breakpoint junctions; some of these were quite extensive, from 1-50 nucleotides in length.

We next used whole exome sequencing (WES) to determine if small indels or single nucleotide variants (SNVs) contributed to the development of pre-T LBL. WES results revealed a total of 104 Tier 1 mutations in the 29 pre-T LBL, or an average of 3/tumor. Only 6 of the 29 samples had mutations in genes known or suspected to be involved in pre-T LBL, three Notch1, two Tp53, and one Ezh2 mutation (24, 26) (Supplemental Table S6). All of these mutations were confirmed (Supplemental Figure 6) by Sanger sequencing of tumor and normal tissue.

**Ongoing deletions in Mcm2cre/cre cell lines.**
We established immortal, cytokine-independent cell lines from 10 $Mcm2^{cre/cre}$ pre-T LBL samples. Since the cell lines were uncontaminated by any normal tissue, we used genomic DNA from the cell lines to validate loss of DNA and protein expression (Supplemental Figure 7, A-C). The anticipated homozygous loss for $Zfp36L2$, $Tcf3$, and $Pten$ were confirmed (Supplemental Figure 7A), as well as loss of $Pten$ protein (Supplemental Figure 7C). However, sample 2773 remained positive in the PCR assay but did not produce $Pten$ protein. A diagram of the deletions (Supplemental Figure 7B) demonstrates that although the deletions from #2773 were not overlapping, each allele had deleted significant portions of $Pten$ coding sequence. $Mcm2$ protein expression was similar in the $Mcm2$ deficient cell lines and primary samples (Supplemental Figure 7D). To determine 1) if the cell line that emerged in tissue culture represented the predominant clone in vivo, and 2) if the cell lines continued to undergo genomic deletions, we compared CNAs in primary tumor, early passage (TCE, 1-1.5 months in culture) and late passage (TCL, 2-6 months in culture) pre-T LBL cell lines. In all cases, the global pattern of deletions was similar to the primary tumor, indicating that the cell line that emerged in tissue culture was representative of the primary tumor (Figure 3). Deleted regions were scored for each trio and summarized in Table 1. A total of 135 deletions were identified; 90 were stable and present in all three samples, suggesting that these were truncal lesions. 22 deletions were acquired at the early passage stage; these could have been acquired during tissue culture or could have been present as a minor clone in the primary tumor. Only 10 new deletions were acquired at the late passage stage, compared to 90 that were present initially and persistent.

To gain further insight into the frequency of ongoing deletions in the $Mcm2^{cre/cre}$ cell lines, we obtained single cell clones by plating late passage 2696 cell line at limiting dilution. Five of 96 wells expanded and were analyzed for CNAs. Supplemental Figure 7E shows new deletions present in the single cell clones; An average of five new deletions per clone compared to the late passage cell line. These deletions involved $Rpl5$, $Cntnap2$ and $Ccdc$ genes, which are not recurrently deleted in other $Mcm2^{cre/cre}$ pre-T LBL samples; Taken together, these findings suggest that although the “deleter” phenotype had become less
pronounced during in vitro culture, ongoing interstitial deletions could be identified in the Mcm2cre/cre cell lines.

**Large Clone Capture sequence identifies frequent Notch1 interstitial deletions**

Mutations involving the extracellular heterodimerization (HD) domain and/or the C-terminal PEST domain of NOTCH1 are found in >50% of both human and murine pre-T LBL (26) (27). In addition, a recurrent deletion of Notch1 exon 1 and the adjacent 5’ regulatory sequences has been identified in mice, but not humans (28, 29). This deletion leads to use of alternate Notch1 transcript initiation 5’ of exon 27, and translation initiation at M1727 within exon 28; resulting in production of a truncated Notch1 protein that lacks the extracellular ligand binding domain, retains the TM domain and sensitivity to gamma secretase inhibitors, and is functionally similar to a NOTCH1 HD mutation (28). Given these findings, we were puzzled that only three Notch1 mutations were identified by WES. Since the WGS revealed an interstitial deletion in one of three samples, we hypothesized that smaller (5-125 kb) interstitial deletions not identified in the sparse sequence CNA assay might be common events in these pre-T LBL. “Large Clone Capture” (LCC)-Seq, is a technique that allows for the custom capture and next generation sequencing of specific genomic regions of interest, similar in concept to whole exome sequencing (30). To search for Notch1 deletions, as well as identify precise breakpoints involved in deletions, we obtained BAC clones that covered regions of chromosome 2 (Notch1, Nup214-Abl1 and Ptn1), 7 (Bcl7c), 10 (Tcf3), and 19 (Pten).

LCC-Seq identified 28 deletions involving Notch1 in 60 independent samples (Supplemental Table S7). Those deletions led to loss of Notch1 exons 2, 1-2, 1-4, 1-19, 1-23, 1-24, 1-27, 3-27, 16-26, or 16-27 (Supplemental Table S7 and Figure 4A). Aberrant Notch1 mRNA splice junctions were detected in samples that had deletions which retained exon 1 (Figure 4B and Supplemental Figure 8). Importantly, all of the aberrant splice forms (3-27, 16-27) retained exon 28, which contains an alternate translation initiation site at Notch1 M1727 (28). Western blot and real-time PCR indicated that samples with Notch1
interstitial deletion produced an intracellular domain of Notch1 (ICN) protein and high Hes1 expression (Figure 4, C-D).

To determine whether sustained high level Notch1 signaling and was required for vigorous growth of the Mcm2\text{cre/cre} TCL, we treated Notch1-deleted cell lines with Compound E, a gamma secretase inhibitor that prevents cleavage of the membrane-bound ICN and resultant transport of the transcriptionally active ICN to the nucleus(26). Compound E did not affect growth of 2795 TCL, a cell line that was wildtype for Notch1 (Supplemental Figure 9A), but significantly inhibited the growth and ICN expression of Mcm2\text{cre/cre} pre-T LBL cell lines with Notch1 mutations (Supplemental Figure 9, B-F). These results indicated that Mcm2\text{cre/cre} TCL require ongoing Notch1 signaling for active proliferation in vitro.

**Mcm2\text{cre/cre} NHD13\text{*} mice develop BCP-ALL.**

In addition to the pre-T LBL detected in the thymus, 12 of 46 Mcm2\text{cre/cre} NHD13\text{*} mice had an unexpected B cell expansion in the BM and spleen (Figure 5A and Supplemental Table S1). None of the Mcm2\text{cre/cre} NHD13\text{*} mice showed this phenotype, in which a predominant CD19\text{*}/B220\text{dim} or CD19\text{*}/B220\text{dim} population was identified in BM and spleen, while the thymus was infiltrated with CD4\text{*}/CD8\text{*}pre-T LBL (Figure 5A). Histologic analysis showed infiltration of BM, spleen, and kidney with clusters of B220\text{*} cells; conversely, the thymus did not stain for B220, but stained for CD3 (Figure 5B) and BM was replaced with lymphoblasts (Figure 5C). The diagnosis of BCP-ALL was further supported by the presence of clonal IgH gene rearrangements (Figure 5D and Supplemental Table S8). In addition to BCP-ALL in Mcm2\text{cre/cre} NHD13\text{*} mice, 10\% of Mcm2\text{cre/wt} NHD13\text{*} mice and 11\% of Mcm2\text{wt/wt} NHD13\text{*} mice developed BCP-ALL (Figure 5E and Supplemental Table S8). The ratio of AML to BCP-ALL in Mcm2\text{cre/wt} NHD13\text{*} and Mcm2\text{wt/wt} NHD13\text{*} mice was 7:1 and 7:1 respectively, similar to previously published findings for NHD13 mice(14). However, the ratio of AML to BCP-ALL in the Mcm2\text{cre/cre} NHD13\text{*} mice was zero, as no Mcm2\text{cre/cre} NHD13\text{*} mice developed AML. The skewed proportion of BCP-ALL in Mcm2\text{cre/cre} NHD13\text{*} mice compared to Mcm2\text{wt/wt} NHD13\text{*} indicated that the
Mcm2 hypomorph dramatically accelerated the development of BCP ALL, but not AML, in NHD13+ background (Table 2).

To gain insight into the leukemic transformation of Mcm2\textsuperscript{cre/cre}NHD13+ B cell precursors, we assessed the differentiation of B and T cells from Mcm2\textsuperscript{cre/cre}NHD13+ mice at one month of age, prior to overt leukemic transformation. Flow cytometry showed a modest decrease in B220\textsuperscript{+}CD19\textsuperscript{+} precursors in Mcm2\textsuperscript{wt/wt}NHD13+ and Mcm2\textsuperscript{cre/cre}NHD13+ compared to WT mice; however, there was a marked decrease in B220\textsuperscript{+}CD19\textsuperscript{+} precursors in Mcm2\textsuperscript{cre/cre}NHD13+ BM, suggesting impaired B cell maturation. Sub-fractionation of B cell precursors revealed a 7-fold decrease (21. vs 3.%) of B220\textsuperscript{+}CD43\textsuperscript{-} pre-B cells in the Mcm2\textsuperscript{cre/cre}NHD13+ BM, while the proportion of pro-B cells was similar in all 4 genotypes (Supplemental Figure 10, A-C), suggesting a block in differentiation at the pro-B to pre-B stage of maturation. In contrast, we detected no clear abnormalities of T cell differentiation at one month of age in any thymocyte and splenocyte of the genotypes studied (Supplemental Figure 10, D-E).

**Recurrent CNAs in BCP-ALL**

Similar to pre-T LBL samples, we identified recurrent, acquired CNAs in Mcm2\textsuperscript{cre/cre}NHD13+ BCP-ALL (Figure 6A and Supplemental Figure 11A). We also assessed CNA in four Mcm2\textsuperscript{cre/wt}NHD13+ samples; as opposed to the pattern of 125-1000 kb deletions seen in Mcm2\textsuperscript{cre/cre} malignancies (both pre-T LBL and BCP-ALL), there were few 125-1000 kb deletions in the Mcm2\textsuperscript{cre/wt}NHD13+, consistent with prior observations that mice heterozygous for an Mcm2\textsuperscript{cre} allele are not prone to the deletor phenotype. Comparison of CNAs from pre-T LBL and BCP-ALL from the same mouse showed no similarities between the pairs (Supplemental Figure 11, B-D), indicating that the T and B cell malignancies did not arise from a common precursor.

Five of seven Mcm2\textsuperscript{cre/cre}NHD13+ BCP-ALL showed deletion involving Pax5, one of the most frequently mutated genes in human BCP-ALL (Figure 6B and Supplemental Table S9). Moreover, the same five samples showed deletion (four of five homozygous) of a region that encompassed Cebpb and Ptpn1.
(Figure 6, C-D and Supplemental Table S9). However, *Cebpb* is not highly expressed in normal B cell precursors, whereas *Ptpn1* is highly expressed in B cell precursors. Figure 6E shows marked downregulation of *Ptpn1* in four samples with putative homozygous deletions.

**Recurrent Nup214-Abl1 fusion gene in Mcm2\textsuperscript{cre/creNHD13\textsuperscript{*}} BCP-ALL.**

Four *Mcm2\textsuperscript{cre/creNHD13\textsuperscript{*}}* BCP-ALL showed copy number gain (estimated 1-2 copies) of a region bounded by *Abl1* and *Nup214* (Figure 7A and Supplemental Table S9). Given that *NUP214-ABL1* fusions have been identified in patients with pre-T LBL(31) and BCP-ALL(32), we hypothesized that *Nup214-Abl1* gains could result in a *Nup214-Abl1* fusion. This putative fusion could be produced by an episome, as shown in human pre-T LBL(31), or by a tandem duplication (Figure 7B). We designed primers to amplify a fusion between *Nup214* exon 23, 29, 31, 32, or 34 and *Abl1* exon 2 based on the fusions in human pre-T LBL(31). A fusion PCR product was identified in all four samples that showed *Nup214-Abl1* copy number gain (samples 2842, 2811, 2725, and 2703). In addition, a *Nup214-Abl1* fusion was identified in one sample (2905) that did not show an amplified region, and in 2 of 3 samples that were not assayed for CNA (Figure 7C). *Nup214-Abl1* fusion genes were not identified in pre-T LBL, AML or *Mcm2\textsuperscript{cre/wtNHD13\textsuperscript{*}}* BCP-ALL (Figure 7, C-D). Nucleotide sequence demonstrated a *Nup214* exon 32 to *Abl1* exon 2 fusion in six samples; one sample (#2697) had a *Nup214* exon 31-*Abl1* exon 2 fusion (Table 3). *Crkl*, a direct target of the *Abl1* kinase (31), was phosphorylated in splenocytes that expressed the *Nup214-Abl1* fusion (Figure 7E) indicating that *Nup214-Abl1* acts as an activated *Abl1* kinase. Imatinib, an inhibitor of Ab1 kinase (33), inhibited the growth of splenocytes, which express a *Nup214-Abl1* fusion (Figure 7F).

We used WES to identify Tier 1 (coding) mutations in *Mcm2\textsuperscript{cre/creNHD13\textsuperscript{*}}* mice and found no recurrent mutations in the 11 samples assayed (Supplemental Table S10). In contrast, all three *Mcm2\textsuperscript{cre/wtNHD13\textsuperscript{*}}* samples showed at least one Tier 1 mutation in genes known to be relevant for human BCP-ALL (*Sh2b3*(34), *Jak1, Trp53*, or *Flt3*(35)) (Supplemental Table S10 and Figure 12).
**Nucleotide level resolution of deletion breakpoints.**

As seen in Figure 2, the genomic regions most susceptible to the 10-1000 kb deletions triggered by the Mcm2 hypomorph are not random, but occur within a limited number of regions. To determine if the breakpoint junctions are focal and follow any clear rules or patterns (such as VDJ recombination), we used LCC-Seq to capture and analyze the breakpoints of several selected regions (Pten, Notch1, Tcf3, Bcl7c, Ptpn1, and Nup214-Abl1) at a nucleotide level. In total, we identified 245 deletions in 91 samples (Supplemental Table S11). Identical deletions found in primary tumor and cell lines were considered as a single independent event. Detailed analysis of the sequence at the deletion junction of 245 independent deleted regions revealed that 155 of the 245 junctions had 1-7 bp of microhomology (Supplemental Table S11). 70 junctions contained non templated insertions at the breakpoint junctions; some of these were quite extensive, from 1-54 nucleotides in length. Given the precedent set by programmed VDJ rearrangement, in which signal sequences can occur at a distance (up to 38 bp) from the breakpoint junction, we analyzed nucleotide sequences in a 200 bp window (100 bp 5’ and 100 bp 3’) flanking both the upstream and downstream breakpoints, in both the WGS data set (Supplemental Table S4) and the LCC-Seq data set (Supplemental Table S11). Comparison of 156 windows from the WGS to 7332 randomly selected 200 bp windows revealed an increased likelihood of mononucleotide repeats in the Mcm2 tumor breakpoints identified by WGS (p=0.009456; Supplemental Figure 13A and Tables S12-13). Moreover, comparison of 490 windows from the 245 independent breakpoints identified by LCC-Seq with 7530 randomly selected 200 bp windows from chr 2, 7, 10, 19 (corresponding to the LCC-Seq chromosomes) also showed an increased frequency of mononucleotide repeats in the Mcm2 tumor breakpoints (p=0.009671; Supplemental Figure 13B and Tables S14-15). These findings are consistent with the observation that upon replication stress, DNA breaks are frequently generated at mononucleotide repeats(36).
Discussion

Replication stress has been linked to many forms of cancer (1, 3). It has been speculated that replicative stress leads to DNA DSB, with subsequent repair by non-homologous end joining (NHEJ), resulting in indel mutations and structural variations, including large interstitial deletions. It is thought that these mutations, such as deletions of important tumor suppressor genes, ultimately cause malignancy. These frequent DNA DSB caused by increased replicative stress lead to activation of Chk1; Chk1 inhibitors have recently entered the clinic (37) as an approach to target unique vulnerabilities in the cancer cell. In this study, we show that mutation of a single gene (Mcm2) leads to a “deletor” phenotype, and that resultant combinations of deletions involving several critical genes act in concert and are selected in vivo as T or B lymphoid malignancies.

Hypomorphs of several members of the Mcm2-7 DNA replicative helicase complex show malignancy or stem cell defects (9, 38-40) which is thought to be associated with replicative stress; in some cases, these malignancies were linked to recurrent interstitial deletions (7, 9, 41). We reasoned that tumor suppressor genes which are important for transformation of MDS to AML could be identified by crossing the Mcm2 mice with a “deletor” phenotype to NHD13 mice. However, all Mcm2\textsuperscript{cre/cre}NHD13\textsuperscript{+} developed T and/or B cell malignancies, but not AML within 5 months of age. We also predicted that Mcm2\textsuperscript{cre/wt} mice, which express approximately 30-40% less Mcm2 protein than WT mice might have a subtle predisposition to malignancy that can be revealed by a cooperating oncogenic event (such as the NHD13 transgene). However, although Mcm2\textsuperscript{cre/wt}NHD13\textsuperscript{+} developed AML, the incidence, age of onset, and immunophenotype was similar to Mcm2\textsuperscript{wt/wt}NHD13\textsuperscript{+} (Figure 1A), suggesting that a single copy of Mcm2\textsuperscript{cre} allele did not predispose to AML. Furthermore, CNA showed that the Mcm2\textsuperscript{cre/wt}NHD13\textsuperscript{+} AML samples did not have recurrent deletions, reinforcing the idea that Mcm2\textsuperscript{cre/wt} mice are not predisposed to malignancy. Although it was previously shown that Mcm2\textsuperscript{cre/cre} mice are not prone to recurrent deletions (9), it remains possible that Mcm2\textsuperscript{cre/wt} mice are prone to more infrequent deletions, that are not evident because they are polyclonal and not discernable in a polyclonal population of cells. However, the
observation that clonal $Mcm2^{cre/wt}NHD13^+$ AML samples do not show deletions argues strongly against that possibility. Given that all $Mcm2^{cre/cre}NHD13^+$ mice died of pre-T LBL or BCP-ALL by 5 months of age, it remains possible that these mice did not develop AML because they died before AML had a chance to develop.

There were 4-32 deletions per tumor. Nucleotide sequence analysis revealed that there was an increase likelihood of mononucleotide repeats at or near the deletion breakpoints, consistent with the hypothesis that polymerase pausing or stalling at sites of mononucleotide repeats leads to DNA DSB (42, 43); improper repair of two such breaks via NHEJ leads to the interstitial deletions. Given that studies of human pre-T LBL has suggested that there maybe 2-4 “driver” mutation per tumor (24), we considered that many of the deletions, were likely to be passenger deletions. Recurrent deletions may still be passenger deletions in this model, selected by virtue of susceptibility to deletion. Some regions had numerous homozygous deletions, centered over a single gene, such as $Pten$. Alternatively, there were regions where the individual deletion were spread over a larger 2-3 Mb region, with no clear common deleted region, and few homozygous deletions, such as the region encompassing $Bcl7a$. We suspect the former are more likely to be driver deletions, and the latter more likely to be passenger deletions. Using these criteria, the leading candidates for driver deletions were $Pten$ and $Tcf3$. These candidates are supported by experimental studies that show homozygous $Pten$ and $Tcf3$ deletions strongly predispose mice to pre-T LBL(21, 44). Moreover, $PTEN$ deletions/mutation are quite common in human pre-T LBL (16%, (20)). In addition, although $TCF3$ deletions are uncommon in human pre-T LBL, enforced expression of $TAL1$ by a large variety of genomic rearrangements is quite common(24); in this context, it is important to note that enforced expression of $TAL1$ leads to a functional inactivation of $TCF3$(45). Of note was a significantly increased likelihood of $Dnmt3a$ mutations, especially homozygous $Dnmt3a$ mutations(3/11 vs 0/18; p=0.0096), in $Mcm2^{cre/cre}NHD13^+$ mice compared to the $Mcm2^{cre/cre}NHD13^+$ mice. A plausible explanation for this observation is that NUP98 fusion proteins have been shown to increase stem cell self-renewal(46), as has $Dnmt3a$ deletion(47). Thus a homozygous $Dnmt3a$ deletion
which promotes stem cell self-renewal may be redundant in the context of an NHD13+ cell which also leads to increased self-renewal potential.

*Notch1* deletions represent a special case in terms of thymocyte transformation. In contrast to *Pten* and *Tcf3*, in which homozygous deletion of the genes predisposes to pre-T LBL, *Notch1* mutations are most commonly mono-allelic point mutations of the heterodimerization domain (HD) or PEST domain in both humans and mice(26, 27). Mutations in the HD lead to loss of requirement for extracellular Notch1 ligand binding, and PEST domain mutations result in decreased degradation of the transcriptionally active intracellular Notch1 (ICN), resulting in sustained activation of Notch1 target genes(26). A recurrent interstitial deletion of *Notch1* exons 1-2, leads to translation initiation at an internal methionine residue(M1727), production of a protein that lacking the extracellular domain and ligand-independent activation of Notch1(28). In this study, we show recurrent *Notch1* genomic deletions remove the Notch1 extracellular domain and retain the transcriptionally active ICN.

Consistent with the notion that driver mutations were produced by recurrent interstitial deletions in this model, WES studies showed relatively few SNVs or indels. The only recurrent Tier 1 mutations identified in 29 samples were four *Notch1*, one *Ezh2* and two *Tp53* mutations. These results stand in contrast to our prior WES studies of murine leukemia and lymphoma on a WT Mcm2 background, in which 60-100% of malignancies had acquired Tier 1 mutations in genes known to be mutated in hematologic malignancy(48, 49).

Serial analysis of the *Mcm2*cre/cre pre-T LBL cell lines suggests that the *Notch1* mutations are relatively late events, consistent with sequencing studies of human pre-T LBL, in which almost half of the *Notch1* mutations identified were thought to be subclonal(24). An overall scheme for cooperative lesions that result in transformation in this model is outlined in Supplemental Figure 14A. Increased stem cell self-renewal is conferred by either the NHD13 transgene or *Dnmt3a* inactivation. *Tcf3* deletion results in a block to thymocyte differentiation, and *Pten* deletion leads to hyperproliferation. Finally, *Notch1*
mutations lead to ligand independent growth; the cells no longer require Notch1 ligands supplied by thymic epithelial cells, and are able to metastasize as well as expand in vitro.

CNA analysis of pre-T LBL cell lines indicated that the overall frequency of new deletions seemed to decrease with time (Table 1). The decreased frequency of new deletions could reflect accommodation to the Mcm2 protein deficiency, alternatively, it is conceivable that there are relatively few exquisitely sensitive regions susceptible to deletion in Mcm2 hypomorph pre-T LBL, and these are deleted relatively early. The ongoing Mcm2 deficiency suggests that these cell lines could be useful for the study of Mcm2 biology, or evolutionary pressures in cancer. For instance, CNA analysis of cell lines selected for resistance to clinically relevant chemotherapy agents (for instance, vinca alkaloids or alkylating agents), could be used to identify genes and pathways important for this acquired resistance.

Approximately 26% of Mcm2Δcre/Δcre NHD13+ developed concurrent BCP-ALL and pre-T LBL. CNA analysis of BCP-ALL was in general simpler than the pre-T LBL, with fewer deletions per sample. Pax5 and Ptpn1 were recurrently deleted, and these deletions were invariably accompanied by Nup214-Abl1 fusion similar to that seen in a subset of human pre-T LBL and BCP-ALL (31, 32). The Nup214-Abl1 fusions were functional, as shown by Crkl phosphorylation, and sensitivity to the imatinib. WES showed no recurrent Tier 1 SNV, consistent with the notion that driver mutations in Mcm2 hypomorph mice were primarily CNAs. Taken together, the complementary pathways involved in this model match those predicted for human BCP-ALL(50). In this context, the collaborative pathways and genes are: 1) increased stem cell self-renewal conferred by the NHD13+ transgene, 2) impaired B cell differentiation caused by Pax5 deletion, and 3) hyperproliferation conferred by a Nup214-Abl1 fusion gene (Supplemental Figure 14B). Although the role of Ptpn1 deletion is unclear, it is interesting to note Ptpn1 deletion accelerates lymphoma onset in Trp53 deficient mice(51), conceivably by enforcing active kinase signaling(52), and deletions of the closely related PTPN2 are common in pre-T LBL patients with NUP214-ABL1 fusion. In both cases, the PTPN deletion is thought to be oncogenic through enforcing hyperactive kinase signaling (52, 53).
Taken together, these findings demonstrate that replicative stress can result in cancer through the generation of chromosomal rearrangements, most commonly interstitial deletions of ~50-1000 kb. The genetic lesions selected in vivo (Dnmt3a, Pten, Tcf3, and Notch1 for pre-T LBL and Pax5 and Abl1 for BCP-ALL) are frequent events in the corresponding human lymphoid leukemias, reinforcing the idea that collaborative pathways leading to lymphoid leukemias are similar between mice and humans. Combinations of CNAs involving tumor suppressor genes or oncogenes are selected due to fitness advantage, and provide an in vivo mammalian model for evolution and selection within a time frame of months.
Methods

Mice and Genotyping

*Mcm2*<sup>cre/cre</sup> and *NHD13* transgenic mice were generated as previously reported; congenic *Mcm2* on a C57Bl6 background mice were generated as described (6, 9, 14). The genotyping of *NHD13* and *Mcm2* mice were performed as previously described (6, 9, 14) with primers listed in Supplemental Table S16. Complete blood count (CBC) were performed on peripheral blood using a HEMAVET Multispecies Hematology Analyzer (CDC Technologies). Diagnosis of hematologic malignancy was based on previously published consensus guidelines (54).

Flow Cytometry, IHC and Immunoblots

Flow cytometry was performed as described previously (29) with the following antibodies: from eBiosciences: Mac1(CD11b)-PE(12-0112-82), CD4-PE(12-0043-82), Gr1-FITC(11-5931-85), Ter119-FITC(11-5921-85), e-Kit (CD117)-FITC(11-1171-82), CD43-FITC(11-0431-82), CD44-FITC(11-0441-82), CD25-PE(12-0251-82), CD4-APC (17-0042-82) and CD8-APC-780(47-0081-82); from BD Biosciences: CD8-FITC(553031), CD71-PE(553267), Sca-1-PE(553108), B220 (CD45R)-FITC(553088) and CD19-PE(553786). IHC and immunoblotting were performed as previously described (29). Formalin-fixed paraffin-embedded tissue sections were stained with hematoxylin and eosin (H&E), MPO (A0398; Dako), CD3(MCA1477; Bio-Rad) and B220 (553086; BD Biosciences). Stained sections were scanned and imaged as previous report (29). For immunoblots, proteins were separated by 7.5% SDS PAGE gel (Bio-Rad Laboratories) and transferred to nitrocellulose membrane (Thermo Scientific). Primary antibodies used were anti-Mcm2 (610701, Transduction Laboratories), anti-PTEN (D4.3) XP® (#9188, Cell Signaling Technology), anti-Phospho-CrkL (Tyr207) (#3181, Cell Signaling Technology), anti-Cleaved Notch1 (Val1744) (D3B8) (#4147, Cell Signaling Technology), anti-beta-Actin(A5316, Sigma-Aldrich) and anti-α-Tubulin (#2125S, Cell Signaling Technology). After application of appropriate
secondary antibodies conjugated to HRP, signals were visualized using Pierce ECL Western Blotting Substrate (Thermo Scientific) and Amersham Hyperfilm ECL (GE Healthcare Ltd.).

**Cell Culture and Cell Lines**

Pre-T LBL(2875, 2854, 2880, 2730, 2773, 2696, 2795, 2869, 2883, 2641, 2739 and 2973 TCL) cell lines were established from single-cell suspensions prepared from thymus (1 × 10^6 cells) of sick mice and maintained in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 15% FBS, 100 mmol/L l-glutamine, and 100 μg/mL penicillin/streptomycin (Invitrogen) without supplemental cytokines.

Splenocytes from mouse #2725 which expressed a Nup214-Abl1 fusion were maintained in RPMI 1640 medium supplemented with 10% heat-inactivated FBS, 1 mM l-glutamine, 100 U/ml streptomycin, 100 μg/ml penicillin, 50 μM 2-ME(all from Gibco), and 10 ng/ml IL-7(217-17; Peprotech).

**Assessment of Compound E and Imatinib Mesylate treatment**

Compound E (ALX-270-415; Enzo Life Science) was dissolved in DMSO and evaluated at a final concentration of 1μM. 5×10^4/ml pre-T LBL cell lines were seeded in IMDM medium supplemented with 15% FBS, 100 mmol/L l-glutamine, and 100 μg/mL penicillin/streptomycin. The cells were treated with 1μM Compound E or vehicle (DMSO) only for 4 days, and cell number was determined by trypan blue exclusion on a daily basis (TC20™ Automated Cell Counter, Bio-Rad Laboratories, Inc.). Imatinib Mesylate (S1026; Selleckchem.com) was dissolved in DMSO and evaluated at 1, 5 and 10 μM. 5×10^4/ml splenocytes from mouse # 2725 were seeded in RPMI 1640 medium supplemented with 10% heat-inactivated FBS, 1 mM l-glutamine, 100 U/ml streptomycin, 100 μg/ml penicillin, 50 μM 2-ME, and 10 ng/ml IL-7. The cells were treated with Imatinib Mesylate or vehicle only for 4 days, and cell number was determined by trypan blue exclusion on a daily basis.

**PCR and Sanger sequencing**

Genomic DNA was extracted with the DNeasy® Blood & Tissue (Qiagen) kit according to the manufacturer’s recommended protocol. PCR was performed using HiFi Taq polymerase mix (10790-020;
Invitrogen) and primers (Invitrogen) as listed in Supplemental Table S16. Clonal *Igh* and *Tcrb* DJ or VDJ segments were identified using previously described PCR-based assays(29) with primers listed in Table S16. RNA was extracted using TRIzol (Invitrogen) and the manufacturer's recommended protocol. cDNA was synthesized by reverse transcriptase using 1 μg RNA with SuperScript III enzyme and reagents (Invitrogen). cDNA splice sites for Notch1, Ikzf1, and Nup214-Abl1 were identified by PCR amplification of cDNA and confirmed by Sanger sequencing. Selected mutations identified by WES or LCC-seq were PCR amplified, purified, and confirmed by Sanger sequencing. Real-time quantitative PCR was performed with a *Ptpn1* Taqman primer-probe (Mm00448427_m1, ThermoFisher Scientific) and *Hes1* Taqman primer-probe (Mm01342805_m1, ThermoFisher Scientific) sets and ABI Fast Universal PCR Master Mix on the ABI Fast7500 system (Applied Biosystems/Life Technologies). Samples were normalized to endogenous 18S rRNA with TaqMan Ribosomal RNA Control Reagents. Sanger sequencing was performed at the NCI Sequencing MiniCore facility.

**Sparse WGS for CNA**

Preparation of genomic libraries for Sparse WGS was performed as previously described(19). In brief, 1ug of genomic DNA was sonicated using the Covaris instrument followed by end repair and A-tail addition. TruSeq dual index adaptors were subsequently ligated to DNA molecules, with ligated products enriched via PCR amplification. Indexed libraries were pooled and sequenced in multiplex fashion while targeting ~ 4 million reads per sample. Data analysis was performed as described previously(55) with the exception that new, higher resolution, bin-boundaries were constructed to facilitate the analysis. Specifically, mouse genome build mm9 was divided into 120k bins, while accounting for unique mappability. This allowed analysis of the CNAs at a segment resolution of ~125kb (five consecutive bins for segmentation) and breakpoint resolution of ~25kb (average width of each bin).

**WES, WGS and LCC-Seq sequencing and sequencing analysis**

Library Preparation
Sequencing libraries were prepared through tagmentation using the Nextera DNA Library Kit (Illumina, 
Inc.) according to the manufacturer’s instructions with the following modifications. PCR primers in the 
kit were replaced with i5 primer (5’-AATGATACGGCGACCACCAGATCTACACNNNNNNNTCGGCAGCGTC-3’) and i7 
primer(5’-CAAGCAGAAGACGGCATACGAGATNNNNNNNGTCTCGTGGGCTCGG-3’), at a final 
concentration of 50nM, where the N’s represent standard Nextera index sequences. Exome and Targeted 
Genomic CaptureExome capture was carried out using the SureSelect XT Mouse All Exon, 49.6Mb Kit 
(Agilent Technologies, Inc.) following the manufacturer’s protocol. BAC clones tiling the genomic 
regions of interest (Supplemental Table S17) were obtained (BACPAC Resources Center) and DNA was 
prepared using the QIAGEN Large-Construct Kit (Qiagen). BACs were pooled and capture bait prepared 
using biotinylated random hexamers and Klenow enzyme in a random primed extension reaction. Bait 
coupled to streptavidin magnetic beads was hybridized to library pools for 24 hrs followed by stringency 
washes and PCR using primers homologous to the outermost adapter sequences of the sequencing 
libraries (5’-AATGATACGGCGACCACC-3’ and 5’-CAAGCAGAAGACGGCATACGAGAT-3’). Sequencing 
Captured libraries were quantified by qPCR (KAPA Biosystems) and sequenced on a HiSeq 2500 System 
(Illumina, Inc.).

Sequencing analysis

Data processing and variant calling procedure followed the Best Practices workflow recommended by the 
Broad Institute(56, 57). Briefly, the raw sequencing reads were mapped to mouse genome build 10 
(mm10) by the Burrows-Wheeler Aligner(58)followed by local realignment using the GATK suite(59) 
from the Broad Institute and duplicated reads were marked by the Picard tools 
(http://broadinstitute.github.io/picard/).

Somatic variant calling was performed by comparison of tumor to wild type(WT) samples using the 
MuTect2 (60) somatic variant caller in the GATK suite. SnpEff(61)variant annotation and effect
LUMPY(62) was used for structural variant discovery in the WGS and LCC-Seq.

Filtering criteria for the somatic variants in the WES data: variants were first filtered with the GATK recommended filtering criteria (https://software.broadinstitute.org/gatk/documentation/article.php?id=3225) and then filtered with the following additional filters: (1) Minimum fraction of altered reads in a tumor is 0.3; (2) Minimum number of altered reads in a tumor is 2; (3) Minimum log_fisher is 0.2; (4) Impact effect is ‘High’ or ‘Moderate’ for Tier 1 indels and missense mutation; (5) Exclude SNPs reported in dbSNP build 137 or previously identified as germline variants in the NIH C57Bl6 colony.

Filtering criteria for the structure variants in the WGS data: (1) ALT=DEL; (2) SVLEN <= -1000; (3) Normal AO=0; (4) Manual check if deletions are present in the bam files.

Filtering criteria for the structure variants in the LCC-Seq data: (1) Only keep SVs overlapping the bait regions; (2) Maximum number of altered reads in the WT sample is 4; (3) Minimum fraction of altered reads in a tumor is 0.1; (4) Manual check if deletions are present in the bam files. Primary NGS sequence data is available at SRA, accession numbers PRJNA565491, PRJNA565494, and PRJNA565492.

Analysis of repeat sequences associated with tumor breakpoints

The breakpoints mapped in recurrent deletions using WGS and LCC-Seq are extended 100bp on either side- 100bp into the deletion and 100 bp into the region intact in the tumor. Each deletion is associated with 2 break points. The DNA sequence of this 200bp window was obtained using bedtools getfasta command(63). Tandem Repeat Finder (TRF) was used to locate the mononucleotide repeat sequences around each breakpoint(64).

For each dataset (WGS and LCC-Seq), a random dataset of 200bp window was generated using bedtools shuffle-chrom command to maintain a similar chromosome distribution of the random dataset as tumor deletion dataset(63). The DNA sequence and repeat sequence are mapped as explained earlier.
For the tumor deletion and random dataset, the enrichment of mononucleotide repeats was evaluated using Fisher’s Exact Test for Counts Data using R. The tumor deletion dataset showed a significant enrichment of mononucleotide repeats with respect to the random dataset.

**Statistics**

Data are displayed as mean ± standard deviation. Significance values were calculated using Student’s t-Test (non-paired, one-tailed distribution) in Excel sheet. The number of independent experiments, statistical tests, and P values and are indicated in figure legends. Survival was analyzed using GraphPad Prism 7.01 software and log-rank (Mantel-Cox) test. P < 0.05 was considered significant.

**Study approval**

All studies involving mice were approved by the National Cancer Institute (Bethesda) Intramural Animal Care and Use Committee and were performed according to protocols approved by the National Cancer Institute (Bethesda) Intramural Animal Care and Use Committee.
Conflicts of interest

PDA receives royalties from the NIH Technology Transfer Office for the invention of NHD13 mice.
Author contributions

MY, PDA and SCP conceived and designed the project. MY, TB, RLW, SWL and PDA developed the methods. MY, TB, RLW, SWL, AF and TM performed experiments. MY, TB, YJZ, SS, AN, PSM and PDA analyzed the data. MY wrote the initial draft of the manuscript. MY, TB, SCP, AN, TM, YJZ, SS, RLW, SWL, PSM and PDA reviewed and edited the manuscript.
Acknowledgments

The authors thank current and former members of Aplan lab for insightful discussions. We also thank the Cold Spring Harbor Laboratories sequencing facility, the NCI Sequencing Minicore for Sanger sequencing, the NCI Transgenic Core for generation of transgenic mice, the NCI Flow cytometry core for cell sorting, Shelley Hoover, Mark Simpson and Jennifer E. Dwyer of the NCI Molecular Pathology Unit for assistance with slide imaging, and Maria Jorge for excellent animal husbandry. This work was supported by the Intramural Research Program of the National Cancer Institute, National Institutes of Health (to PSM and PDA, grant numbers ZIA SC0130378, SC010379, BC010982), the William C. and Joyce C. O’Neil Charitable Trust, Memorial Sloan Kettering Single Cell Sequencing Initiative (to TB), and NIH R01 CA130995 (to SCP). SWL is an investigator for the Howard Hughes Medical Institute and the Geoffrey Beene Chair of Cancer Biology.


Figure 1. *Mcm2cre/creNHD13* mice do not develop AML. (A) Survival curve for all six genotypes, compared by log-rank test. *P<0.05;**P<0.01;***P<0.001. (B) Representative flow cytometry profile of malignant thymocytes from *Mcm2cre/creNHD13* and *Mcm2cre/creNHD13* mice with pre-T LBL. (C) Immunophenotype summary of pre-T LBL samples from thymus of *Mcm2cre/creNHD13* (n=33) and *Mcm2cre/creNHD13* (n=45) mice.
Figure 2. *Mcm2*^{cre/cre} mice show recurrent deletions. (A) Whole genome view of copy number alteration (CNA) analysis for germline, pre-T LBL (*Mcm2*^{cre/cre}NHD13^{−} (n=11) and *Mcm2*^{cre/cre}NHD13^{+} (n=18) mice) and AML (*Mcm2*^{cre/wt}NHD13^{+} (n=8)); samples were >70% tumor tissue (thymus or BM) based on flow cytometry. Mouse chromosomes 1-19 and X are indicated. (B-C) Zoomed in view of (B) *Pten* (chromosome 19) and (C) *Bcl7a* (chromosome 5) regions. Red indicates gain; blue indicates loss. Copy number loss is proportional to color; darker blue is consistent with homozygous loss, lighter blue suggests heterozygous loss. * indicate the gain or deletions present in germline samples compared to reference C57Bl6 genome (mm9).
Figure 3. Stable and ongoing CNAs in Mcm2crecre pre-T LBL. CNAs of trios from mouse 2875, 2854, 2880, 2730, 2773, 2696, 2795, 2869, 2883, and 2641 primary pre-T LBL tumor (PTT), pre-T LBL cell line at early passage (1-1.5 months; TCE) and late passage (2-6 months; TCL). Color code as in Figure 2A.
Figure 4. Interstitial deletions lead to activation of Notch1. (A) Summary of Notch1 deletions. Notch1 (ENSMUSG00000026923) exons are indicated in blue. Functional domains (EGF-like ligand binding domain, HD, and ICN) as shown. Deleted exons of Notch1, frameshift mutations, missense mutations and in-frame insertion are indicated. (B) Aberrant Notch1 mRNA splice forms in 2880,2963,2748, 2973,2008, 2897 thymus and 2773 TCL. RT-PCR for WT (WT thy1), 2880,2963, 2008, 2897 thymus and 2773 TCL were performed by using primers located in exon 1 and 29 of Notch1 to detect the junction of exon 2 and exon 28. RT-PCR for WT(WT thy2), 2748 and 2973 thymus were performed by using primers located in exon 13 and 29 of Notch1 to detect the junction of exon 15 and exon 28. The lanes were grouped from different gels as indicated by the vertical lines. (C) ICN expression in pre-T LBL primary tumor (upper) and pre-T LBL cell line (lower). The lanes were grouped from different gels as indicated by the vertical lines. (D) Real-time-PCR analysis of Notch1 target Hes1 expression in WT thymus and pre-T LBL with Notch1 mutations. Error bars represent standard deviation of three technical replicates (SD; n=3).
Figure 5. *Mcm2*<sup>cre/cre*NHD13*<sup> mice develop BCP-ALL. (A) Flow cytometry plots of BM and spleen (SP) stained with CD19 and B220, thymocytes (thy) stained with CD4 and CD8 from *Mcm2*<sup>cre/cre*NHD13*<sup> mice with BCP-ALL. WT BM, SP and thy are controls. (B) Hematoxylin and eosin (H&E), B220 and CD3 immunohistochemistry of infiltrated BM, thymus and kidney of mouse 2725 with concurrent BCP-ALL and pre-T LBL. Scale bar, 200 μm. (C) May-Grunwald Giemsa (MGG) stained BM lymphoblasts from mouse with BCP-ALL (#2811). Scale bar, 50 μm. (D) IgH gene rearrangements assay for #2697 BM; Clonal fragments indicated by a red asterisk were sequenced. (E) Frequency of leukemia subtypes by genotype (*Mcm2*<sup>cre/cre*NHD13*<sup> (n=33), *Mcm2*<sup>cre/cre*NHD13*<sup> (n=46), *Mcm2<sup>wt/wt*NHD13*<sup> (n=18) and *Mcm2<sup>cre/wt*NHD13*<sup> (n=41)).
Figure 6. CNA analysis of BCP-ALL. (A) CNA analysis of germline and BCP-ALL ($Mcm2^{cre/cre}NHD13^*$ (n=7) and $Mcm2^{cre/w}NHD13^*$ (n=4) mice); samples were >60% tumor tissue(BM) based on flow cytometry. The ID of $Mcm2^{cre/cre}NHD13^*$ mice are indicated on the right side. Color code as in Figure 2A. (B-C) CNA for Pax5 region (B) and Cebpb/Ptpn1 region (C). Color code as in Figure 2A. (D) PCR amplification of Ptpn1 deleted region. Faint signals in 2703 and 2725 could be due to haplo-insufficiency or contamination with non-malignant cells. (E) Ptpn1 mRNA expression, samples and copy number of Ptpn1(CNA) are indicated. Error bars represent standard deviation of three technical replicates (SD; n=3). mRNA expression is markedly decreased in samples with 2 copies lost.
Figure 7. Recurrent Nup214-Abl1 fusion gene detected in BCP-ALL. (A) CNAs for the Abl1 and Nup214 containing region of Chr2. Color code as in Figure 2A. (B) Schematic for putative Nup214-Abl1 fusion gene produced via episome or tandem duplication. (C) RT-PCR detection of Nup214-Abl1 fusion mRNA using Nup214 exon 29 forward primer and Abl1 exon 2 reverse primer. Pre-T LBL samples are from mice with genotype Mcm2$^{cre/cre}$NHD13$^+$ and Mcm2$^{cre/cre}$NHD13$^-$, BCP-ALL samples from mice with genotype Mcm2$^{cre/cre}$NHD13$^-$, and AML samples from mice with genotype Mcm2$^{cre/wt}$NHD13$^+$. B: Bone marrow; T: Thymus; LN: Lymph node; SP: Spleen. (D) RT-PCR does not show Nup214-Abl1 fusion mRNA in BCP-ALL from Mcm2$^{cre/wt}$NHD13$^+$ mice; #2842 is positive control. BM: Bone marrow; Thy: Thymus; LN: Lymph node; SP: Spleen. (E) Phosphorylation of Crkl in the spleen of mice (WT, 2842, 2703, 2725, 2905, 2811, 2888 and 2799) with Nup214-Abl1 fusion; v-Abl transformed B cell line is a positive control. (F) Imatinib treatment inhibited the growth of 2725 splenocytes which express a Nup214-Abl1 fusion. Growth curves of 2725 splenocytes treated with Imatinib for 2 days. Cell number was counted by trypan blue exclusion. Error bars represent standard deviation of three technical replicates (SD; n=3).
Table 1 Summary of deletion patterns in *Mcm2crecre* pre-T LBL cell lines.

<table>
<thead>
<tr>
<th></th>
<th>PTT</th>
<th>-</th>
<th>-</th>
<th>-</th>
<th>+</th>
<th>+</th>
<th>+</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCE</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>TCL</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total with indicated pattern</td>
<td>90</td>
<td>22</td>
<td>10</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

+ indicates deletion present, - indicates deletion not present.
Table 2 Total number of AML and BCP-ALL in mice with indicated genotype.

<table>
<thead>
<tr>
<th></th>
<th>$Mcm2^{wt/wt}$ NHD13$^+$</th>
<th>$Mcm2^{cre/ht}$ NHD13$^+$</th>
<th>$Mcm2^{cre/cre}$ NHD13$^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AML</td>
<td>14</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>BCP-ALL</td>
<td>2</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>*p Value (vs $Mcm2^{wt/wt}$ NHD13$^+$)</td>
<td>N/A</td>
<td>1</td>
<td>0.00001</td>
</tr>
</tbody>
</table>

* calculated by Fisher’s exact test.
<table>
<thead>
<tr>
<th>ID</th>
<th>Nup214ex32-Abl1ex2</th>
<th>Nup214ex31-Abl1ex2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2905</td>
<td>CAAATAAAAACAGAAGCCCTGCA</td>
<td></td>
</tr>
<tr>
<td>2697</td>
<td></td>
<td>GCTGCCCAAAGCCCTGCA</td>
</tr>
<tr>
<td>2811</td>
<td>CAAATAAAAACAGAAGCCCTGCA</td>
<td></td>
</tr>
<tr>
<td>2725</td>
<td>CAAATAAAAACAGAAGCCCTGCA</td>
<td></td>
</tr>
<tr>
<td>2703</td>
<td>CAAATAAAAACAGAAGCCCTGCA</td>
<td></td>
</tr>
<tr>
<td>2842</td>
<td>CAAATAAAAACAGAAGCCCTGCA</td>
<td></td>
</tr>
<tr>
<td>2944</td>
<td>CAAATAAAAACAGAAGCCCTGCA</td>
<td></td>
</tr>
</tbody>
</table>