Published in final edited form as:

Genes Chromosomes Cancer. 2017 June ; 56(6): 472–483. doi:10.1002/gcc.22451.

Somatic mutations in murine models of leukemia and lymphoma: disease specificity and clinical relevance

Liat Goldberg¹, Sheryl M Gough¹, Fan Lee¹, Christine Dang¹, Robert L. Walker¹, Yeulin J. Zhu¹, Sven Bilke¹, Marbin Pineda¹, Masahiro Onozawa², Yang Jo Chung¹, Paul S. Meltzer¹, and Peter D. Aplan¹

¹Genetics Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda

²Center for Medical Education/Department of hematology, Hokkaido University Graduate school of Medicine Hokkaido, Japan

Abstract

Malignant transformation is a multistep process that is dictated by acquisition of multiple genomic aberrations that provide growth and survival advantage. During the post genomic era, high throughput genomic sequencing has advanced exponentially, leading to identification of countless cancer associated mutations with potential for targeted therapy. Mouse models of cancer serve as excellent tools to examine the functionality of gene mutations and their contribution to the malignant process. However, it remains unclear whether the genetic events that occur during transformation are similar in mice and humans. To address that, we chose several transgenic mouse models of hematopoietic malignancies and identified acquired mutations in these mice by means of targeted re-sequencing of known cancer-associated genes as well as whole exome sequencing. We found that mutations that are typically found in acute myeloid leukemia or T cell acute lymphoblastic leukemia patients are also common in mouse models of the respective disease. Moreover, we found that the most frequent mutations found in a mouse model of lymphoma occur in a set of epigenetic modifier genes, implicating this pathway in the generation of lymphoma. These results demonstrate that genetically engineered mouse models (GEMM) mimic the genetic evolution of human cancer and serve as excellent platforms for target discovery and validation.

Introduction

Genetically engineered animal models of human cancer have proven to be useful tools for investigating cancer initiation, progression, and therapy (Frese and Tuveson 2007; Hanahan, et al. 2007; Smith and Muller 2013). A common strategy employed in these studies is to generate mice that have one or more defined mutation(s) in the germline, and observe these mice for the onset of cancer over a defined study period. In these experiments, it is anticipated that the mutation(s) engineered into the germline have disrupted one or more critical pathways required for malignant transformation, and that complementary pathways

Corresponding author: Dr. Peter D Aplan. 37 Convent drive, Bethesda, MD 20892; Fax: 301-496-0047, Tel: 301-435-5005.

The authors have no competing interests in relation to the work described in this manuscript.

will become activated with time, possibly due to spontaneous somatic mutations. The fact that additional mutations can collaborate with transgenes to induce malignant transformation has been shown by acceleration of disease onset in transgenic mice through the use of mutagens such as *N*-ethyl-*N*-nitrosourea (ENU) or retroviral insertions (Dang, et al. 2015; Huser, et al. 2014).

Investigators have traditionally used two approaches to identify genes and/or pathways that can complement a genetically engineered mutation. A targeted or candidate gene approach studies a limited number of genes that are known or suspected to be involved in malignant transformation, using PCR amplification and Sanger sequencing of target genes. Using this strategy, we have previously identified spontaneous mutations in one third of the leukemias that arose in *NUP98-HOXD13* transgenic mice (Novak, et al. 2012; Slape, et al. 2008). These mutations most commonly involved *Ras* pathway genes, such as *Nras*, *Kras*, *Ptpn11* and *Cbl*. A major limitation of this strategy however is that novel gene mutations would not be identified, as only a limited number of genes is assayed.

An alternate approach involves mutagenizing the genome, and allowing for biological selection to identify a malignant clone. One limitation of this approach is that it does not necessarily mimic malignancies that arise spontaneously, in that the complementary mutations are forced using chemicals or retroviral insertions. However, this approach has clearly been successful, as chemical mutagenesis screens implicated *APC* as a critical gene involved in colon cancer (Moser, et al. 1990), and retroviral insertional mutagenesis identified *Hoxa9*, *Meis1*, *Nf1*, and *Evi1* as genes important for leukemic transformation (Largaespada, et al. 1995; Mucenski, et al. 1988; Nakamura, et al. 1996).

Over the past decade, whole genome sequencing (WGS) has emerged as a major tool for surveying the entire genome in order to identify cancer associated mutations. A complementary method, whole exome sequencing (WES), sequences only the exonic portion of the genome (~3%); these two techniques have revolutionized the search for mutations in cancer genomes (Ley, et al. 2008; Parsons, et al. 2008; Thirman, et al. 1993)

The ability of genetically engineered mice to accurately model human malignancy has been hotly debated (Frese and Tuveson 2007). In this study we used several transgenic mouse models of hematologic malignancies and examined the genetic events that occur during the generation of the disease by targeted sequencing and WES. We studied two different murine models of leukemogenic *NUP98* fusions, *NUP98-HOXD13* and *NUP98-PHF23*, that were established in our lab (Gough, et al. 2014; Lin, et al. 2005). Although *NUP98* fusions have been considered rare events in patients with AML, recent studies show that *NUP98* fusions are present in approximately 5% of all pediatric AML patients, and as many as 15% of “normal karyotype” pediatric AML patients (Bisio, et al. 2016; Gough, et al. 2011; Ostronoff, et al. 2015; Struski, et al. 2016). A third murine model of AML that we studied was generated by expression of a *CALM-AF10* transgene (Caudell, et al. 2007). *CALM-AF10* fusions are associated with AML and immature T-ALL in humans; the AML in both human patients and engineered murine models are characterized by clonal rearrangements of both IG and TCR genes (Asnafi, et al. 2003; Caudell and Aplan 2008; Deshpande, et al. 2006), suggesting that the cell of origin is multipotent. Both *NUP98* fusions, which have

recently been shown to be dependent on *MLL1* for leukemic transformation (Xu, et al. 2016), as well as *CALM-AF10* fusions lead to overexpression of *HOXA7/9/10* (referred to as a “*HOXA*” cluster), which is thought to be a final common pathway leading to poor prognosis AML in both humans and mice (Caudell and Aplan 2008; Collins and Hess 2016; Golub, et al. 1999; Gough, et al. 2011). To broaden the models of hematopoietic malignancy evaluated, we studied a murine model of peripheral T cell lymphoma, not otherwise specified (PTCL-NOS) that was induced by overexpression of *Lin28b* (Beachy, et al. 2012). Overexpression of *LIN28B*, or the closely related *LIN28A*, is seen in over 50% of patients with PTCL-NOS (Beachy, et al. 2012). In this report, we show that certain gene mutations are highly recurrent in both human and mouse leukemia, validating that these mouse models are relevant to the human disease. These results also provide insight into gene pathways that are recurrently involved in specific malignancies and may serve as attractive targets for therapeutics.

Methods

Genetically engineered mouse models (GEMMs)

Construction of transgenic mice expressing *NUP98-HOXD13* (*NHD13*), *NUP98-PHF23* (*NP23*), *CALM-AF10* (*CA10*) and *Lin28b* and the diagnoses of malignancies that arise in these mice have been described elsewhere (Beachy, et al. 2012; Caudell, et al. 2007; Gough, et al. 2014; Lin, et al. 2005).

Multiplex targeted re-sequencing

High molecular weight genomic DNA was isolated from tissues by standard salting-out procedures (Chervinsky, et al. 2001) or Qiagen Blood and Tissue DNA Extraction Kit (Qiagen). Regions of interest were targeted by PCR amplification using multiplexed pools of primers tiling whole exons. Targeting primers were tailed with partial Illumina adapter sequences. Amplification and pooling of target amplicons was accomplished in a 48.48 Access microfluidic array on a Fluidigm Biomark System. Primer Wells were loaded with 5µl containing 1X Access Array Loading Reagent and 12-plex pools of primer pairs (6.25µM each). Sample Wells were loaded with 5µl containing 1X Advantage HD Buffer (Clontech), 0.2mM each dNTP, 0.025U/µl Advantage HD Polymerase, 1X Access Array Loading Reagent, 0.5X ROX Reference Dye (Invitrogen), 0.5X EvaGreen Dye (Biotium), and 25ng template DNA. Thermocycler conditions were: 98°C for 3 min., 40 cycles of 98°C for 10 sec., 55°C for 30 sec., 72°C for 15 sec., followed by 72°C for 5 min. Illumina adapter sequences were completed and sample specific indexes were added in a 50µl PCR containing: 1µl 100X diluted amplicon pool, 0.125µM each primer (Index Primer F: 5'-AAT GAT ACG GCG ACC ACC GAG ATCT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT-3', Index Primer R: 5'-CAA GCA AAG ACG GCA TAC GAG ATx xxx xxx xGT GAC TGG AGT TCA G-3' where “x” indicates an eight nucleotide barcode sequence), and 1X Phusion High-Fidelity PCR Master Mix with HF Buffer (New England Biolabs). Thermocycler conditions were: 95°C for 30 sec., 18 cycles of 95°C for 10 sec., 65°C for 30 sec., 72°C for 30 sec., followed by 72°C for 5 min. Multiplex PCR Pools were sequenced on an Illumina Genome Analyzer IIx.

The raw sequences were aligned to MM9 mouse genome using BWA version 0.7.12-r1039. Reads not aligning to the location expected by the PCR reaction were removed from further analysis. From the remaining sequences, the known, constant primer sub-sequences were removed and alignment positions adjusted accordingly. The resulting BAM files were analyzed using the Strelka version 1.0.14 for somatic variant calling.

Whole exome capture and sequencing analysis

Mouse Illumina DNA libraries were prepared and captured using the Agilent SureSelectXT Mouse All Exon Kit according to the manufacturer's instructions. In brief, DNA was fragmented on a Covaris S1 sonicator followed by end repair and phosphorylation. Blunt fragments were adenylated, ligated to Illumina Y-adapters and PCR amplified. Bait hybridization proceeded for 48 hours, followed by recovery of captured exome fragments by PCR. Captured exomes were sequenced on an Illumina HiSeq 2000.

Data processing and variant calling procedure mainly followed the Best Practices workflow recommended by the Broad Institute (<http://www.broadinstitute.org/gatk/guide/best-practices>). Briefly, the raw sequencing reads were mapped to mouse genome build 10 (mm10) by the Burrows-Wheeler Aligner (BWA, (Li and Durbin 2009)) followed by local realignment using the GATK suite from Broad Institute and duplicated reads were marked by Picard tools (<http://picard.sourceforge.net>).

Somatic variant calling was performed by comparison of tumor to wild type samples by the Strelka somatic variant caller (Saunders, et al. 2012) and germline variant calling was done with the UnifiedGenotyper from the Broad Institute (<https://www.broadinstitute.org/gatk/>). SnpEff (Cingolani, et al. 2012) and dbSNP 137 (NCBI) were used to annotate and predict effects of the variants.

The following filtering criteria were used for germline variation calls: (1) minimum read depth is 5; (2) minimum altered read number is 3; (3) minimum fraction of altered reads is 0.01; (4) impact is 'High' or 'Moderate'. The following filtering criteria were used for somatic variation calls: (1) minimum read depth is 5; (2) maximum altered read number in reference samples ('Normal') is 2 or maximum fraction of altered reads in reference samples ('Normal') is 0.01.

Additional filtering is described in the results.

Validation of sequencing results

Genomic DNA from paired tumor-normal (tail or kidney) tissues was used to PCR amplify regions where variants were identified by targeted re-sequencing or WES. Amplicons were purified and Sanger sequenced at the NCI CCR genomics core. To validate expression of variants, RNA was extracted from tumor tissue using TRIzol reagent according to the manufacturer's guidelines (Ambion). cDNA was synthesized using SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen) and then PCR amplified and sequenced.

Results

Targeted re-sequencing of mouse hematologic malignancies identifies acquired mutations in genes commonly mutated in human cancer

We have previously characterized four GEMMs of hematologic malignancy that develop a wide spectrum of disease, including myelodysplastic syndrome (MDS), acute myeloid leukemia (AML), T-cell acute lymphoblastic leukemia T-ALL, B-cell acute lymphoblastic leukemia (B-ALL), and peripheral T-cell lymphoma (PTCL) (Beachy, et al. 2012; Caudell, et al. 2007; Gough, et al. 2014; Lin, et al. 2005). These models include mice that express *NHD13*, *NP23*, *CA10*, or *Lin28b* (Table S1) transgenes driven by *Vav1* regulatory elements, which directs lineage specific expression to the hematopoietic compartment (Ogilvy, et al. 1999). We used targeted re-sequencing to search for spontaneous somatic mutations in a set of genes known to be recurrently mutated in hematologic cancers (Tables S2, S3, (Cancer Genome Atlas Research 2013; Lindqvist, et al. 2016). This gene set included the 14 genes most commonly mutated in human AML (*NPM1*, *FLT3*, *DNMT3A*, *IDH1*, *IDH2*, *NRAS*, *KRAS*, *RUNX1*, *TET2*, *TP53*, *CEBPA*, *WT1*, *PTPN11* and *KIT*), as well as the 4 genes most commonly mutated in T-ALL (*NOTCH1*, *PTEN*, *PHF6* and *FBXW7*), supplemented with 6 additional genes less commonly mutated in human hematopoietic malignancy (*ASXL1*, *CBL*, *JAK2*, *EZH2*, *MPL* and *GATA1*). Genomic DNA from tumor samples and healthy wild type littermates was used for multiplex PCR amplification, barcoding and sequencing. Since the GEMMs were generated in either C57Bl/6 or FVB strains, wild type (WT) controls with matching genetic backgrounds were included in the analyses to eliminate germline polymorphisms. Single nucleotide variants (SNVs) and insertions/deletions (indels) were validated using Sanger sequencing of matching tumor and normal tissues. Out of 24 gene targets, we identified and validated mutations in eight genes: *Flt3*, *Kras*, *Nras*, *Cbl*, *Trp53*, *Ptpn11*, *Idh1* and *Notch1* (Figure 1A, Table S4). Most of the mutations were similar to mutations previously identified in human leukemia. For instance, we found numerous mutations in both *Kras* and *Nras* at amino acid positions 12, 13, 61 and 63, which are frequently mutated in cancer patients and considered mutational hotspots. In most cases (12/18), the variant allele frequencies (VAF) of the *ras* mutations were close to or greater than 0.5, suggesting that this mutation is a relatively early event (Table S4). In addition, several samples had VAF >0.7, suggesting reduction to homozygosity in a tumor subclone. Conversely, most mice with mutations in *Ptpn11*, which encodes a phosphatase known to affect *ras* signaling, had a low VAF (<0.25), suggesting that the mutation was only present in a minor subclone (Table S4).

Targeted re-sequencing of only 24 genes identified spontaneous mutations in more than half of the T-ALL samples and almost half of the AML samples (Figure 1A). The most common acquired mutation in the T-ALL samples involved the heterodimerization (HD) or proline, glutamic acid, serine, threonine-rich (PEST) domain of *Notch1* (Figure 1B), similar to findings in human patients with T-ALL (Tzoneva and Ferrando 2012). In contrast, the most common acquired mutations identified in the murine AML samples involved *Ras* or tyrosine kinase signaling pathways, similar to findings in human AML in general (Scholl, et al. 2008), as well as *NUP98*-rearranged AML specifically (Bisio, et al. 2016; Ostronoff, et al. 2015; Struski, et al. 2016; Thol, et al. 2013).

We have previously shown that *Notch1* PEST domain mutations in murine primary pre-T LBL tumors have a high occurrence at two hot spots; R2361 and S2398 (Lin, et al. 2006). Six of the 14 *Notch1* PEST domain mutants identified in this screen mapped to one of those two “hotspots” for mutation.

Although we identified recurrent missense mutations and one in frame deletion in the *Flt3* kinase domain (KD), we identified no internal tandem duplication (ITD) mutations in mouse AML samples using this multiplex PCR/next-gen sequencing approach. However, *Flt3*-ITD mutations were previously identified in two CA10 mice (Novak, et al. 2012). One of these mice (#7468) was included in this study, however, the assay did not detect the previously documented 39 bp insertion, suggesting that this assay may have missed large indels in *Flt3*. Therefore, to determine if these large insertions might have been missed due to alignment errors, we analyzed 22 NP23 AML samples for *Flt3*-ITD mutations using PCR primers that flank the previously identified murine Flt3-ITD (primers muFlt3LM For and muFlt3LM Rev (Novak, et al. 2012)) and found no ITD mutations (data not shown).

Mutations in epigenetic modifier genes, which are frequently found in AML patients (Naoe and Kiyoi 2013) were typically absent from the mouse leukemias, with the exception of mutations in *Idh1* (Table S4, highlighted). An *Idh1* R132G was initially identified in two primary NHD13 AML samples in addition to two AMLs that arose in mice transplanted from an NHD13 donor that showed no clinical or hematologic evidence of AML (Chung, et al. 2008) (Table S4). In the case of the two primary AML samples, the VAF of *Idh1* mutations was close to 0.5, indicating these mutations were most likely early events, as was recently proposed in the context of AML patients (Chan and Majeti 2013; Shlush, et al. 2014). However, the *Idh1* R132G VAF in the secondary leukemias was 0.25 or 0.36 (Table S4), suggesting that this mutation may have occurred in a subclone which arose at a later stage. This observation suggested that *Idh1* mutations with low VAF might be present in other NHD13 AML samples that did not pass the designated threshold of VAF=0.1. Subsequently, we identified three additional NHD13 AML samples (1138, 1155, and 2420) in which *Idh1* R132G occurred at VAF of 0.08, 0.03 and 0.02 (Figure 2A, Table S4). Using an allele specific PCR assay, we were able to confirm that the *Idh1* R132G mutation was present in two of these samples (1138, 2420, Figure 2B); we suspect the frequency of the *Idh1* mutation in sample 1155 was below the level of detection of the assay.

Whole exome sequencing reveals novel somatic and germline mutations in mouse models of hematologic malignancies

In many (94/154) of the samples analyzed by targeted re-sequencing, no mutations were detected in the selected target genes (Figure 1, Table S1). Of note, only 2 out of 31 Lin28b samples contained a mutation in any of the target genes, suggesting that the selected target genes were not common targets for the Lin28b PTCL model, indicating a need for an unbiased approach to unravel the mutational processes in these malignancies. Therefore, we performed whole exome sequencing (WES) on 17 Lin28b PTCL samples and 18 AML samples from NP23 mice in order to gain new insights into the genetic events leading to these malignancies. These results were analyzed using the approach summarized in Figure 3A.

Tier 1 variants were defined as non-synonymous mutations in coding regions with a VAF of >0.25 compared to a WT sibling. Figure 3B shows that the distribution of Tier I variants in NP23 and Lin28b tumors is very similar, with the median number of variants per sample being 13 for Lin28b and 10 for NP23. These values are similar to those in clinical series of WES for patients with AML or PTCL (Ley, et al. 2008; Palomero, et al. 2014). Using this approach, we identified 40 recurrent variants; 18 in Lin28b and 22 in NP23 samples, respectively (Table S5).

Since our study was based on inbred mice, we initially assumed, as have others (Kode, et al. 2014), that germline mutations would be filtered out by our analysis. However, for 28 of the 40 recurrent variants, identical SNVs were documented in several samples (Table S6). We suspected that these may be germline mutations and selected seven of the most common SNVs to validate by Sanger sequencing of non-malignant tissue. In all seven cases, we found the SNV in both the tumor and the normal tissue from the same mouse, demonstrating that these SNVs represented new germline mutations that had occurred within the colony (Table S6). All of the mutations presented in figure 5 and table 1 have been documented to be acquired mutations by sequencing germline tissue (tailor non-infiltrated kidney) from the same mouse.

Interestingly, one of the germline mutations we identified, a R250* in the ETS homologous factor (*Ehf*) gene, demonstrated a frequent, nearly complete, loss of heterozygosity in tumor tissue (Figure 4A). We sequenced a total of 20 paired NP23 tumor and normal tissues from the same founder line (C10), and found LOH in 12 tumor samples; in all cases, the WT allele was lost, while the mutant R250* allele was retained. An additional seven samples from two other founder lines (B10, G8) did not show an *Ehf* mutation (Figure 4B). Furthermore, NP23 mice with AML that displayed a LOH of the WT *Ehf* allele had a shorter survival rate compared to mice with no LOH (Figure 4C), suggesting that loss of a functional Ehf protein may provide a survival advantage.

Given that the stated criteria of $\text{VAF} > 0.25$ may have eliminated important mutations present in subclonal populations, we expanded our search to include $\text{VAF} > 0.1$ for genes that are known to be important for leukemic transformation. Using this approach, we identified and validated additional variants in relevant genes such as *Hist1h3b*, *Dot1l* and *Asx1l*. In sum, a total of 36 somatic variants were identified by WES with $\text{VAF} > 0.1$ and validated by Sanger sequencing of Lin28b and NP23 paired tumor and normal germline tissue (Table 1, Table S7). Using the same filtering approach, we also examined non-coding SNVs with a $\text{VAF} > 0.25$. We found 66 and 119 unique non-coding SNVs in Lin28b and NP23 samples respectively (Table S8 and S9). Of these, a total of 12 non-coding SNVs were recurrent, identified in 2–13 samples. However, these SNVs were precisely identical across the samples and, similar to the identical coding SNVs we identified, are suspected to be germline mutations within the colony (Table S8 and S9).

Acquired mutations in both models could be placed into several large functional categories (Table 1). Within the Lin28b cohort, a large group of mutations were found in epigenetic regulators, similar to findings in lymphoma patients (Pasqualucci, et al. 2011b). In some cases, we found these mutations to occur at the same position as ones previously reported in

human malignancy, such as in *Hist1h3b*, *Hist1h3d* and *Asx1l* genes, while other mutations affected different residues within the same functional domains as seen in human malignancy (Table 1). It should be noted that the *Hist1h3b* and *Hist1h3d* mutations occur at an amino acid residue (K27) well-known to be a critical residue for epigenetic regulation of transcription (Rice and Allis 2001). The *Asx1l* frameshift mutation was due to an expansion of a G repeat tract (Figure S1); this form of mutation (addition of a G within a G repeat tract encoding a frameshift at codon 642) is the single most common *ASXL1* mutation in MDS and AML patients (Abdel-Wahab, et al. 2011; Gelsi-Boyer, et al. 2009).

According to our study, one Lin28b-driven PTCL sample had acquired a nonsense mutation (Q516*) in *Crebbp* (Table 1), which truncates a large portion of the protein, resulting in loss of both the DNA and protein interacting domains, and the acetyltransferase domain. It seems likely that Q516* results in a loss of function, similar to mutations found in lymphoma and ALL patients. The second Lin28b-driven PTCL sample acquired a *Crebbp* mutation at a splice acceptor site (Chr16: 4107517; AG > AA) upstream of exon 19 (Table 1), and would be predicted to omit exon 19 in the mature transcript. Reverse transcription PCR (RT-PCR) of exons 18–20 revealed two fragments; a wild type-sized allele (~340bp) and a second, smaller fragment (~250bp) consistent with the omission of exon 19 (Figure S2A). Sequence analysis of these two bands revealed 3 splice variants: the first deletes 6 bp of exon 19, the second deletes 18 bp of exon 19, while the third deletes exon 19 completely, resulting in a frame shift and premature stop codon (Figure S2). The amino acids missing in the first two variants are part of a domain of unknown function (DUF902) that is conserved in several transcriptional co-activators (pFAM database). The short mutant isoform encodes a truncated protein, lacking the histone acetyltransferase domain. Since *Crebbp* is suggested to act as a haploinsufficient tumor suppressor (Mullighan, et al. 2011; Pasqualucci, et al. 2011a), the splicing mutation in the Lin28b PTCL may contribute to tumor formation.

In summary, using targeted re-sequencing and whole exome sequencing, we identified and validated acquired mutations in 37 genes (Figure 5). Examination of the recurrent mutations revealed that some events were unique to a specific type of malignancy. This implies that such mutations collaborate with the transgene to drive leukemia/lymphoma in a specific cell context.

Discussion

We studied collaborative mutations using mouse models of hematologic malignancies in which a single genetic aberration is introduced in the mouse germline and spontaneous, complementary mutations accumulate over time to generate a malignancy (Beachy, et al. 2012; Caudell, et al. 2007; Gough, et al. 2014; Lin, et al. 2006). It seems feasible that the disease latency reflects the mutational processes required to generate cancer- the stronger the driver mutation is, the fewer additional mutations are needed and the shorter the disease latency will be. Of note, *NUP98* fusions are likely to be early, initiating events in human AML (Struski, et al. 2016), similar to murine models in which transgenes engineered into the murine germline are initiating events.

We used targeted resequencing of selected candidate genes and WES to identify spontaneous, acquired mutations in murine hematopoietic malignancies. An important observation of this study is that, similar to findings in human patients, specific mutations are recurrently found in specific types of mouse leukemia and lymphoma (Figure 5). For instance, the most frequent mutation we identified in mouse T-ALL involved *Notch1* (Figure 1); 14 of 27 T-ALL cases had *Notch1* mutations in the HD or PEST domains (Table S4). Similarly, over half of T-ALL patients carry activating mutations in *Notch1* (Tzoneva and Ferrando 2012), indicating that *Notch1* is a major driver of both human and mouse T-ALL. Interestingly, we found a *Notch1* PEST domain mutation in a CALM-AF10 AML sample (Table S4) that had a clonal rearrangement of the *Tcrb* and *Igh* loci (Caudell, et al. 2007), suggesting that this leukemia may have emerged from an immature and uncommitted progenitor cell.

A working hypothesis for collaborative mutations in AML suggests that myeloid leukemias have at least two complementary mutations; a “Class I” mutation that regulates cell proliferation and/or survival, and a “Class II” mutation that impairs differentiation or increases stem cell self-renewal (Gilliland and Tallman 2002; Naoe and Kiyoi 2013). Of note, most of the spontaneous somatic mutations identified by targeted re-sequencing in the current study are considered to be “Class I” mutations, such as those involving *Ras* and *Flt3*. Conversely, “Class II” mutations, such as those involving *Npm1*, *Runx1* and *Cebpa*, as well as mutations in key epigenetic modifiers such as *Dnmt3a* and *Tet2*, were not detected in this study. The reason for this mutational pattern is most likely related to the choice of the initiating mutation, namely, the transgene. Expression of CALM-AF10 or Nup98-fusion proteins has been shown to induce expression of the *HoxA* gene cluster, leading to impaired differentiation (Caudell, et al. 2007; Gough, et al. 2014; Lohr, et al. 2012). Therefore, acquisition of additional differentiation-impairing mutations would likely provide little selective advantage, whereas mutations contributing to proliferation and survival are positively selected. The findings in these murine models are similar to those in *NUP98*-rearranged AML, in which the most frequent co-occurring mutations are in *NRAS*, *KRAS*, *FLT3*, *WT1* and *CEBPA*. In addition, mutations in epigenetic modifiers, such as *DNMT3A*, *TET2*, *ASXL1* and *IDH1/2*, as well as *NPM1*, are absent or rare in both human and murine *NUP98*-fusion AML (Bisio, et al. 2016; Ostronoff, et al. 2015; Struski, et al. 2016; Thol, et al. 2013).

An important exception to the above observation is the recurrent *Idh1* mutations in NHD13 AML samples (Figure 2, Table S4). *IDH* genes are recurrently mutated in a wide variety of tumors, including glioblastoma multiforme and AML (Losman and Kaelin 2013; Marcucci, et al. 2010; Parsons, et al. 2008). These recurrent mutations in *IDH1/2* genes lead to hypermethylation of both cytosine residues and histone lysine residues accompanied by a block in differentiation (Figueroa, et al. 2010; Lu, et al. 2012; Waterfall, et al. 2014). In addition, *IDH2* mutations are thought to be early, pre-leukemic events that lead to expansion of hematopoietic stem and progenitor cells (Shlush, et al. 2014). Two NHD13-driven AML samples had clonal *Idh1* mutations (VAF of ~ 0.5), indicating that these were early events in disease development. In contrast, at least four other NHD13-driven AML samples had subclonal *Idh1* mutations, suggesting a late mutational event. Out of 74 AML samples taken from three mouse models (NHD13, NP23 and CA10), targeted re-sequencing identified *Idh1*

mutations in seven NHD13 samples, six of which were validated (Figure 2, Table S4). The exclusivity of these mutational events suggests that NHD13 and mutant *Idh1* may collaborate in leukemia initiation or maintenance.

FLT3 mutations are common in unselected AML patients, and were recently found to occur in 36–63% of *NUP98*-rearranged AML (Bisio, et al. 2016; Struski, et al. 2016; Thol, et al. 2013). We found *Flt3* mutations in the murine AML samples, but not the murine PTCLs. Interestingly, we found several missense mutations and one in-frame deletion in the TKD of *Flt3*, but no ITD mutations in this sample set. Of note, no ITD was identified in CA10 # 7468, a sample that we previously reported had a 39 bp *Flt3*-ITD identified by PCR amplification and Sanger sequencing (Novak, et al. 2012). This finding highlights a limitation of NGS in ascertainment of large indels. As mentioned above, although uncommon, *Flt3*-ITD mutations can be acquired in mice (Novak, et al. 2012), and a *Flt3*-ITD collaborates potently with NHD13 to cause leukemia in a murine model (Greenblatt, et al. 2012). Although speculative, it may be that the murine *Flt3* locus is less susceptible to ITD events than the human *FLT3* locus.

In our WES studies, control DNA was obtained from wild type littermates of the inbred transgenic mice. This approach decreases the cost of WES, as WES data is not obtained from germline tissue of all mice with malignancy, but only a single wild-type mouse. However, a limitation of this approach is that a germline mutation within the colony will be mis-interpreted as a somatic, acquired SNV. We suspected that some of the tier 1 mutations may arise from germline mutations within the colony, as precisely identical SNVs were present in numerous independent animals. PCR amplification and Sanger sequence analysis of matched tumor and normal tissues demonstrated that almost all identical, recurrent SNVs were spontaneous germline mutations, such as the *Ehf* mutation in *NP23* AML samples (Figure 4, Table S6). Exceptions to this generalization were found in well-known “hot spot” mutations of established cancer genes, such as *Idh1*, *Flt3*, *Nras*, and *Kras*. Acquired SNVs were previously noted within a colony of congenic mice that had been backcrossed 10+ generations (Wartman, et al. 2011). This and the current report highlight the importance of sequencing non-malignant tissue, even when studying inbred mice, in order to distinguish somatic mutations from germline mutations that can arise within a mouse colony. Of note, some germline mutations are associated with predisposition to cancer and may be functionally relevant (Godley 2014; Noetzli, et al. 2015).

WES of murine Lin28b PTCL samples revealed recurrent mutations in genes that are involved in epigenetic regulation of transcription, such as *Hist1h3b*, *Hist1h3d*, *Asx11*, *Mll2*, *Dot11*, *Ezh2* and *Crebbp* (Table 1). Mutations in epigenetic modifiers have been widely reported in patients with hematopoietic malignancies, and, in the case of *Asx11* and histone genes, are identical to those found in murine Lin28b PTCL (Abdel-Wahab, et al. 2011; Gelsi-Boyer, et al. 2009; Rice and Allis 2001).

In summary, parallels between the mutational evolution of mouse and human malignancies supports the sequencing of mouse tumors as a discovery tool for collaborative mutations relevant to human disease. In this study, we identified at least one likely collaborative mutation in >60% of our samples. We have shown that several epigenetic factors are subject

to mutation in Lin28b-driven PTCL, identifying them as candidate genes that may play a role in human PTCL. This important finding indicates that aberrant chromatin remodeling is a critical process during development of hematopoietic malignancies, and highlights epigenetic regulatory proteins as promising targets for therapeutic intervention.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Michael Kuehl, Ross Levine, and current and former members of the Aplan lab for insightful discussions. We also thank the NCI Sequencing core for Sanger sequencing, the NCI Flow Cytometry core for assistance with immunophenotype analysis, the NCI Transgenic Core for generation of transgenic mice, 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.

References

- Abdel-Wahab O, Adli M, LaFave LM, Gao J, Hricik T, Shih AH, Pandey S, Patel JP, Chung YR, Koche R, Perna F, Zhao X, Taylor JE, Park CY, Carroll M, Melnick A, Nimer SD, Jaffe JD, Aifantis I, Bernstein BE, Levine RL. ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer Cell*. 2012; 22(2):180–193. [PubMed: 22897849]
- Abdel-Wahab O, Kilpivaara O, Patel J, Busque L, Levine RL. The most commonly reported variant in ASXL1 (c.1934dupG;p.Gly646TrpfsX12) is not a somatic alteration. *Leukemia*. 2010; 24(9):1656–1657. [PubMed: 20596031]
- Abdel-Wahab O, Pardanani A, Patel J, Wadleigh M, Lasho T, Heguy A, Beran M, Gilliland DG, Levine RL, Tefferi A. Concomitant analysis of EZH2 and ASXL1 mutations in myelofibrosis, chronic myelomonocytic leukemia and blast-phase myeloproliferative neoplasms. *Leukemia*. 2011; 25(7):1200–1202. [PubMed: 21455215]
- Asnafi V, Radford-Weiss I, Dastugue N, Bayle C, Leboeuf D, Charrin C, Garand R, Lafage-Pochitaloff M, Delabesse E, Buzyn A, Troussard X, Macintyre E. CALM-AF10 is a common fusion transcript in T-ALL and is specific to the TCRgammadelta lineage. *Blood*. 2003; 102(3):1000–1006. [PubMed: 12676784]
- Attieh Y, Geng QR, Dinardo CD, Zheng H, Jia Y, Fang ZH, Ganani-Gomez I, Yang H, Wei Y, Kantarjian H, Garcia-Manero G. Low frequency of H3.3 mutations and upregulated DAXX expression in MDS. *Blood*. 2013; 121(19):4009–4011. [PubMed: 23660862]
- Beachy SH, Onozawa M, Chung YJ, Slape C, Bilke S, Francis P, Pineda M, Walker RL, Meltzer P, Aplan PD. Enforced expression of Lin28b leads to impaired T-cell development, release of inflammatory cytokines, and peripheral T-cell lymphoma. *Blood*. 2012; 120(5):1048–1059. [PubMed: 22723554]
- Berger SL. The complex language of chromatin regulation during transcription. *Nature*. 2007; 447(7143):407–412. [PubMed: 17522673]
- Bisio V, Zampini M, Tregnago C, Manara E, Salsi V, Di Meglio A, Masetti R, Togni M, Di Giacomo D, Minuzzo S, Leszl A, Zappavigna V, Rondelli R, Mecucci C, Pession A, Locatelli F, Basso G, Pigazzi M. NUP98-fusion transcripts characterize different biological entities within acute myeloid leukemia: a report from the AIEOP-AML group. *Leukemia*. 2016
- Cancer Genome Atlas Research N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013; 368(22):2059–2074. [PubMed: 23634996]
- Caudell D, Aplan PD. The role of CALM-AF10 gene fusion in acute leukemia. *Leukemia*. 2008; 22(4):678–685. [PubMed: 18094714]
- Caudell D, Zhang Z, Chung YJ, Aplan PD. Expression of a CALM-AF10 fusion gene leads to Hoxa cluster overexpression and acute leukemia in transgenic mice. *Cancer Res*. 2007; 67(17):8022–8031. [PubMed: 17804713]

- Chan SM, Majeti R. Role of DNMT3A, TET2, and IDH1/2 mutations in pre-leukemic stem cells in acute myeloid leukemia. *Int J Hematol*. 2013; 98(6):648–657. [PubMed: 23949914]
- Chervinsky DS, Lam DH, Zhao XF, Melman MP, Aplan PD. Development and characterization of T cell leukemia cell lines established from SCL/LMO1 double transgenic mice. *Leukemia*. 2001; 15(1):141–147. [PubMed: 11243382]
- Chung YJ, Choi CW, Slape C, Fry T, Aplan PD. Transplantation of a myelodysplastic syndrome by a long-term repopulating hematopoietic cell. *Proc Natl Acad Sci U S A*. 2008; 105(37):14088–14093. [PubMed: 18768819]
- Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012; 6(2):80–92. [PubMed: 22728672]
- Collins CT, Hess JL. Deregulation of the HOXA9/MEIS1 axis in acute leukemia. *Curr Opin Hematol*. 2016; 23(4):354–361. [PubMed: 27258906]
- Dang J, Wei L, de Ridder J, Su X, Rust AG, Roberts KG, Payne-Turner D, Cheng J, Ma J, Qu C, Wu G, Song G, Huether RG, Schulman B, Janke L, Zhang J, Downing JR, van der Weyden L, Adams DJ, Mullighan CG. Pax5 is a tumor suppressor in mouse mutagenesis models of acute lymphoblastic leukemia. *Blood*. 2015; 125(23):3609–3617. [PubMed: 25855603]
- Deshpande AJ, Cusan M, Rawat VP, Reuter H, Krause A, Pott C, Quintanilla-Martinez L, Kakadia P, Kuchenbauer F, Ahmed F, Delabesse E, Hahn M, Lichter P, Kneba M, Hiddemann W, Macintyre E, Mecucci C, Ludwig WD, Humphries RK, Bohlander SK, Feuring-Buske M, Buske C. Acute myeloid leukemia is propagated by a leukemic stem cell with lymphoid characteristics in a mouse model of CALM/AF10-positive leukemia. *Cancer Cell*. 2006; 10(5):363–374. [PubMed: 17097559]
- Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumar A, Fernandez HF, Tallman MS, Sun Z, Wolniak K, Peeters JK, Liu W, Choe SE, Fantin VR, Paietta E, Lowenberg B, Licht JD, Godley LA, Delwel R, Valk PJ, Thompson CB, Levine RL, Melnick A. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell*. 2010; 18(6):553–567. [PubMed: 21130701]
- Frese KK, Tuveson DA. Maximizing mouse cancer models. *Nat Rev Cancer*. 2007; 7(9):645–658. [PubMed: 17687385]
- Gelsi-Boyer V, Trouplin V, Adelaide J, Bonansea J, Cervera N, Carbuccia N, Lagarde A, Prebet T, Nezri M, Sainty D, Olschwan S, Xerri L, Chaffanet M, Mozziconacci MJ, Vey N, Birnbaum D. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. *Br J Haematol*. 2009; 145(6):788–800. [PubMed: 19388938]
- Gilliland DG, Tallman MS. Focus on acute leukemias. *Cancer Cell*. 2002; 1(5):417–420. [PubMed: 12124171]
- Godley LA. Inherited predisposition to acute myeloid leukemia. *Semin Hematol*. 2014; 51(4):306–321. [PubMed: 25311743]
- Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*. 1999; 286(5439):531–537. [PubMed: 10521349]
- Gough SM, Lee F, Yang F, Walker RL, Zhu YJ, Pineda M, Onozawa M, Chung YJ, Bilke S, Wagner EK, Denu JM, Ning Y, Xu B, Wang GG, Meltzer PS, Aplan PD. NUP98-PHF23 is a chromatin-modifying oncoprotein that causes a wide array of leukemias sensitive to inhibition of PHD histone reader function. *Cancer Discov*. 2014; 4(5):564–577. [PubMed: 24535671]
- Gough SM, Slape CI, Aplan PD. NUP98 gene fusions and hematopoietic malignancies: common themes and new biologic insights. *Blood*. 2011; 118(24):6247–6257. [PubMed: 21948299]
- Greenblatt S, Li L, Slape C, Nguyen B, Novak R, Duffield A, Huso D, Desiderio S, Borowitz MJ, Aplan P, Small D. Knock-in of a FLT3/ITD mutation cooperates with a NUP98-HOXD13 fusion to generate acute myeloid leukemia in a mouse model. *Blood*. 2012; 119(12):2883–2894. [PubMed: 22323452]

- Hanahan D, Wagner EF, Palmiter RD. The origins of oncomice: a history of the first transgenic mice genetically engineered to develop cancer. *Genes Dev.* 2007; 21(18):2258–2270. [PubMed: 17875663]
- Huser CA, Gilroy KL, de Ridder J, Kilbey A, Borland G, Mackay N, Jenkins A, Bell M, Herzyk P, van der Weyden L, Adams DJ, Rust AG, Cameron E, Neil JC. Insertional mutagenesis and deep profiling reveals gene hierarchies and a Myc/p53-dependent bottleneck in lymphomagenesis. *PLoS Genet.* 2014; 10(2):e1004167. [PubMed: 24586197]
- Klinakis A, Lobry C, Abdel-Wahab O, Oh P, Haeno H, Buonamici S, van De Walle I, Cathelin S, Trimarchi T, Araldi E, Liu C, Ibrahim S, Beran M, Zavadil J, Efstratiadis A, Taghon T, Michor F, Levine RL, Aifantis I. A novel tumour-suppressor function for the Notch pathway in myeloid leukaemia. *Nature.* 2011; 473(7346):230–233. [PubMed: 21562564]
- Kode A, Manavalan JS, Mosialou I, Bhagat G, Rathinam CV, Luo N, Khiabani H, Lee A, Murty VV, Friedman R, Brum A, Park D, Galili N, Mukherjee S, Teruya-Feldstein J, Raza A, Rabadan R, Berman E, Kousteni S. Leukaemogenesis induced by an activating beta-catenin mutation in osteoblasts. *Nature.* 2014; 506(7487):240–244. [PubMed: 24429522]
- Largaespada DA, Shaughnessy JD Jr, Jenkins NA, Copeland NG. Retroviral integration at the Evi-2 locus in BXH-2 myeloid leukemia cell lines disrupts Nf1 expression without changes in steady-state Ras-GTP levels. *J Virol.* 1995; 69(8):5095–5102. [PubMed: 7609078]
- Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M, Cook L, Abbott R, Larson DE, Koboldt DC, Pohl C, Smith S, Hawkins A, Abbott S, Locke D, Hillier LW, Miner T, Fulton L, Magrini V, Wylie T, Glasscock J, Conyers J, Sander N, Shi XQ, Osborne JR, Minx P, Gordon D, Chinwalla A, Zhao Y, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson M, Baty J, Ivanovich J, Heath S, Shannon WD, Nagarajan R, Walter MJ, Link DC, Graubert TA, DiPersio JF, Wilson RK. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature.* 2008; 456(7218):66–72. [PubMed: 18987736]
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009; 25(14):1754–1760. [PubMed: 19451168]
- Lin YW, Nichols RA, Letterio JJ, Aplan PD. Notch1 mutations are important for leukemic transformation in murine models of precursor-T leukemia/lymphoma. *Blood.* 2006; 107(6):2540–2543. [PubMed: 16282337]
- Lin YW, Slape C, Zhang Z, Aplan PD. NUP98-HOXD13 transgenic mice develop a highly penetrant, severe myelodysplastic syndrome that progresses to acute leukemia. *Blood.* 2005; 106(1):287–295. [PubMed: 15755899]
- Lindqvist CM, Lundmark A, Nordlund J, Freyhult E, Ekman D, Carlsson Almlof J, Raine A, Overnas E, Abrahamsson J, Frost BM, Grander D, Heyman M, Palle J, Forestier E, Lonnherholm G, Berglund EC, Syvanen AC. Deep targeted sequencing in pediatric acute lymphoblastic leukemia unveils distinct mutational patterns between genetic subtypes and novel relapse-associated genes. *Oncotarget.* 2016; 7(39):64071–64088. [PubMed: 27590521]
- Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, Sougnez C, Cruz-Gordillo P, Knoechel B, Asmann YW, Slager SL, Novak AJ, Dogan A, Ansell SM, Link BK, Zou L, Gould J, Saksena G, Stransky N, Rangel-Escareno C, Fernandez-Lopez JC, Hidalgo-Miranda A, Melendez-Zajgla J, Hernandez-Lemus E, Schwarz-Cruz y Celis A, Imaz-Rosshandler I, Ojesina AI, Jung J, Pedamallu CS, Lander ES, Habermann TM, Cerhan JR, Shipp MA, Getz G, Golub TR. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc Natl Acad Sci U S A.* 2012; 109(10):3879–3884. [PubMed: 22343534]
- Losman JA, Kaelin WG Jr. What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes Dev.* 2013; 27(8):836–852. [PubMed: 23630074]
- Lu C, Ward PS, Kapoor GS, Rohle D, Turcan S, Abdel-Wahab O, Edwards CR, Khanin R, Figueroa ME, Melnick A, Wellen KE, O'Rourke DM, Berger SL, Chan TA, Levine RL, Mellinghoff IK, Thompson CB. IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature.* 2012; 483(7390):474–478. [PubMed: 22343901]
- Marcucci G, Maharry K, Wu YZ, Radmacher MD, Mrozek K, Margeson D, Holland KB, Whitman SP, Becker H, Schwind S, Metzeler KH, Powell BL, Carter TH, Koltz JE, Wetzler M, Carroll AJ, Baer MR, Caligiuri MA, Larson RA, Bloomfield CD. IDH1 and IDH2 gene mutations identify

- novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol*. 2010; 28(14):2348–2355. [PubMed: 20368543]
- Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, Johnson NA, Severson TM, Chiu R, Field M, Jackman S, Krzywinski M, Scott DW, Trinh DL, Tamura-Wells J, Li S, Firme MR, Rogic S, Griffith M, Chan S, Yakovenko O, Meyer IM, Zhao EY, Smailus D, Moksa M, Chittaranjan S, Rimsza L, Brooks-Wilson A, Spinelli JJ, Ben-Neriah S, Meissner B, Woolcock B, Boyle M, McDonald H, Tam A, Zhao Y, Delaney A, Zeng T, Tse K, Butterfield Y, Birol I, Holt R, Schein J, Horsman DE, Moore R, Jones SJ, Connors JM, Hirst M, Gascoyne RD, Marra MA. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature*. 2011; 476(7360):298–303. [PubMed: 21796119]
- Moser AR, Pitot HC, Dove WF. A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. *Science*. 1990; 247(4940):322–324. [PubMed: 2296722]
- Mucenski ML, Taylor BA, Ihle JN, Hartley JW, Morse HC 3rd, Jenkins NA, Copeland NG. Identification of a common ecotropic viral integration site, Evi-1, in the DNA of AKXD murine myeloid tumors. *Mol Cell Biol*. 1988; 8(1):301–308. [PubMed: 2827004]
- Mullighan CG, Zhang J, Kasper LH, Lerach S, Payne-Turner D, Phillips LA, Heatley SL, Holmfeldt L, Collins-Underwood JR, Ma J, Buetow KH, Pui CH, Baker SD, Brindle PK, Downing JR. CREBBP mutations in relapsed acute lymphoblastic leukaemia. *Nature*. 2011; 471(7337):235–239. [PubMed: 21390130]
- Nakamura T, Largaespada DA, Shaughnessy JD Jr, Jenkins NA, Copeland NG. Cooperative activation of Hoxa and Pbx1-related genes in murine myeloid leukaemias. *Nat Genet*. 1996; 12(2):149–153. [PubMed: 8563752]
- Naoe T, Kiyoi H. Gene mutations of acute myeloid leukemia in the genome era. *Int J Hematol*. 2013; 97(2):165–174. [PubMed: 23359299]
- Noetzi L, Lo RW, Lee-Sherick AB, Callaghan M, Noris P, Savoia A, Rajpurkar M, Jones K, Gowan K, Balduini CL, Pecci A, Gnan C, De Rocco D, Doubek M, Li L, Lu L, Leung R, Landolt-Marticorena C, Hunger S, Heller P, Gutierrez-Hartmann A, Xiayuan L, Pluthero FG, Rowley JW, Weyrich AS, Kahr WH, Porter CC, Di Paola J. Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nat Genet*. 2015; 47(5):535–538. [PubMed: 25807284]
- Novak RL, Harper DP, Caudell D, Slape C, Beachy SH, Aplan PD. Gene expression profiling and candidate gene resequencing identifies pathways and mutations important for malignant transformation caused by leukemogenic fusion genes. *Exp Hematol*. 2012; 40(12):1016–1027. [PubMed: 22885519]
- Ogilvy S, Metcalf D, Gibson L, Bath ML, Harris AW, Adams JM. Promoter elements of vav drive transgene expression in vivo throughout the hematopoietic compartment. *Blood*. 1999; 94(6):1855–1863. [PubMed: 10477714]
- Ostronoff F, Ries RE, Gerbing RB, Marra MA, Yussanne M, Long W, Zong S, Mungall K, Andrew A, Gerhard DS, Smith MA, Patee G, Davidsen TM, Hermida LC, Farrar JE, Auvil JG, Raimondi SC, Hirsch BA, Kolb EA, Gamis AS, Alonzo TA, Meshinchi S. Rearrangements in Nucleoporin Family of Genes in Childhood Acute Myeloid Leukemia: A Report from Children Oncology Group and NCI/COG Target AML Initiative. *Blood*. 2015; 126(23):169.
- Palomero T, Couronne L, Khiabanian H, Kim MY, Ambesi-Impiombato A, Perez-Garcia A, Carpenter Z, Abate F, Allegratta M, Haydu JE, Jiang X, Lossos IS, Nicolas C, Balbin M, Bastard C, Bhagat G, Piris MA, Campo E, Bernard OA, Rabadan R, Ferrando AA. Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas. *Nat Genet*. 2014; 46(2):166–170. [PubMed: 24413734]
- Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, McLendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz LA Jr, Hartigan J, Smith DR, Strausberg RL, Marie SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW. An integrated genomic analysis of human glioblastoma multiforme. *Science*. 2008; 321(5897):1807–1812. [PubMed: 18772396]
- Pasqualucci L, Dominguez-Sola D, Chiarenza A, Fabbri G, Grunn A, Trifonov V, Kasper LH, Lerach S, Tang H, Ma J, Rossi D, Chadburn A, Murty VV, Mullighan CG, Gaidano G, Rabadan R,

- Brindle PK, Dalla-Favera R. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature*. 2011a; 471(7337):189–195. [PubMed: 21390126]
- Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A, Wells VA, Grunn A, Messina M, Elliot O, Chan J, Bhagat G, Chadburn A, Gaidano G, Mullighan CG, Rabadan R, Dalla-Favera R. Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet*. 2011b; 43(9):830–837. [PubMed: 21804550]
- Rice JC, Allis CD. Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Curr Opin Cell Biol*. 2001; 13(3):263–273. [PubMed: 11343896]
- Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012; 28(14):1811–1817. [PubMed: 22581179]
- Scholl C, Gilliland DG, Frohling S. Deregulation of signaling pathways in acute myeloid leukemia. *Semin Oncol*. 2008; 35(4):336–345. [PubMed: 18692684]
- Schwartzentruber J, Korshunov A, Liu XY, Jones DT, Pfaff E, Jacob K, Sturm D, Fontebasso AM, Quang DA, Tonjes M, Hovestadt V, Albrecht S, Kool M, Nantel A, Konermann C, Lindroth A, Jager N, Rausch T, Ryzhova M, Korbel JO, Hielscher T, Hauser P, Garami M, Klekner A, Bognar L, Ebinger M, Schuhmann MU, Scheurlen W, Pekrun A, Fruhwald MC, Roggendorf W, Kramm C, Durken M, Atkinson J, Lepage P, Montpetit A, Zakrzewska M, Zakrzewski K, Liberski PP, Dong Z, Siegel P, Kulozik AE, Zapatka M, Guha A, Malkin D, Felsberg J, Reifemberger G, von Deimling A, Ichimura K, Collins VP, Witt H, Milde T, Witt O, Zhang C, Castelo-Branco P, Lichter P, Faury D, Tabori U, Plass C, Majewski J, Pfister SM, Jabado N. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature*. 2012; 482(7384):226–231. [PubMed: 22286061]
- Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V, Kennedy JA, Schimmer AD, Schuh AC, Yee KW, McLeod JL, Doedens M, Medeiros JJ, Marke R, Kim HJ, Lee K, McPherson JD, Hudson TJ, Brown AM, Yousif F, Trinh QM, Stein LD, Minden MD, Wang JC, Dick JE. Consortium HP-LGP. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature*. 2014; 506(7488):328–333. [PubMed: 24522528]
- Slope C, Liu LY, Beachy S, Aplan PD. Leukemic transformation in mice expressing a NUP98-HOXD13 transgene is accompanied by spontaneous mutations in Nras, Kras, and Cbl. *Blood*. 2008; 112(5):2017–2019. [PubMed: 18566322]
- Smith HW, Muller WJ. Transgenic mouse models--a seminal breakthrough in oncogene research. *Cold Spring Harb Protoc*. 2013; 2013(12):1099–1108. [PubMed: 24298026]
- Struski S, Lagarde S, Bories P, Puiseux C, Prade N, Cuccuini W, Pages MP, Bidet A, Gervais C, Lafage-Pochitaloff M, Roche-Lestienne C, Barin C, Penther D, Nadal N, Radford-Weiss I, Collonge-Rame MA, Gaillard B, Mugneret F, Lefebvre C, Bart-Delabesse E, Petit A, Leverger G, Broccardo C, Luquet I, Pasquet M. NUP98 is rearranged in 3.8% of pediatric AML forming a clinical and molecular homogenous group with a poor prognosis. *Leukemia*. 2016
- Thirman MJ, Gill HJ, Burnett RC, Mbangkollo D, McCabe NR, Kobayashi H, Ziemann-van der Poel S, Kaneko Y, Morgan R, Sandberg AA, et al. Rearrangement of the MLL gene in acute lymphoblastic and acute myeloid leukemias with 11q23 chromosomal translocations. *N Engl J Med*. 1993; 329(13):909–914. [PubMed: 8361504]
- Thol F, Kolking B, Hollink IH, Damm F, van den Heuvel-Eibrink MM, Michel Zwaan C, Bug G, Ottmann O, Wagner K, Morgan M, Hofmann WK, Gohring G, Schlegelberger B, Krauter J, Ganser A, Heuser M. Analysis of NUP98/NSD1 translocations in adult AML and MDS patients. *Leukemia*. 2013; 27(3):750–754. [PubMed: 22929522]
- Tzoneva G, Ferrando AA. Recent advances on NOTCH signaling in T-ALL. *Curr Top Microbiol Immunol*. 2012; 360:163–182. [PubMed: 22673746]
- Wartman LD, Larson DE, Xiang Z, Ding L, Chen K, Lin L, Cahan P, Klco JM, Welch JS, Li C, Payton JE, Uy GL, Varghese N, Ries RE, Hooch M, Koboldt DC, McLellan MD, Schmidt H, Fulton RS, Abbott RM, Cook L, McGrath SD, Fan X, Dukes AF, Vickery T, Kalicki J, Lamprecht TL, Graubert TA, Tomasson MH, Mardis ER, Wilson RK, Ley TJ. Sequencing a mouse acute promyelocytic leukemia genome reveals genetic events relevant for disease progression. *J Clin Invest*. 2011; 121(4):1445–1455. [PubMed: 21436584]

- Waterfall JJ, Killian JK, Meltzer PS. The role of mutation of metabolism-related genes in genomic hypermethylation. *Biochem Biophys Res Commun.* 2014; 455(1–2):16–23. [PubMed: 25111818]
- Wu G, Broniscer A, McEachron TA, Lu C, Paugh BS, Becksfort J, Qu C, Ding L, Huether R, Parker M, Zhang J, Gajjar A, Dyer MA, Mullighan CG, Gilbertson RJ, Mardis ER, Wilson RK, Downing JR, Ellison DW, Zhang J, Baker SJ. St Jude Children’s Research Hospital-Washington University Pediatric Cancer Genome P. Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat Genet.* 2012; 44(3):251–253. [PubMed: 22286216]
- Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, McMichael JF, Schmidt HK, Yellapantula V, Miller CA, Ozenberger BA, Welch JS, Link DC, Walter MJ, Mardis ER, Diersio JF, Chen F, Wilson RK, Ley TJ, Ding L. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med.* 2014; 20(12):1472–1478. [PubMed: 25326804]
- Xu H, Valerio DG, Eisold ME, Sinha A, Koche RP, Hu W, Chen CW, Chu SH, Brien GL, Park CY, Hsieh JJ, Ernst P, Armstrong SA. NUP98 Fusion Proteins Interact with the NSL and MLL1 Complexes to Drive Leukemogenesis. *Cancer Cell.* 2016; 30(6):863–878. [PubMed: 27889185]

Figure 1.

A. Distribution of mutations in mouse T-ALL and AML samples as detected by targeted re-sequencing. Ras pathway genes refers to *Kras*, *Nras* and *Ptpn11*. B. Schematic representation of the recurrent mutations in *Notch1* detected by targeted re-sequencing in mouse T-ALL and PTCL samples.

A

<i>AML mouse model</i>	<i># Samples</i>	<i># Samples with IDH1R132 mutation</i>	<i>Mutation frequency (%)</i>
NHD13	34	7	20.6
NP23	24	0	0
CA10	16	0	0

B

Figure 2.

Detection of *Idh1* mutations in *NHD13* mice. **A.** *Idh1* R132 mutations were detected by targeted re-sequencing. The table summarizes all mutations with over 3 reads and over 1% allele frequency found in *NHD13*, *NP23* and *CA10* mice (Fisher's exact test $p=0.003$ for *NHD13*-driven AML vs non *NHD13*-driven AML). **B.** *Idh1* R132G low frequency mutations were confirmed in mice 1138 and 2420 by use of allele specific PCR. Sensitivity of the assay was measured using serial dilutions of genomic DNA from AML sample of mouse 1135 (standard, VAF=0.62). Total DNA of the standard samples was adjusted to 100 ng using genomic DNA from a WT mouse. NTC, no template control.

Figure 3.

A. Scheme summarizing the filtering approach used for whole exome sequencing data analysis. B. Waterfall plot depicting the distribution of Tier 1 variants in *Lin28b* and *NP23* tumor samples. Horizontal lines represent medians (solid line for *Lin28b*, dashed line for *NP23*).

Figure 4. Germline mutations in the *Ehf* gene with high frequency of LOH. Whole exome sequencing identified a R250* mutation in 8 *NP23* AML samples originating from the C10 founder line. A. Sanger sequencing of both tumor and normal kidney tissue validated that these were germline mutations with loss of heterozygosity of the wild type allele in the tumor tissue. This analysis was performed on 20 *NP23* C10 samples, 6 *NP23* B10 samples and 1 G8 *NP23* sample (B) and showed exclusivity of these mutations in the C10 founder line. C. Survival curve showing shorter survival of *NP23* AML bearing mice carrying the *Ehf* mutation with LOH of the WT allele compared to mice with no LOH.

Figure 5. Summary of the Acquired somatic mutations identified by either targeted resequencing or whole exome sequencing in GEMMs of hematologic malignancies. Oncoprints generated using the cBioPortal oncoprint tool (Gao et al. *Sci. Signal.* 2013 & Cerami et al. *Cancer Discov.* 2012) are presented for AML samples (A), T-ALL samples (B) and PTCL (C). Samples highlighted in Blue were sequenced only by targeted re-sequencing while the rest were sequenced either by WES or by both. Mutations marked green are missense events, and mutations marked black are truncating events.

Table 1

Validated acquired SNVs and indels identified by WES

Sample Name	POS	REF	ALT	Effect	AACChange	Gene	Exon	VAF	Functional class
C57BL6_Lin28b_6297	13:23575884	A	T	Missense	K28M [*]	Hist1h3d	1	0.52	Epigenetic regulation of transcription
C57BL6_Lin28b_6266	13:23752462	A	T	Missense	K28M [*]	Hist1h3b	1	0.222	
C57BL6_Lin28b_6412	15:98851917	G	A	Nonsense	Q2632 [*]	Mll2	32	0.597	
C57BL6_Lin28b_6316	15:98854095	G	A	Nonsense	R1998 [*]	Mll2	30	0.464	
C57BL6_Lin28b_6502	16:4107517	C	T	splicing acceptor mutation Frame shift		Crebbp	19	0.545	
C57BL6_Lin28b_6206	16:4128468	G	A	Nonsense	Q516 [*]	Crebbp	6	0.35	
C57BL6_Lin28b_6220	10:80786409	A	T	Missense	S603C	Dot1l	14	0.197	
C57BL6_Lin28b_6502	11:11754064	T	C	Missense	L160P	Ikzf1	5	0.426	
C57BL6_Lin28b_6554	11:98467122	A	T	Missense	V463E	Ikzf3	8	0.278	
C57BL6_Lin28b_6297	18:78858500	T	G	Missense	K604Q	Setbp1	4	0.431	
C57BL6_Lin28b_6502	6:47532952	C	T	Missense	G613E	Ezh2	15	0.78	Proliferation, survival and cytokine signaling
C57BL6_Lin28b_6206	X:8070752	A	G	Missense	C222R	Suv39h1	3	0.8	
C57BL6_Lin28b_6266	2:153399445	A	AG	Frame shift	639	Asx1l	13	0.2	
C57BL6_Lin28b_6358	4:101156405	C	A	Missense	S1042I	Jak1	22	0.196	
C57BL6_Lin28b_6297	4:101163673	C	A	Missense	V657F	Jak1	14	0.625	
C57BL6_Lin28b_6208	4:148540355	C	A	Missense	P2141Q	Mtor	46	0.529	
C57BL6_Lin28b_6358	4:148550170	C	A	Missense	A2416D	Mtor	53	0.467	
C57BL6_Lin28b_6297	15:61988029	A	T	Missense	S170C	Myc	2	0.81	
C57BL6_Lin28b_6206	3:98147049	C	T	Nonsense	Q2343 [*]	Notch2	34	0.4	
C57BL6_Lin28b_6502	7:141192012	C	T	Missense	G159R	Hras1	5	0.343	

Sample Name	POS	REF	ALT	Effect	AAChange	Gene	Exon	VAF	Functional class
C57BL6_Lin28b_6226	11:4794482	C	T	Missense	R158Q	Nf2	7	0.442	
C57BL6_Lin28b_6206	15:78565454	T	C	Missense	N92S	Rac2	4	0.495	
C57BL6_Lin28b_6180	9:120952926	A	C	Missense	K335T	Cttnb1	7	0.233	
C57BL6_Lin28b_6412	10:19750816	C	T	Missense	T196M	Il20ra	4	0.273	
C57BL6_Lin28b_6226	13:116305440	A	T	Missense	F86I	Isl1	3	0.364	
C57BL6_Lin28b_6226	X:12048473	TC	T	Frame shift	-1229	Bcor	8	0.645	Transcription regulation
C57BL6_Lin28b_6206	11:101523955	C	T	Missense	D1118N	Breal	10	0.288	DNA damage repair
C57BL6_Lin28b_6502	11:84263324	G	T	Missense	R957L	Acaca	21	0.31	Fatty acid synthesis
C57BL6_NP23_653	6:145246771	C	T	Missense	G12D	Kras	2	0.816	Cell proliferation, DNA replication and cytokine signaling
C57BL6_NP23_G8	6:145234351	C	T	Missense	E63K	Kras	3	0.16	
C57BL6_NP23_C10	5:147341238	T	C	Missense	D842G	Flt3	20	0.387	
C57BL6_NP23_709	5:147341238	T	C	Missense	D842G	Flt3	20	0.93	
C57BL6_NP23_G8	5:147341238	T	C	Missense	D842G	Flt3	20	0.11	
C57BL6_NP23_856	18:61129084	A	C	Missense	D804A	Csflr	18	0.241	Translation regulation
C57BL6_NP23_173	X:87890154	C	T	Missense	A28T	Il1rap1l	2	0.432	
C57BL6_NP23_784	17:84186771	C+38nt	C	Frame shift	del at S133	Zfp36l2	2	0.5 [†]	
C57BL6_NP23_C10	17:84186580	A+40nt	A	Frame shift	del at Y210	Zfp36l2	2	0.311	
C57BL6_NP23_961C	2:67512860	G	A	Missense	R1815K	Xirp2	7	0.442	
C57BL6_NP23_824	2:67511677	G	T	Nonsense	E1421 [*]	Xirp2	7	1	Cardiac development, function and disease

^{*} After post-translational removal of the initiator methionine, the mutation is in K27.

[†] VAF estimated by Sanger sequencing. Variant had VAF<0.25 in WES.