Quantitative expression profiling guided by common retroviral insertion sites reveals novel and cell type–specific cancer genes in leukemia

Martin Sauvageau^{1,2}, Michelle Miller¹, Sébastien Lemieux³, Julie Lessard¹, Josée Hébert^{1,4,5}, and Guy Sauvageau^{1,4,5}

¹Molecular Genetics of Stem Cells Laboratory, Institute for Research in Immunology and Cancer (IRIC), Montreal, QC

²Division of Experimental Medicine, McGill University, Montreal, QC

³Functional and Structural Bioinformatics Laboratory, IRIC, Montreal, QC

⁴Leukemia Cell Bank of Quebec and Division of Hematology, Maisonneuve-Rosemont Hospital, Montreal, QC

⁵Department of Medicine, University of Montreal, QC

Abstract

Proviral insertional mutagenesis is a powerful tool for the discovery of cancer-associated genes. The ability of integrated proviruses to affect gene expression over long distances combined with the lack of methods to determine the expression levels of large numbers of genes in a systematic and truly quantitative manner have limited the identification of cancer genes by proviral insertional mutagenesis. Here, we have characterized a new model of proviral insertional mutagenesis-induced lymphoid tumors derived from *Eed* Polycomb group gene mutant mice and quantitatively determined the expression levels of all genes within 100 kb of 20 different retroviral common insertion sites (CISs) identified in these tumors. Using high-throughput quantitative reverse transcription–polymerase chain reaction (Q-RT-PCR), we document an average of 13 CIS-associated genes deregulated per tumor, half of which are leukemia subtype–specific, while the others are coordinately deregulated in the majority of tumors analyzed. Interestingly, we find that genes located distantly from common proviral integration sites are as frequently deregulated as proximal genes, with multiple genes affected per integration. Our studies reveal an unsuspected conservation in the group of genes deregulated among phenotypically similar subtypes of lymphoid leukemias, and suggest that identification of common molecular determinants of this disease is within reach.

Correspondence: Guy Sauvageau, Institute for Research in Immunology and Cancer (IRIC), University of Montreal, CP 6128, Downtown Station, Montreal, QC, Canada H3C 3J7; guy.sauvageau@umontreal.ca.

Authorship

Contribution: M.S. designed and performed research, analyzed data, and wrote the paper; M.M., J.L., and J.H. performed research; S.L. did statistical analysis; and G.S. designed research and helped with writing.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

The online version of this article contains a data supplement.

Introduction

Proviral insertion mutagenesis screens have proven useful tools for identification of oncogenes and tumor suppressors involved in cancer.¹⁻³ With the advent of high-throughput retroviral insertion site cloning, identification of large numbers of common insertion sites (CISs) is readily accessible and holds great promise for elucidation of cancer-promoting pathways.⁴⁻⁹ A CIS is a genomic region where retroviral insertions are found in close proximity in multiple tumors. The assumption is made that such events cannot be accounted for by a random process, but are likely due to a selective advantage conferred to cells by aberrant expression of cancer genes surrounding these viral integrations. Large-scale studies have used statistical parameters to define a CIS.^{6,9} Suzuki et al⁶ established genomic interval cutoffs of 30 kb for 2 integrations, 50 kb for 3 insertions, and 100 kb for 4 or more insertions for a panel of 1200 integrations to label a locus as a CIS.

To date, more than 500 CISs have been identified, and numerous others are expected as new screens are conducted in different mouse models (Retroviral Tagged Cancer Gene Database [RTCGD]).¹⁰ This constitutes an important list of candidate cancer genes and a rich source of information to improve our understanding of the molecular bases of cancer. The full potential of proviral insertional mutagenesis in the identification of oncogenic determinants remains limited by several factors. First, the retrovirus tropism limits the range of tumor types that arise after infection. Proviral insertional mutagenesis screens have mainly used Moloney murine leukemia virus (MMLV) or mammary tumor virus (MTV), which induce hematopoietic and mammary tumors, respectively. Solid tumors, which are more prevalent in human, are more difficult to study by proviral insertion mutagenesis. Recently, the use of a modified MMLV virus encoding the platelet-derived growth factor (PDGF) β -chain allowed for identification of several CISs in brain tumors.¹¹ Transposons have also recently been adapted to insertional mutagenesis screens, which included hematopoietic and solid tumors.¹²⁻¹⁴ Second, the enhancer elements in the proviral long terminal repeats (LTRs) can influence gene expression over large distances,⁴ and may possibly affect several genes at once. The magnitude of this effect has not been extensively studied. Third, retroviral integration is not a random process,^{8,15} and therefore some cancer-associated genes may not be targeted. Inversely, this selectivity in viral integrations could result in incorrect labeling of some genes as “true CISs” contributing to tumor formation. To date, the number of well-documented cases of CIS-associated genes that are aberrantly expressed or involved in tumor development remains relatively small.¹⁶⁻¹⁸ An extensive and quantitative analysis of gene expression near common integration sites would thus be of great value to identify the relevant genes affected. Affymetrix-type expression arrays are of interest, but sensitivity and specificity issues may complicate the analysis. Moreover, the apparently aberrant gene expression detected by this approach has to be validated using quantitative reverse transcription–polymerase chain reaction (Q-RT-PCR). Recent development in microwell high-throughput Q-RT-PCR approaches and systematic Taqman probe set design for all mouse genes provides a new possibility of quantitatively assessing gene expression over large distances around CISs.

As a first test to this approach, we conducted proviral insertion mutagenesis screens for leukemia formation in wild-type and in a *Eed* Polycomb group (PcG) gene sensitized

background where chromatin is expected to be altered. The *Eed*^{PcG} mutant mouse was chosen because heterozygosity for a null allele of *Eed* (*Eed*^{3354/+}) or homozygosity for a hypomorph allele (*Eed*^{1989/1989}) causes the development of hyperproliferative primitive progenitor cells evolving into myeloid and lymphoid diseases, indicating that *Eed* possesses antiproliferative roles in the hematopoietic system.¹⁹ *Eed* mutant mice also show accelerated T-cell tumor development when exposed to *N*-methyl-*N*-nitrosourea.²⁰ The studies presented here constitute the most comprehensive analysis of changes in expression levels of genes surrounding retroviral insertions. They reveal the unpredictable effect of proviral insertional mutagenesis in altering the expression levels of the genes surrounding CISs and document the potential of this approach for elucidating cancer-associated genes activated in hematopoietic tumors.

Methods

Mice and genotyping

Eed^{1989SB} hypomorphic (*Eed*^{hypo}, 101/R1 × C3Hf/R1 hybrid background) and *Eed*^{3354SB} null (*Eed*^{null}) mutant mice have been described before^{21,22} and were maintained as described.²³ *Eed*^{3354SB} (*Eed*^{null}) mutant mice were backcrossed 13 to 15 times in a C57Bl/6J background. Animals were housed and handled in accordance with the guidelines of the Animal Care and Use Committee at the Institut de Recherche Clinique de Montréal (IRCM) and IRIC.

Retroviral insertional mutagenesis

Eed^{hypo/hypo}, *Eed*^{hypo/+} newborn mice and wild-type littermates were injected intraperitoneally with 10⁵ infectious units of MMLV suspended in 50 µL of Dulbecco modified Eagle medium (DMEM).³ Cells from bone marrow, spleen, thymus, and lymph nodes of moribund mice were isolated for DNA and RNA extraction and for histologic analysis.

Histologic staining and morphologic analysis of lymphomas

Morphologic features of leukemic cell populations were examined under 100-fold magnification of Wright stain–dyed cytopspins or touch preparations of bone marrow, spleen, liver, lung, and kidney of diseased mice. Images were acquired using a Leica DMIRB microscope with an HCXPL FluotarL 40×/0.6 numeric aperture objective (Leica, Wetzlar, Germany) and a Retiga EKI camera (Q-Imaging, Burnaby, BC). Images were transformed directly into TIFF files using Adobe Photoshop version 6.0 software (Adobe, San Jose, CA). Percentage of infiltration of blasts in hematopoietic and nonhematopoietic organs was determined for every diseased mouse. For subsequent analyses, tissue samples containing more than 75% blasts were selected (Table S1, available on the *Blood* website; see the Supplemental Materials link at the top of the online article).

Inverse PCR

Amplification of viral insertion sites was carried out using inverse PCR as described by Suzuki et al.⁶ Inverse PCR products were separated on a 1% agarose gel, purified using the QIAEX II Gel Extraction Kit (Invitrogen, Carlsbad, CA), and directly sequenced using

nested primers in the MMLV LTRs at the Montreal Genome Center. Retroviral insertion sites were mapped on the mouse genome (February 2006 assembly) using the BLAT alignment tool from the University of California, Santa Cruz (UCSC) Genome Browser. The primers used for the primary and secondary Inverse PCR and sequencing primers are available online in Document S1.

Statistical significance of CISs

The number of retroviral integrations found for each CIS in *Eed* mutant and wild-type tumors was used to determine the statistical significance of the difference in the number of integrations between both genotypes. The Fisher exact test was applied. The adjusted *P* value was estimated by applying the Benjamini-Hochberg correction to correct for multiple testing.²⁴ A *P* value of .05 or less was considered significant.

Q-RT-PCR

Total RNA was isolated using Trizol according to the manufacturer's protocol (Invitrogen) and treated with DnaseI (Invitrogen) before cDNA synthesis. Reverse transcription of total RNA was performed using the MMLV-RT and random hexamer primers according to the manufacturer's protocol (Invitrogen). The relative quantification of genes was performed by quantitative real-time PCR using 384-well Taqman low-density arrays (Applied Biosystems, Foster City, CA) and custom Taqman assays (designed with the Primer Xpress program from Applied Biosystems) for genes not available with low-density arrays (*Lenep*, *Cks1b*, *BC061194*, and *Prm3*). Amplification efficiency of custom assays was measured over a 5-log dilution range. Primer and probe sequences are available on request. Q-RT-PCR reactions were done on a high-throughput ABI 7900HT Fast Real-Time PCR System (Applied Biosystems) using the method described in the User Bulletin No. 2 (P/N 4303859B).⁴³ Briefly, the Ct (threshold cycle) values of each gene were normalized to the endogenous control gene 18S rRNA (Applied Biosystems; $\rho\text{CT} = \text{Ct}_{\text{target}} - \text{Ct}_{\text{endogenous}}$) and compared with the corresponding wild-type nontransformed tissue (calibrator sample) using the “Ct” method ($\text{CT} = \text{Ct}_{\text{Sample}} - \text{Ct}_{\text{Calibrator}}$). Relative fold difference values for all samples are provided in Table S3.

PCR reactions were carried out using 25 to 50 ng of cDNA and Taqman Universal Master Mix according to the manufacturer's recommendations (Applied Biosystems). Q-RT-PCR cycling conditions were 2 minutes at 50°C and 10 minutes at 95°C, followed by 40 cycles of 15 seconds at 95°C and 1 minute at 59°C. All reactions were done in triplicate. Average values were used for quantification.

To select genes for Q-RT-PCR analyses, the median chromosomal location of each CIS found in 3 or more tumors was determined. The UCSC Genome browser Tables tool was used to identify every gene found within a window of 50 kb upstream and downstream of the median CIS position, according to the RefSeq genes track on the mouse genome February 2006 assembly.

URLs

Further information about the retroviral insertion sites and tumors is available at the Retroviral Tagged Cancer Gene Database (<http://RTCGD.ncicrf.gov>).¹⁰

GenBank accession numbers

Accession numbers for the 321 flanking sequences of the retroviral insertions in the wild-type and *Eed*^{hypo/hypo} tumors are DQ285683 through DQ285997.

Results

Increased incidence of irradiation-induced leukemias in *Eed* mutant mice

To further document the tumor repressive function of *Eed*,²⁰ large cohorts of *Eed*^{hypo/hypo} and *Eed*^{null/+} mice were given a single dose of ionizing radiation (600 cGy) at 5 weeks of age and were monitored for leukemia development (Figure S1). Heterozygous *Eed*^{null/+} and homozygous *Eed*^{hypo/hypo} mutant mice developed T- and B-cell leukemia with higher penetrance and shorter latency than control littermates (Table 1; Figure S1). Interestingly, the median survival time of *Eed*^{null/+} mutants was almost twice as long as homozygous *Eed*^{hypo/hypo} mice. Together, this suggests that loss of function of *Eed* increases susceptibility of cells to oncogenic insult. A detailed characterization of the leukemias observed in these mice is provided in Figure S1 and in Document S1. In agreement with previous genotoxic studies performed in the *Eed* mutant background,²⁰ our studies demonstrate that *Eed* exerts tumor suppressive activity in lymphoid cells.

Retroviral insertional mutagenesis in *Eed* mutant mice

Insertional mutagenesis screens in sensitized backgrounds have been conducted using mice mutant for cell-cycle regulators.^{5,25} However, the impact of a sensitized background for a chromatin modifier with tumor suppressive activity, such as *Eed*, has not yet been evaluated in these screens. Cohorts of heterozygous and homozygous *Eed*^{hypo} mice and wild-type littermates (controls) were therefore injected with MMLV and subsequently monitored for disease occurrence (Figure 1A). In accordance with the irradiation studies (Figure S1), MMLV-infected *Eed*^{hypo/+} mice developed tumors more rapidly than wild-type mice, and loss of the second *Eed* allele further accelerated the rate of tumor onset (Figure 1B; blue survival curve). As shown in Table 1, *Eed*^{hypo/hypo} mice displayed a median survival time of 128 days compared with 151 and 190 days for *Eed*^{hypo/+} and wild-type mice, respectively ($P < .001$). Interestingly, we could not detect loss of heterozygosity in 10 of 10 *Eed*^{hypo/+} tumors analyzed (Figure 1C).

Mice from all 3 genotypes (*Eed*^{hypo/hypo}, *Eed*^{hypo/+} and wild-type) developed T-cell leukemias at approximately the same frequency, varying between 64% and 73%. B-cell leukemias were also observed, but at much lower frequencies (10% *Eed*^{hypo/hypo}, 13% *Eed*^{hypo/+}, and 19% wild-type). Flow cytometric analysis confirmed the T and B lymphoid origin of the tumors (data not shown). Histologic analysis of the hematopoietic tissues of these mice showed infiltration by blastlike lymphoid cells reminiscent of acute lymphoid leukemia (Figure 1D). In most cases, tumor cells had infiltrated the bone marrow, spleen and thymus, liver, lung, and kidney of the animals (Figure 1D; Table S1). Analysis of the T-cell

receptor β chain and the immunoglobulin KC light chain rearrangement patterns by Southern blot indicated that the majority of MMLV-induced lymphomas were monoclonal (data not shown). The phenotype of 5 *Eed^{hypo/hypo}*, 11 *Eed^{hypo/+}*, and 3 wild-type hematopoietic tumors could not be precisely characterized, although histologic staining confirmed death from lymphoid leukemia (Table S1; data not shown).

Identification of CIS in *Eed* leukemia/lymphoma

Inverse PCR was used to locate MMLV integrations in a panel of 25 *Eed^{hypo/hypo}* and 26 wild-type lymphoid tumors (Figures 2A, S2). Mapping by UCSC BLAT searches on the mouse genome revealed 321 retroviral insertion sites (RISs) (Figure S2). To determine which insertions constitute a CIS, we used criteria proposed by Suzuki et al.⁶ When considering a total of 321 insertions, the probability of 2 or more insertions to occur in a window of 41 kb is less than .005. Based on this analysis, 33 CISs were identified (Table 2; Figure 2A), of which 14 were novel, including the *Fgfr3*, *Sh2d3c*, and *Zbtb7b* loci (Table 2). The *Bmi1*, *Zbtb7b*, and *Zfpn1a1* loci were specifically targeted in the *Eed^{hypo/hypo}*-sensitized background (12%–16% of the tumors analyzed), whereas the *Nnp1* and *Sh2d3c* loci were targeted only in wild-type tumors. However, expression of these genes did not differ significantly between *Eed* mutant and wild-type tumors (Figure 2; see next paragraph). The CIS near the *Pim1* gene showed a bias, although not statistically significant, for the *Eed^{hypo/hypo}* genotype (44% *Eed^{hypo/hypo}* tumors vs 12% wild-type; Table 2; see adjusted *P* value). However, *Pim1* overexpression does not cooperate with loss of function of *Eed* in tumor development, because no acceleration of the tumor onset was observed in compound *Pim1* Tg/*Eed^{hypo}* mutant mice compared with *Pim1* transgenic (Tg) mice after more than 400 days (*Pim1* Tg, *n* = 6 mice; *Pim1* Tg/*Eed^{hypo/+}*, *n* = 26 mice; *Pim1* Tg/*Eed^{hypo/hypo}*, *n* = 9 mice; data not shown). For all the other CISs analyzed, the difference between the number of integrations found in *Eed* mutant and wild-type tumors was not statistically significant either (Table 2). It remains possible that some of the identified CISs, like *Zbtb7b* or *Fgfr3*, effectively cooperates with the loss of function of *Eed*. Further functional studies will be required to clarify this possibility. Therefore, we conclude that the frequency and specificity of viral integrations in the genome is not affected by the functional status of *Eed*.

Common and cell type-specific expression of CIS-associated genes in lymphoid tumors

Because viral enhancers can affect gene expression over long distances,⁴ we determined the median genomic location of each CIS identified in 3 or more *Eed^{hypo/hypo}* and wild-type tumors and designed Q-RT-PCR assays for each gene (hereafter called CIS-associated genes; *n* = 48) located within 50 kb of distance from the median integration site (Figure 2B). The basal expression level of 48 CIS-associated genes was first determined in sorted nontransformed T cells (CD4⁺, CD8⁺, and CD4⁺CD8⁺) and B cells isolated from healthy *Eed^{hypo/hypo}* mice (before MMLV infection), and compared with that of wild-type littermates. Results from these studies indicated that the loss of 2 functional alleles of *Eed* does not alter the expression level of any of these CIS-associated genes (Figure 2C; compare wild-type [WT] vs Mutant lanes).

Tissues that were massively (75%–100%) infiltrated with neoplastic cells (Table S1) were also evaluated for expression of CIS-associated genes relative to untransformed (control)

tissues. Analyses of these Q-RT-PCR studies performed in triplicate are presented in Figure 2D (levels are expressed as fold difference compared with untransformed cells; light green and bright red indicates 5- or 10-fold above values determined in control cells shown in Figure 2C, respectively). Strikingly, most tumors contained a large number of aberrantly expressed CIS-associated genes (average = 13 with 5-fold difference). Overall, and in agreement with the results recently obtained by Theodorou et al,²⁶ approximately half (ie, 54%) of the recurrent and proven retroviral integrations lead to detectable deregulation of at least 1 CIS-associated gene (5-fold difference). In contrast, several CIS-associated genes were deregulated in tumors for which no proviral integrations in the corresponding CISs could be identified. For example, proviral integrations in the *Myo5b* gene were found in 7 different tumors (Table 2), of which only 2 (tumors 6Th and 72Th; Figure 2D) showed deregulated expression of this gene. However, significant overexpression of this gene was detected in 10 other tumors, for which no proviral insertions in this region could be detected (for details on RIS and expression levels of genes in each individual tumors, see Tables S2–S3). Similar findings were also reported by Theodorou et al²⁶ for 4 different CIS-associated genes. Overexpression of these genes in absence of nearby retroviral integrations could be caused by the activation of upstream regulators in the same pathway.

As previously mentioned, it is difficult to predict aberrant gene expression based on integration data alone (Figure 3A–C; see MMLV plus signs and arrows in left and right panels and compare impact of expression of the genes for each tumor). Nevertheless, the Q-RT-PCR studies showed that genes located near a CIS are more frequently overexpressed in tumors than genes found in nontargeted regions of the genome (Figures 2D, S3A). The data also show that multiple genes are often deregulated for a given CIS. For the studies presented in Figure 3A,B, 58% and 54% of the tumors aberrantly express 2 or more genes within approximately 100 kb of the *Zbtb7b* (Figure 3A) and *Sh2d3c* (Figure 3B) loci, respectively. For all the other groups of CIS-associated genes tested (Figure 2D; see color code in last row), less than 16% of the tumors showed deregulated expression of 2 or more CIS-associated genes.

Another important finding was that the aberrant expression of the CIS-associated genes is independent of *Eed*'s activity in tumor cells. As shown in Figure 2D, no obvious differences in gene expression patterns were observed between *Eed* mutant and wild-type tumors. However, we observed a remarkable degree of similarity in gene expression patterns within tumor subgroups (eg, CD4⁺ vs CD8⁺; Figure 2D). For example, the *Fgfr3* gene, which is expressed at very low levels in nontransformed T or B cells (CT value > 26; Figure 2C), is highly expressed (up to 1000-fold) in 92% of CD8⁺ tumors, while only 33% of CD4⁺ and no double-positive (T DP) tumors show expression of this gene (Figure 2D). *Eng*, *Myo5b*, and *Thy1* show a similar pattern in nontransformed and transformed pre-B cells (Figure 2D). When arbitrarily considering a more than 5-fold difference threshold as significant, the desintegrin metalloproteinase *Adam15* is overexpressed in 54% CD8⁺ tumors, but not in other lymphoid cell types (Figure 2D; red bars in Figure 3A). Interestingly, the *Adam15* gene is located approximately 45 kb upstream of the median CIS chromosomal location (Figure 3A right). The CIS-associated genes *Zbtb7b* and *Pbxip1* (Figure 3A right) are located at 43 and 86 kb away from *Adam15* and show very different expression patterns. While *Zbtb7b* is down-regulated in most CD4⁺ tumors analyzed (8 of 9; see CD4⁺ in Figure

3A left), its expression is not significantly affected in the majority of CD8⁺ samples (Figure 3A). In contrast, *Pbxip1* is down-regulated (11-fold average) in CD8⁺ (77%) tumors, and to a lesser extent in CD4⁺ tumors (6-fold average in 56% of tumors).

A similar analysis is shown in Figure 3B,C for 2 additional sets of CIS-associated genes, *Sh2d3c/Eng/Ttc16* and *Gfi1/Evi5* (Figure 3B,C, respectively). *Sh2d3c/Eng/Ttc16* distribution around a CIS located between exons 1 and 2 of *Sh2d3c* is shown in Figure 3B (right). All CD4⁺ tumors exhibit an average of 309-fold increase in the tetratricopeptide repeat gene *Ttc16* compared with nontransformed controls, whereas only a fraction (less than 20%) of CD8⁺ tumors show overexpression (less than 9-fold average; red bars in Figure 3B indicate a more than 40-fold difference in expression levels). *Sh2d3c* (targeted in 4 tumors; Figure 3B right) is down-regulated on average 9-fold in 69% of CD8⁺ and 44% of CD4⁺ tumors, whereas *Eng* (Figure 3B white bars) is strongly down-regulated in double positives and in several CD8⁺ tumors. The expression analysis of the *Gfi1/Evi5* group of CIS-associated genes provides further evidence for cell type-specific regulation of gene expression (Figure 3C). For example, *Gfi1* is overexpressed in 88% of CD4⁺ and in less than 4% CD8⁺ tumors, while *Evi5* expression shows the opposite pattern. These data suggest that the cellular context has an important impact on CIS-associated gene expression in the tumors analyzed.

Our studies also led to the identification of genes for which expression is affected in a large spectrum of lymphoid tumors. *Jundm2* is overexpressed in almost every lymphoid tumor analyzed (88%), independently of the subtype (Figure 2C). Although the *Fgfr2* gene is expressed at relatively low levels in tumors (Table S3), it is overexpressed in 54% of CD8⁺ and 67% of CD4⁺ tumors compared with nontransformed cells (Figure 2C). The *Pim2* gene is down-regulated in 92% of CD8⁺, 100% of CD4⁺, and 2 of 3 double-positive (CD4⁺CD8⁺) leukemias. This suggests that these genes may have a broader role in T-cell tumor formation.

We next performed a series of control experiments to determine whether the expression levels of genes not associated with CISs differ between tumor and control cells. For this, we randomly selected 23 genes across the genome, located distantly from known common integration sites, as well as 8 Polycomb (and PcG-related) genes and assessed their expression levels in both nontransformed cells and MMLV-infected lymphoid tumors. Results from these studies showed that 16% (5 of 31) of these randomly selected genes were down-regulated in tumorigenic cells, whereas none of these genes were consistently overexpressed (except for *Ece2*, Figure S3B). This is in contrast with most of the CIS-associated genes tested, which were consistently up-regulated in tumor cells (8 of 50 CIS-associated genes vs 1 of 31 randomly selected genes). According to the Kolmogorov-Smirnov test, this increase in the proportion of CIS-associated genes that are up-regulated in tumorigenic cells is statistically significant ($P < .001$; Document S1; Figure S3C). Importantly, using an alternative threshold-based test, this result was confirmed and shown to be independent from the specific threshold used (Document S1; Figure S3D). Together, these data indicate that heterogeneity between the tumor cells and their corresponding normal counterparts does not explain the global up-regulation in CIS-associated gene expression.

The results presented here identify *Fgfr3*, *Fgfr2*, *Ttc16*, *Adam15*, *Evi5*, *Jundm2*, and *Myo5b* as candidate leukemia-associated genes, some of which are novel. Q-RT-PCR on CIS-associated genes thus reveals a series of candidate leukemia-associated genes and suggests an unsuspected complexity in the regulation of defined gene clusters with oncogenic relevance. Data presented in this report also suggests conservation in the genetic programs leading to lymphoid tumor formation.

Discussion

In this study, we have developed a combinatorial approach which exploits proviral insertional mutagenesis, bioinformatics, and high-throughput Q-RT-PCR for the efficient identification of 7 candidate leukemia-associated genes (*Fgfr3*, *Fgfr2*, *Ttc16*, *Adam15*, *Evi5*, *Jundm2*, and *Myo5b*) from a list of 50 prospects (located within 50 kb in average of common integration sites). The power of this new approach is validated by studies which confirmed implication of at least 4 of these 7 genes in human cancers (*Fgfr3* in multiple myeloma, *Fgfr2* in breast cancer, and *Adam15* in breast, prostate, and stomach cancers)^{27–32} and in the transformation of chicken embryonic fibroblasts (*Jundm2*).³³ Results from our high-throughput expression analysis also revealed an unsuspected complexity in the deregulation of genes at candidate cancer loci, and underline the unpredictable effect of proviral insertional mutagenesis on the expression levels of CIS-associated genes. Most interestingly, we identified genes which deregulation is common to all subtypes of lymphoid leukemias analyzed (*Jundm2*, *Fgfr2*), thus suggesting the existence of shared genetic alterations in these tumors. Finally, we successfully identified 14 new CISs in this study (Table 2). The distribution of insertions found in our wild-type tumors differs substantially from most tumor sets in the RTCGD database. The use of a hybrid (101/C3Hf) background mouse strain, different from other reported proviral insertional mutagenesis screens, likely explains this high number of new CISs identified and the difference in the distribution of the insertions in this study.

The power of insertional mutagenesis in predicting a region containing cancer genes appears relatively high because 95% (19 of 20) of CIS clusters analyzed in this study (Figure 2D right, color-coded groups) contained at least 1 gene that was aberrantly expressed in a large proportion of our leukemias (summarized in Table 3). Moreover, results presented in Figure 3 indicate that several cancer genes are often clustered in a given CIS, and that aberrant expression frequently correlates with tumor type. Thus, a CIS identified in a CD8⁺ leukemia sometimes implies deregulation of a different gene than if identified in a CD4⁺ tumor. Aberrant expression of *Ttc16* in CD4⁺ leukemia and of *Eng* in CD8⁺ leukemia best illustrates this point for the *Sh2d3c* CIS (Figure 3B). These conclusions are obviously dependent on the difference in gene-expression levels between normal and neoplastic cells, and explain our initial focus on lymphoid tumors for which normal (untransformed) counterparts are easily purified. Indeed, the expression levels of the CIS-associated genes did not vary much between the 3 different types of normal T cells (Figure 2C; CD8⁺ vs CD4⁺ and T DP) and pre-B cells, except for the T-cell marker *Thy1*. This kind of analysis might be difficult to perform with myeloid tumors, although recent purification strategies for stem and myeloid progenitors may provide new possibilities for extending these studies.³⁴ When interpreting the gene-expression results, one must also keep in mind that retroviral

insertions can create truncated dominant gain-of-function mutations that might contribute to leukemogenesis without obvious alterations in transcript levels. Point mutations acquired during tumorigenic process can also affect normal gene function. A parallel screen assessing transcript integrity of CIS-associated genes will likely uncover new oncogenes and will be complementary to expression studies to further understand the events leading to leukemia.

The systematic exploitation of quantitative expression arrays as shown herein may set new standards in retroviral insertional mutagenesis screens. Importantly, special care in selecting only tumor samples that are massively infiltrated (75%–100%) with neoplastic cells also helps to prevent contamination from untransformed cells in the expression analyses. In this study, the expression of the CIS-associated genes found in only 2 independent tumors (Table 2) was not analyzed by Q-RT-PCR, even if we knew of the potential importance of some of the genes involved. For example, the histone acetyltransferase *Myst4* gene (also known as *Morf*) is fused to *Cbp* in acute myeloid leukemia,³⁵ and transgenic mice overexpressing the *Ris2* gene (or *Cdt1*) in the absence of p53 develop thymic lymphoblastic lymphoma.³⁶ The role of these “2 hits” CISs is therefore likely to be also important in tumorigenesis. Moreover, extending gene-expression analysis over larger chromosomal distances (> 50 kb in both directions surrounding the CIS) might be informative, because a significant proportion of the retroviral integrations in CISs analyzed in this study was not accompanied by deregulation in gene expression. Extending our protocol to the 510 CISs already identified in the literature should also be informative in revealing additional cancer genes. According to the RTCGD retroviral database, a complete study of all genes encoded within 100 kb of distance of the 35 different CISs containing at least 10 hits would include approximately 120 candidate genes. Replicate analysis in a set of 100 tumors from different phenotypes would thus require 24 000 Q-RT-PCR reactions, approximately 3 times the size of this study. Considering that we identified on average 13 deregulated CIS-associated genes per tumor analyzed in our study, 2 to 3 times as many candidate genes would easily be accessible with such an extended approach.

In our gene-expression analysis, we found that the proportion of genes down-regulated in tumorigenic cells was similar between the CIS-associated genes and the randomly selected genes (5 of 50 of the CIS-associated genes compared with 4 of 31 of the randomly genes, a greater than 5-fold difference). On the contrary, a significantly higher proportion of the CIS-associated genes was up-regulated compared with the randomly selected genes (> 5-fold difference; $P < .001$). The exact nature of this discrepancy remains unclear, but underlies the bias of retroviral insertional mutagenesis screens and of our approach for the identification of oncogenes over tumor suppressor genes.

Using radiation and retrovirus-induced tumor strategies, we confirmed in 2 different sensitized mouse models that *Eed* exerts tumor-suppressive activity in lymphoid tissues. Mutation of only 1 *Eed* allele is sufficient to sensitize cells to transformation. A dominant-negative effect of the mutant protein is unlikely, because it is unstable and undetectable in *Eed*^{null/null} cells.³⁷ Moreover, homozygous mutants have a stronger phenotype and develop tumors more rapidly than heterozygotes. This suggests that gene dosage is important, and that slight changes in *Eed* levels may be sufficient to predispose mice to tumor formation. Similarly, targeted deletion of only 1 allele of the *Suv39h1* H3K9 methyltransferase

predisposes mice to develop B-cell lymphomas.³⁸ Although FACS and histologic analysis of tumors suggests that loss of *Eed* function does not change the phenotype of lymphomas compared with wild-type counterparts, there could be some subtle differences in the differentiation state of the lymphomas that have not been detected in this study. The increased susceptibility to oncogenic transformation in *Eed* mutant cells is not well understood. No specific CISs were identified in *Eed* mutant tumors, suggesting that *Eed* might not sensitize cells toward distinct oncogenic pathways, but instead has a more general effect in cellular transformation. Recently, *Eed* was shown to interact with the DDB1-CUL4 complex, which is involved in cell-cycle progression, replication, and DNA damage repair.³⁹ A link between *Eed*'s main interacting partner, *Ezh2*, and the homologous recombination pathway of DNA repair was also established.⁴⁰ Moreover, EZH2 and SUZ12 interact with PARP3, a protein involved in base excision repair.⁴¹ This suggests that increased tumor development in *Eed* mutants may be due to defects in DNA repair. Further experiments are needed to clarify the mechanism underlying the tumor-suppressive role of *Eed*. Moreover, in the era of chromatin modulation for therapeutic targets, it is postulated that the elimination of the Polycomb genes *Eed*, *Suz12*, or *Ezh2* might be highly efficient to induce apoptosis in transformed cells.⁴² In contrast, the studies presented here document that loss of function of *Eed* sensitizes cells to transformation and therefore raises important concerns about potential therapeutic approaches targeting this gene.

In summary, high-throughput Q-RT-PCR to assess expression of CIS-associated genes is a powerful addition to insertional mutagenesis to identify novel cancer-associated genes. We conclude that extension of this approach to other CIS and tumor subtypes will be instrumental in uncovering novel tissue-specific and common cancer gene networks for the design of better-targeted molecular therapies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are grateful to Dr Neal Copeland for critical reading of the manuscript and to Jana Krosi and Jalila Chagraoui for insightful discussions and comments. Thanks to Pierre Chagnon and Jean Duchaine from the IRIC Genomics and High-throughput platforms, Amélie Faubert for help with the analysis of the irradiation-induced tumors, Melanie Frechette for her expertise and help with the maintenance and manipulation of the animals, Danièle Gagné from the IRIC flow cytometry platform, and Martine Dupuis and Éric Massicotte from the IRCM flow cytometry service for assistance in flow cytometry and cell sorting.

This work was supported by a grant from the Canadian Institutes of Health Research (CIHR) to G.S. and by a grant from the Cancer Research Network of Fonds de la Recherche en Santé du Québec to J.H. M.S. is a recipient of a CIHR Canada Graduate Scholarship Doctoral award. G.S. holds a Canada Research Chair in Molecular Genetics of Stem Cells and is a scholar of the Leukemia & Lymphoma Society of America.

References

1. Uren AG, Kool J, Berns A, van Lohuizen M. Retroviral insertional mutagenesis: past, present and future. *Oncogene*. 2005; 24:7656–7672. [PubMed: 16299527]
2. Neil JC, Cameron ER. Retroviral insertion sites and cancer: fountain of all knowledge? *Cancer Cell*. 2002; 2:253–255. [PubMed: 12398888]

3. van Lohuizen M, Verbeek S, Scheijen B, Wientjens E, van der Gulden H, Berns A. Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. *Cell*. 1991; 65:737–752. [PubMed: 1904008]
4. Lazo PA, Lee JS, Tsichlis PN. Long-distance activation of the Myc protooncogene by provirus insertion in Mlvi-1 or Mlvi-4 in rat T-cell lymphomas. *Proc Natl Acad Sci U S A*. 1990; 87:170–173. [PubMed: 1688653]
5. Lund AH, Turner G, Trubetskoy A, et al. Genome-wide retroviral insertional tagging of genes involved in cancer in Cdkn2a-deficient mice. *Nat Genet*. 2002; 32:160–165. [PubMed: 12185367]
6. Suzuki T, Shen H, Akagi K, et al. New genes involved in cancer identified by retroviral tagging. *Nat Genet*. 2002; 32:166–174. [PubMed: 12185365]
7. Welsh JB, Sapinoso LM, Su AI, et al. Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Res*. 2001; 61:5974–5978. [PubMed: 11507037]
8. Wu X, Li Y, Crise B, Burgess SM. Transcription start regions in the human genome are favored targets for MLV integration. *Science*. 2003; 300:1749–1751. [PubMed: 12805549]
9. Wu X, Luke BT, Burgess SM. Redefining the common insertion site. *Virology*. 2006; 344:292–295. [PubMed: 16271739]
10. National Cancer Institute. [Accessed June 2007] Retroviral Tagged Cancer Gene Database. Available at <http://rtcgd.ncif-crf.gov>
11. Johansson FK, Brodd J, Eklof C, et al. Identification of candidate cancer-causing genes in mouse brain tumors by retroviral tagging. *Proc Natl Acad Sci U S A*. 2004; 101:11334–11337. [PubMed: 15273287]
12. Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature*. 2005; 436:221–226. [PubMed: 16015321]
13. Collier LS, Carlson CM, Ravimohan S, Dupuy AJ, Largaespada DA. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature*. 2005; 436:272–276. [PubMed: 16015333]
14. Dupuy AJ, Jenkins NA, Copeland NG. Sleeping beauty: a novel cancer gene discovery tool. *Hum Mol Genet*. 2006; 15:R75–R79. [PubMed: 16651372]
15. Bushman FD. Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. *Cell*. 2003; 115:135–138. [PubMed: 14567911]
16. Popken-Harris P, Kirchhof N, Harrison B, Harris LF. Gene expression array analyses predict increased proto-oncogene expression in MMTV induced mammary tumors. *Virus Res*. 2006; 119:177–186. [PubMed: 16469401]
17. Recchia A, Bonini C, Magnani Z, et al. Retroviral vector integration deregulates gene expression but has no consequence on the biology and function of transplanted T cells. *Proc Natl Acad Sci U S A*. 2006; 103:1457–1462. [PubMed: 16432223]
18. Erkeland SJ, Verhaak RG, Valk PJ, Delwel R, Lowenberg B, Touw IP. Significance of murine retroviral mutagenesis for identification of disease genes in human acute myeloid leukemia. *Cancer Res*. 2006; 66:622–626. [PubMed: 16423987]
19. Lessard J, Schumacher A, Thorsteinsdottir U, van Lohuizen M, Magnuson T, Sauvageau G. Functional antagonism of the Polycomb-Group genes *eed* and *Bmi1* in hemopoietic cell proliferation. *Genes Dev*. 1999; 13:2691–2703. [PubMed: 10541555]
20. Richie ER, Schumacher A, Angel JM, Holloway M, Rinchik EM, Magnuson T. The Polycomb-group gene *eed* regulates thymocyte differentiation and suppresses the development of carcinogen-induced T-cell lymphomas. *Oncogene*. 2002; 21:299–306. [PubMed: 11803473]
21. Schumacher A, Faust C, Magnuson T. Positional cloning of a global regulator of anterior-posterior patterning in mice. *Nature*. 1996; 384:648. [PubMed: 8984348]
22. Rinchik EM, Carpenter DA. N-ethyl-N-nitrosourea-induced prenatally lethal mutations define at least 2 complementation groups within the embryonic ectoderm development (*eed*) locus in mouse chromosome 7. *Mamm Genome*. 1993; 4:349–353. [PubMed: 8358168]
23. Faust C, Schumacher A, Holdener B, Magnuson T. The *eed* mutation disrupts anterior mesoderm production in mice. *Development*. 1995; 121:273–285. [PubMed: 7768172]

24. Benjamini Y, Hochberg T. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc B*. 1995; 85:289–300.
25. Hwang HC, Martins CP, Bronkhorst Y, et al. Identification of oncogenes collaborating with p27Kip1 loss by insertional mutagenesis and high-throughput insertion site analysis. *Proc Natl Acad Sci U S A*. 2002; 99:11293–11298. [PubMed: 12151601]
26. Theodorou V, Kimm MA, Boer M, et al. MMTV insertional mutagenesis identifies genes, gene families and pathways involved in mammary cancer. *Nat Genet*. 2007; 39:759–769. [PubMed: 17468756]
27. Carl-McGrath S, Lendeckel U, Ebert M, Roessner A, Rocken C. The disintegrin-metalloproteinases ADAM9, ADAM12, and ADAM15 are upregulated in gastric cancer. *Int J Oncol*. 2005; 26:17–24. [PubMed: 15586220]
28. Chesi M, Nardini E, Brents LA, et al. Frequent translocation t(4;14)(p16.3;q32.3) in multiple myeloma is associated with increased expression and activating mutations of fibroblast growth factor receptor 3. *Nat Genet*. 1997; 16:260–264. [PubMed: 9207791]
29. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet*. 2007; 39:870–874. [PubMed: 17529973]
30. Kuefer R, Day KC, Kleer CG, et al. ADAM15 dis-integrin is associated with aggressive prostate and breast cancer disease. *Neoplasia*. 2006; 8:319–329. [PubMed: 16756724]
31. Li Z, Zhu YX, Plowright EE, et al. The myeloma-associated oncogene fibroblast growth factor receptor 3 is transforming in hematopoietic cells. *Blood*. 2001; 97:2413–2419. [PubMed: 11290605]
32. Plowright EE, Li Z, Bergsagel PL, et al. Ectopic expression of fibroblast growth factor receptor 3 promotes myeloma cell proliferation and prevents apoptosis. *Blood*. 2000; 95:992–998. [PubMed: 10648414]
33. Blazek E, Wasmer S, Kruse U, Aronheim A, Aoki M, Vogt PK. Partial oncogenic transformation of chicken embryo fibroblasts by Jun dimerization protein 2, a negative regulator of TRE- and CRE-dependent transcription. *Oncogene*. 2003; 22:2151–2159. [PubMed: 12687017]
34. Akashi K, Traver D, Miyamoto T, Weissman IL. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*. 2000; 404:193–197. [PubMed: 10724173]
35. Murati A, Adelaide J, Mozziconacci MJ, et al. Variant MYST4-CBP gene fusion in a t(10;16) acute myeloid leukaemia. *Br J Haematol*. 2004; 125:601–604. [PubMed: 15147375]
36. Seo J, Chung YS, Sharma GG, et al. Cdt1 transgenic mice develop lymphoblastic lymphoma in the absence of p53. *Oncogene*. 2005; 24:8176–8186. [PubMed: 16261166]
37. Montgomery ND, Yee D, Chen A, et al. The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. *Curr Biol*. 2005; 15:942–947. [PubMed: 15916951]
38. Peters AH, O'Carroll D, Scherthan H, et al. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*. 2001; 107:323–337. [PubMed: 11701123]
39. Higa LA, Wu M, Ye T, Kobayashi R, Sun H, Zhang H. CUL4-DDB1 ubiquitin ligase interacts with multiple WD40-repeat proteins and regulates histone methylation. *Nat Cell Biol*. 2006; 8:1277–1283. [PubMed: 17041588]
40. Zeidler M, Varambally S, Cao Q, et al. The Poly-comb group protein EZH2 impairs DNA repair in breast epithelial cells. *Neoplasia*. 2005; 7:1011–1019. [PubMed: 16331887]
41. Rouleau M, McDonald D, Gagne P, et al. PARP-3 associates with polycomb group bodies and with components of the DNA damage repair machinery. *J Cell Biochem*. 2007; 100:385–401. [PubMed: 16924674]
42. Hormaeche I, Licht JD. Chromatin modulation by oncogenic transcription factors: new complexity, new therapeutic targets. *Cancer Cell*. 2007; 11:475–478. [PubMed: 17560329]
43. [Accessed June 2005] Applied Biosystems. User Bulletin No. 2, P/N4303859B. <http://docs.appliedbiosystems.com/pebiiodocs/4303859B.pdf>

Figure 1. Increased incidence of MMLV-induced lymphomas in *Eed* mutant mice

(A) Schematic representation of the experimental design. (B) Survival curves showing increased incidence of MMLV-induced tumors in *Eed*^{hypo/+} and *Eed*^{hypo/hypo} mutant mice compared with control littermates. (C) Presence of both *Eed*^{hypo} mutant and wild-type alleles in MMLV-induced *Eed*^{hypo/+} tumors. Top shows the hypomorph T > A point mutation at position 1031 of *Eed*. Bar graphs show the number of clones sequenced, containing the *Eed*^{hypo} or wild-type allele, per *Eed*^{hypo/+} tumor. Clones were obtained by amplifying the *Eed* gene from DNA (middle) or RNA (cDNA; bottom) extracted from heterozygous tumors. (D) Wright-Giemsa stainings of representative cytologic preparations of T and B lymphomas and their healthy counterparts from wild-type and *Eed* mutant mice. *No healthy lymph node preparations could be obtained.

Figure 2. Quantitative analysis of gene-expression levels surrounding CIS clusters in *Eed^{hypo/hypo}* and wild-type tumors

(A) Schematic representation of the retroviral insertional mutagenesis screen. Briefly, DNA was extracted from *Eed* mutant or wild-type MMLV-induced lymphomas, and inverse PCR was used to amplify genomic DNA flanking the retroviral integrations. Amplicons were purified, sequenced, and BLATed against the mouse genome. RIS indicates retroviral insertion site; and CIS, common insertion site. (B) Schematic representation of the strategy used to assay expression of genes surrounding a CIS. The median position for every CIS that was found in at least 3 independent MMLV-induced lymphomas was first determined. Genes encompassed in a region spanning 50 kb upstream and downstream of the median CIS (100-kb window) were then assayed for their expression levels by quantitative RT-PCR using Taqman assays. A total of 48 different genes, plus 2 other CISs unrelated to this study (*Pim2*, *Pim3*) were analyzed. (C) Color gradient representation of the Ct values of CIS-associated genes in nontransformed CD4⁺, CD8⁺, CD4⁺CD8⁺ (T DP), or B200⁺IgM⁺ (B) lymphoid cells from healthy mice. Phenotype and genotypes of the analyzed cells are shown on top. Genes analyzed are shown to the right of panel D. * indicates known CISs unrelated to this study. Reactions were done in triplicate, and average values were calculated. Average

Ct values were determined using the 18S rRNA as the internal endogenous control gene to normalize the levels of target gene expression. Low Ct values denote high expression levels. Average Ct values for the 18S rRNA endogenous control was approximately 9. Mutant indicates *Eed^{hypo/hypo}*; WT, wild-type. (D) Relative fold differences in gene

expression levels of *Eed^{hypo/hypo}* mutant and wild-type MMLV-induced lymphomas compared with nontransformed wild-type lymphoid cells. The phenotype, genotype, and ID of the analyzed tumor cells are shown on top. Genes analyzed are shown to the right. Genes are grouped by CIS cluster, indicated by a colored bar beside the gene name. Reactions were done in triplicate, and average values were calculated. Relative fold differences were determined using the $2^{-\Delta\Delta Ct}$ method. The corresponding wild-type nontransformed cells were used as the reference calibrators to assess relative fold differences in expression levels in tumors. Differences of less than -5-fold or more than 5-fold were considered significant. Tumor tissues taken for quantitative RT-PCR contained 75% to 100% lymphoid blasts. T DP indicates CD4⁺CD8⁺; B, B220⁺IgM⁺; Mutant, *Eed^{hypo/hypo}*; and WT, wild-type.

Figure 3. Cell type-specific expression patterns of CIS-associated genes in lymphomas
Relative fold differences in expression levels of genes surrounding a CIS at the *Zbtb7b/Adam15/Pbxip1* (A), *Sh2d3c/Ttc16/Eng* (B), and *Gfi1/Evi5* loci (C) in MMLV-induced lymphoid tumors from *Eed* mutant and wild-type mice. Expression data corresponds to data in Figure 2D. Tumor IDs and phenotypes are shown at the bottom of each graph. * indicates samples with expression level differences higher than 40-fold; and + indicates MMLV retroviral insertions identified by inverse PCR. Positioning of genes and retroviral insertions (arrows with tumor ID) identified at each CIS cluster are depicted on the right panels. Chromosome number and location (in bp) for each CIS cluster are indicated at the bottom of the right panel. B indicates B220⁺IgM⁺; T DP, CD4⁺CD8⁺.

Table 1
Survival of *Eed* mutant and wild-type mice and phenotypic characterization of their lymphomas

Mouse line	Treatment	n	Median survival, d	Penetrance, %	Lymphoma, %			Tumor-free, %*
					B cell	T cell	Not determined	
Wild-type B6	Irradiated	13	504	62	—	23	38	38
<i>Eed^{null/+}</i>	Irradiated	13	338	92	31	38	23	8
Wild-type 101 × C3Hf	Irradiated	16	> 600	19	—	13	6	81
<i>Eed^{hypo/hypo}</i>	Irradiated	11	184	73	9	46	18	27
Wild-type 101 × C3Hf	MMLV	27	190	100	19	70	11	—
<i>Eed^{hypo/+}</i>	MMLV	47	151	100	13	64	23	—
<i>Eed^{hypo/hypo}</i>	MMLV	30	128	100	9	69	22	—

Penetrance is equal to 100% minus the percentage of mice that died of nonhematologic etiology. B cell indicates B-ALL; T cell, T-ALL; Not determined, lymphoma cell type not determined; and —, not applicable.

* Tumor-free mice were killed at 600 days.

Table 2

MMLV CISs identified in *Eed^{hypo/hypo}* and wild-type lymphomas

CIS candidate gene	Protein family	Mouse chromosome	Nb insertions			P	Adjusted P	Human chromosome	Human diseases [*]
			<i>Eed^{hypo/hypo}</i>	Wild-type	Total insertions				
<i>Notch1</i> [†]	Transmembrane receptor	2qA3	18	12	30	.999	.999	9q34.3	T-ALL
<i>Fgfr3</i> [‡]	NA/Tyr kinase receptor	5qB1	6	9	15	.541	.965	4p16.3	Multiple myeloma, T lymphoma
<i>Pim1</i>	Ser/Thr kinase	17qA3.3	11	3	14	.027	.545	6p21.2	NHL, DLCL
<i>Gadd45g</i>	DNA-damage inducible gene	13qA5	6	6	12	.743	.999	9q22.2	—
<i>Prm2</i> [‡]	Histone protein	16qA1	5	4	9	.999	.999	16p13.13	—
<i>Gfi1</i>	Zinc finger protein	5qE5	2	5	7	.419	.914	1p22.1	—
<i>Myo5b</i> [‡]	Myosin protein	18qE2	4	3	7	.703	.999	18q21.1	—
<i>Nup1</i> [‡]	Nuclear protein	10qC1	0	6	6	.023	.545	21q22.3	—
<i>Jundn2</i>	Transcription factor	12qD1	2	4	6	.668	.999	14q24.3	—
<i>Tnepai</i>	Signal transduction	2qH4	3	3	6	.999	.999	20q13.32	—
<i>Myc</i>	Transcription factor	15qD1	2	3	5	.999	.999	8q24.21	Burkitt lymphoma, B-CLL
<i>Thy1</i> [‡]	Membrane glycoprotein	9qA5.2	1	4	5	.350	.914	11q23.3	—
<i>Rai17</i>	SP-RING Zinc finger protein	14qA3	1	4	5	.350	.914	10q22.3	—
<i>Bmi1</i>	Polycomb protein	2qA3	4	0	4	.110	.755	10p12.2	—
<i>Sh2d3c</i> [‡]	Signal transduction	2qB	0	4	4	.110	.755	9q34.11	—
<i>Grap2</i> [‡]	Signal transduction	15qE1	1	3	4	.999	.999	22q13.1	—
<i>Mys4</i> [‡]	Histone acetyltransferase	14qA2	2	2	4	.999	.999	10q22.2	AML
<i>Zbtb7b</i> [‡]	POZ/Krüppel factor/NA	3qF1	3	0	3	.110	.755	1q22	—
<i>Zfpn1a1</i>	Zinc finger protein	11qA1	3	0	3	.110	.755	7p12.2	ALL, DLBL
<i>Brwd2</i>	WD-repeat protein	7qF3	1	2	3	.999	.999	6q14.1	—
<i>Fil1/Ets1</i>	Transcription factor/transcription factor	9qA4	2	0	2	.235	.914	11q24.3	—
<i>Runx3</i>	Transcription factor	4qD3	2	0	2	.235	.914	1p36.11	—
<i>Hrip3</i> [‡]	NA	7qF3	0	2	2	.490	.914	16p11.2	—
<i>Hhex</i>	Homeobox protein	19qC2	0	2	2	.490	.914	10q23.33	—

CIS candidate gene	Protein family	Mouse chromosome	Nb insertions		P	Adjusted P	Human chromosome	Human diseases*
			<i>Eed^{hypo/hypo}</i>	Wild-type				
<i>Idh3/E2F2</i>	Transcription factor/HLH factor	4qD3	0	2	.490	.914	1p36.12	—
<i>Lfng</i>	Glycosyltransferase	5qG2	0	2	.490	.914	7p22.3	—
<i>Pex5</i>	Peroxisomal targeting signal	6qF2	0	2	.490	.914	12p13.31	—
<i>Sup4h2/Bzap1</i>	Transcription initiation factor/NA	11qC	0	2	.490	.914	17q23.2	—
<i>Emi4</i>	WD-repeat protein	17qE4	1	1	.999	.999	2p21	—
<i>Ris2</i>	DNA replication factor	8qE1	1	1	.999	.999	16q24.3	—
<i>Sup6dh</i> [‡]	Transcription initiation factor	11qB5	1	1	.999	.999	17q11.2	—
<i>Zip217</i>	Zinc finger protein	2qH3	1	1	.999	.999	20q13.2	—
<i>Zip644</i> [‡]	Zinc finger protein	5qE5	1	1	.999	.999	1q22.2	—

Pvalue is based on the Fisher exact test. Adjusted Pvalue applies the Benjamini-Hochberg correction.

T-ALL indicates T-cell acute lymphoblastic leukemia; NA, not available; NHL, non-Hodgkin leukemia; DLCL, diffuse large-cell leukemia; B-CLL, B-cell chronic lymphocytic leukemia; AML, acute myeloid leukemia; ALL, acute lymphoblastic leukemia; DLBL, diffuse large B-cell lymphoma; and —, not identified in Human blood diseases.

* Genes implicated in human blood diseases according to the Cancer Gene Census (www.sanger.ac.uk/genetics/CGP/Census).

[‡] Some tumors contained more than 1 MMLV integration in *Notch1*. Frequency of tumors with *Notch1*/MMLV integration: Wild-type, 10 of 26; *Eed^{hypo/hypo}*, 9 of 25. See Table S2 for details.

[‡]New CIS identified in this screen.

Table 3

Percentage of tumors aberrantly expressing 1 or more gene per CIS cluster

CIS cluster ID*	CD8 ⁺ , %	CD4 ⁺ , %	T DP, %	B, %	Total, %
1 (<i>Bmi1</i>)	57.7	100.0	100.0	33.3	68.3
2 (<i>Fgfr3</i>)	92.3	88.9	66.7	66.7	87.8
3 (<i>Pim1</i>)	19.2	55.6	66.7	66.7	34.1
4 (<i>Pim2</i>)	92.3	100.0	66.7	33.3	87.8
5 (<i>Pim3</i>)	3.8	0.0	0.0	0.0	2.4
6 (<i>Sh2d3c</i>)	96.2	100.0	100.0	100.0	97.6
7 (<i>Zbb7b</i>)	96.2	100.0	66.7	100.0	95.1
8 (<i>Jundm2</i>)	88.5	88.9	66.7	66.7	85.4
9 (<i>Gfi1</i>)	34.6	77.8	33.3	100.0	48.8
10 (<i>Nap1</i>)	7.7	0.0	0.0	0.0	4.9
11 (<i>Brwd2</i>)	61.5	66.7	100.0	33.3	63.4
12 (<i>Thy1</i>)	34.6	11.1	33.3	0.0	26.8
13 (<i>Pim2</i>)	19.2	77.8	100.0	100.0	43.9
14 (<i>Notch1</i>)	50.0	55.6	66.7	33.3	51.2
15 (<i>Grap2</i>)	26.9	66.7	66.7	33.3	39.0
16 (<i>Myc</i>)	0.0	55.6	33.3	33.3	17.1
17 (<i>Zfpn1a1</i>)	23.1	11.1	33.3	0.0	19.5
18 (<i>Tnfrsfai</i>)	46.2	44.4	66.7	0.0	43.9
19 (<i>Mysd4</i>)	7.7	33.3	33.3	0.0	14.6
20 (<i>Rai17</i>)	23.1	0.0	33.3	0.0	17.1
21 (<i>Gadd45g</i>)	15.4	44.4	66.7	0.0	24.4
22 (<i>Myc5b</i>)	65.4	33.3	33.3	100.0	58.5

CD8⁺, n = 26; CD4⁺, n = 9; T DP (CD4⁺CD8⁺), n = 3; and B (B220⁺IgM⁺), n = 3.

* Closest gene to median CIS location.