JOURNAL OF VIROLOGY, Sept. 2005, p. 11443–11456 Vol. 79, No. 17
0022-538X/05/$08.00+0 doi:10.1128/JVI.79.17.11443–11456.2005

Radiation Leukemia Virus Common Integration at the *Kis2* Locus: Simultaneous Overexpression of a Novel Noncoding RNA and of the Proximal *Phf6* Gene

Séverine Landais, Renaud Quantin, and Eric Rassart*

Laboratoire de Biologie Moléculaire, Département des Sciences Biologiques, Université du Québec à Montréal, Québec, Canada

Received 14 March 2005/Accepted 1 June 2005

Retroviral tagging has been used extensively and successfully to identify genes implicated in cancer pathways. In order to find oncogenes implicated in T-cell leukemia, we used the highly leukemogenic radiation leukemia retrovirus VL3 (RadLV/VL3). We applied the inverted PCR technique to isolate and analyze sequences flanking proviral integrations in RadLV/VL3-induced T lymphomas. We found retroviral integrations in c-*myc* and *Pim1* as already reported but we also identified for the first time *Notch1* as a RadLV common integration site. More interestingly, we found a new RadLV common integration site that is situated on mouse chromosome X (XA4 region, bp 45091000). This site has also been reported as an SL3-3 and Moloney murine leukemia virus integration site, which strengthens its implication in murine leukemia virus-induced T lymphomas. This locus, named *Kis2* (Kaplan Integration Site 2), was found rearranged in 11% of the tumors analyzed. In this article, we report not only the alteration of the *Kis2* gene located nearby in response to RadLV integration but also the induction of the expression of *Phf6*, situated about 250 kbp from the integration site. The *Kis2* gene encodes five different alternatively spliced noncoding RNAs and the *Phf6* gene codes for a 365-amino-acid protein which contains two plant homology domain fingers, recently implicated in the Börjeson-Forssman-Lehmann syndrome in humans. With the recent release of the mouse genome sequence, high-throughput retroviral tagging emerges as a powerful tool in the quest for oncogenes. It also allows the analysis of large DNA regions surrounding the integration locus.

Cancer is a complex disease that involves a cascade of events and the collaboration of several genes, including oncogenes and tumor suppressors. Several lines of evidence indicate that cell transformation is under the control of a few major pathways (21, 22, 68). Several key oncogenes have been identified for each of these, such as c-*myc* and *ras*. Such key oncogenes are necessary although not enough to induce tumors, and other genes are required. Secondary events such as the activation of tumor progression genes or inactivation of tumor suppressor genes are necessary to maintain the tumoral state and to allow the appearance of the tumor. Many of those collaborating genes remain to be discovered.

Retrovirus-induced tumors are interesting models to study those different pathways. Indeed, retrovirus pathogenesis mimics classical events in the development of a tumor cell but with a shorter latency. The leukemogenic potential of nontransforming retroviruses is a direct consequence of the random integration of their proviral DNA genome in the host cell DNA during the replicative cycle. Retroviral integration sites, when common to several tumors, are considered significantly associated with tumor induction and/or development. As those loci generally contain oncogenes, retroviruses have been widely used over the years as molecular tools to identify cancer-associated genes. Recently, mouse genome sequencing demonstrated powerful use of retroviruses with the emergence of high-throughput retroviral tagging (30, 32, 45, 50, 68). When performed on a transgenic mouse model overexpressing a known oncogene or a mutated tumor suppressor gene, collaborating oncogenes from a particular pathway could be identified (27, 45, 50). All these data are now listed in the Mouse Retroviral Tagged Cancer Gene Database (http://genome2.ncifcrf.gov/RTCGD).

The radiation leukemia virus (RadLV) is a radiation-induced retrovirus that became leukemogenic after serial passages in mice (61). Originally thymotropic, weakly leukemogenic, nondefective, and ecotropic when isolated from X-ray-induced T-cell lymphomas in C57BL/6 mice (40, 42, 60), serial passages in newborn C57BL/Ka mice gave rise to a nonfibrotropic virus preparation with a very high leukemogenic potential (31, 40, 60). Several lymphoid cell lines were established from tumors induced by the passaged RadLV, including the BL/VL_3 cell line (41). Several molecular clones were obtained from the BL/VL_3 cell line as proviruses and several variants were characterized (59, 61). More particularly, the highly leukemogenic molecular clone V-13 provides an interesting T-cell lymphoma model, mostly because the disease latency is less than 3 months (60, 61).

Until now RadLV/VL3 has been rarely used in retroviral tagging. Previous studies report common integration near the cyclin D2 gene (Vin1 locus), *Pim1*, and c-*myc* in 5%, 30%, and 15%, respectively, of tumors (23, 70). *Pim1* and c-*myc* are both known to be involved in T-cell lymphomas (9, 64, 65, 71). Common integrations in the *Kis1* locus on mouse chromosome

* Corresponding author. Mailing address: Département des Sciences Biologiques, Université du Québec à Montréal, Case Postale 8888 Succ. Centre-ville, Montréal, Canada H3C-3P8. Phone: (514) 987-3000, ext. 3953. Fax: (514) 987-4647. E-mail: Rassart.Eric@UQAM.ca.

2 have also been reported, but, so far, no gene has been associated with this locus (35).

In the present study, we used the CpG island technique of Li et al. (39) in an attempt to identify new proto-oncogenes in RadLV-induced T-cell lymphomas. We provide evidence of RadLV integration in the *Notch1* gene. We report the identification of a new common integration site named *Kis2* (Kaplan Integration Site 2) which is located on the mouse X chromosome. Retroviral integration was observed in 11% of the tumors and it modifies the expression of two genes in the region.

MATERIALS AND METHODS

Viruses, mice, and tumors. Several RadLV variants were used to generate leukemias. These are molecular clones from the BL/VL3 tumor cell line (61). The SIM/R and C57BL/6 mouse strains were obtained from Arthur Axelrad (Toronto, Canada) and from Charles River Canada (St-Constant, Quebec, Canada), respectively. Newborn mice were inoculated intraperitoneally with 0.15 ml of filtered virus suspension obtained from independent clones (10^5 to 10^6 virions/ml). Most mice developed tumors within 3 to 5 months postinoculation, and they were sacrificed when obvious lymphoid organ enlargement was detected through palpation. Mice from both strains had a typical enlarged thymus, often accompanied by enlarged lymph nodes and, more rarely, enlarged spleen (61).

Genomic DNA and RNA analysis. Genomic DNA was extracted from frozen normal and leukemic tissues by standard procedures as described previously (63). For Southern analysis, 15 μg of genomic DNA was digested overnight with suitable restriction enzyme, separated on a 0.8% agarose gel, and transferred to a nylon membrane (Osmonics, Fisher, Ottawa, Ontario, Canada).

Total RNA was extracted from frozen normal and leukemic tissues with the TRIzol reagent (Invitrogen, Burlington, Ontario, Canada) according to the manufacturer's protocol. For Northern analysis, 15 μg of total RNA was separated on a 1% formaldehyde agarose gel as described previously (63), and transferred to a nylon membrane (Osmonics).

All the membranes were prehybridized for 1 h at 42°C in 50% formamide-5× SSPE (3 M NaCl, 20 mM EDTA, 20 mM sodium phosphate pH 6.8)-5× Denhardt's-0.5% sodium dodecyl sulfate (SDS),-0.1 mg/ml denatured salmon sperm DNA. Hybridization was performed overnight at 42°C in the same buffer containing 10% dextran sulfate. The probes were labeled by the random primer extension method using oligohexamers (GE Healthcare, Baie d'Urfé, Quebec, Canada), and approximately 2×10^6 cpm/ml were used. The membranes were washed for 10 min at 42°C in 2× SSC (1× SSC is 0.15 M NaCl-0.015 M sodium citrate), 30 min at 42°C in 2× SSC-0.1% SDS, and finally for 15 min at 65°C in 0.1× SSC-0.1% SDS. Membranes were revealed by exposure and scanning with a phosphorimager (Bio-Rad, Mississauga, Ontario, Canada).

RadLV integration site amplification by inverse PCR. The viral insertion site amplification involving inverse PCR was performed essentially according to the method of Li et al. (39). Briefly, 2 μg of tumor DNA was digested with SacII and allowed to circularize with T4 DNA ligase (GE Healthcare). Due to the presence of a SacII site in the RadLV/VL3 *gag* region, DNA molecules that contained a provirus portion and flanking cellular sequences were specifically PCR amplified with a pair of oligonucleotides complementary to the U3 long terminal repeat and the *gag* region upstream of the SacII site, respectively.

Confirmation of retroviral integration at the *Kis2* locus was done by PCR with the *Taq* polymerase kit (QIAGEN, Mississauga, Ontario, Canada) according to the manufacturer's protocol. Briefly, 200 ng of genomic DNA from tumors was used as the template in a 50-μl reaction, with 10 pmol of either the U3 sense RadLV primer (U3624, 5′-TGGGCCCCGGCTCCGTTAGACATA-3′) or the U3 reverse RadLV primer (U3nsens, 5′-GGCTGGGACTTTCCAGAAACTG TTG-3′), and 10 pmol of either *Kis2* genomic primer sense (255.4s, 5′-CATGC GCACTACCTCACAGCTAGAT-3′) or *Kis2* genomic primer reverse (255.4ns, 5′-ATCTAGCTGTGAGGTAGTGCGCATG-3′). The PCR program was one cycle of 94°C for 2 min and 30 cycles of 94°C for 20 seconds, 57°C for 30 seconds, 72°C for 5 min, and one cycle of 72°C for 10 min.

5′ and 3′ RACE, circularized RNA RT-PCR, and RT-PCR. Both 5′ and 3′ rapid amplification of cDNA ends (RACE) on *Kis2* RNAs were performed using the GeneRacer kit (Invitrogen), according to the manufacturer's protocol. To isolate 3′ ends, the reverse transcription reaction was primed with the oligo(dT) anchor of the kit, and PCR was performed with a combination of the 3′ anchor primer and a *Kis2* primer (oligonucleotide 2, 5′-GTGACATCTCTAGGCCAT CTGTGCGGTATGTG-3′). To isolate 5′ ends, the reverse transcription reaction was primed with a primer internal to *Kis2* (oligonucleotide D, 5′-CTATTC CCGCATCTTCAGCAGCGGCTCCAG-3′), and PCR was performed with the primer specific to the 5′ anchor and a nested *Kis2*-specific primer (oligonucleotide IS, 5′-CAAAACGCGCCAAGCTTAAGCCTCTCCCTCCATTCTC-3′). PCR conditions for both 5′ and 3′ RACE were 94°C for 5 min, and 30 cycles of 94°C for 30 seconds, 63°C for 30 seconds, 72°C for 1 min 30 seconds, and 72°C for 10 min.

Circularized RNA RT-PCR was performed as described by Kuhn (34). Briefly, 10 μg of total RNA from *Kis2* tumors was dephosphorylated, the cap was removed, and the resulting RNA was circularized by ligation. Reverse transcription (Omniscript, QIAGEN) was primed with two different *Kis2* primers, either oligonucleotide D or oligonucleotide IS described above, followed by PCR with nested primers. PCR products were loaded on 1% agarose gel and the bands were eluted, cloned, and sequenced.

RT-PCR for detection of *Phf6* in tumors was performed on 300 ng of total RNA as follows. Reverse transcription was primed with an oligo(dT) and PCR was performed with primers g49rik1 (5′-CTGTCCCACGAACCTAGTTTCCA TTAAG-3′) and g49rik2 (5′-CAGGAAATCAAACAGGTACCCAGGAAAAT G-3′). PCR was performed with one-half of the reverse transcription product reaction: 94°C for 5 min, and 25 cycles of 94°C for 30 seconds, 62°C for 30 seconds, 72°C for 1 min 30 seconds, and 72°C for 10 min.

Nested-RT-PCR to detect *Kis2* expression in normal tissues was performed as follows. Total RNA from normal tissues extracted by TRIzol reagent were subsequently cleaned by phenol-chloroform extraction followed by chloroform extraction to eliminate phenol contamination and ensure reverse transcription reproducibility. reverse transcription was performed with 300 ng of total RNA and primed with the *Kis2* primer (oligonucleotide B, 5′-GACATCGCCAAGG CATACTCCAGGAGTGTAACC-3′). The first PCR was done with oligonucleotide B and oligonucleotide IS4bis (5′-ATCGGGCTTTGGGACGCAGCTCG AG-3′), and the following program was used: 94°C for 5 min, and 15 cycles of 94°C for 30 seconds, 62°C for 30 seconds, 72°C for 2 min, and 72°C for 10 min. The PCR products were then cleaned on Qiaquick PCR purification columns (QIAGEN), and 1/30 of the eluate was used in the nested PCR performed with oligonucleotide IS4bis and a *Kis2* oligonucleotide nested to oligonucleotide B (oligonucleotide B2, 5′-TCTTCCACGAACTCTTCTTTGCCAGATCTGCAG-3′), following the PCR program 94°C for 5 min, and 25 cycles of 94°C for 30 seconds, 64°C for 30 seconds, 72°C for 2 min, and 72°C for 10 min. The PCR products were analyzed on a 2% agarose gel.

Nested RT-PCR to detect *Phf6* transcripts in normal tissues was performed as for *Kis2*. Reverse transcription was primed with oligo(dT). The primers used in the first PCR were oligonucleotide 1-3R1 (5′-TGGAATTCTCGAAACTACCC CTGGCGACAG-3′) and oligonucleotide 5-4Sma1 (5′-TACTGACCCGGGTG TTTCCATTAAGCTGCTG-3′). The two oligonucleotides contained an EcoRI site and a SmaI site, respectively. The PCR program was 94°C for 5 min, and 15 cycles of 94°C for 30 seconds, 56°C for 30 seconds, 72°C for 1 min 30 seconds, and 72°C for 10 min. For the nested PCR, the primers were oligonucleotides 1-3R1 and Phf6TCSma (5′-TACCCCGGGTCCACATTTAGGCCTGGTAT-3′) but the PCR program was for 25 cycles. The PCR products were run on a 2% agarose gel.

The β-actin gene was used as control in all the RT-PCRs performed. The reverse primer was either oligo(dT) or an internal primer (βact2, 5′-GACGGG GTCACCCACACTGTGCCCATCTA-3′). PCR was performed with 1/10 of the reverse transcription product reaction in the presence of primers βact2 and βact1 (5′-CTAGAAGCACTTGCGGTGCACGATGGAGGG-3′). The PCR program was 94°C for 5 min, and 25 cycles of 94°C for 30 seconds, 62°C for 30 seconds, 72°C for 1 min, and 72°C for 10 min.

cDNA cloning of *Phf6*. A cDNA library was prepared from the *Kis2*-rearranged tumor T3 with the Superscript Plasmid System with Gateway Technology for cDNA synthesis and cloning kit (Invitrogen). The library was made using 4 μg of polyadenylated RNA, and cDNA ligation product was precipitated prior to electroporation in electrocompetent *Escherichia coli* DHB10 (Invitrogen); 1 μl of the ligation was used to transform 25 μl of bacteria. A total of 5×10^5 clones were plated on L-broth agar plates (22,5 by 22,5 cm; 8×10^4 clones per plate) containing 100 μg/ml ampicillin, grown overnight at 30°C, transferred on nylon membranes in duplicate and screened with a *Phf6*-specific probe (RT-PCR product eluted from agarose gel). Positives clones were identified on the original plates, isolated and sequenced.

PHF6 localization in NIH 3T3 cells. NIH 3T3 cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% calf serum (Invitrogen). The DNA sequence corresponding to the protein encoded by transcripts t2 and t7 from *Phf6* was PCR amplified using primers 1-3R1 (5′-TGGAATTCTC GAAACTACCCCTGGCGACAG-3′) and 5-4Sma1 (5′-TACTGACCCGGGT GTTTCCATTAAGCTGCTG-3′) for *Phf6*-t7, and primers 5-4R1 (5′-TGGAA TTCTCCTTAAAGTGGCAGTCTCTAAAC 3′) and 5-4Sma1 for *Phf6*-t2. PCR

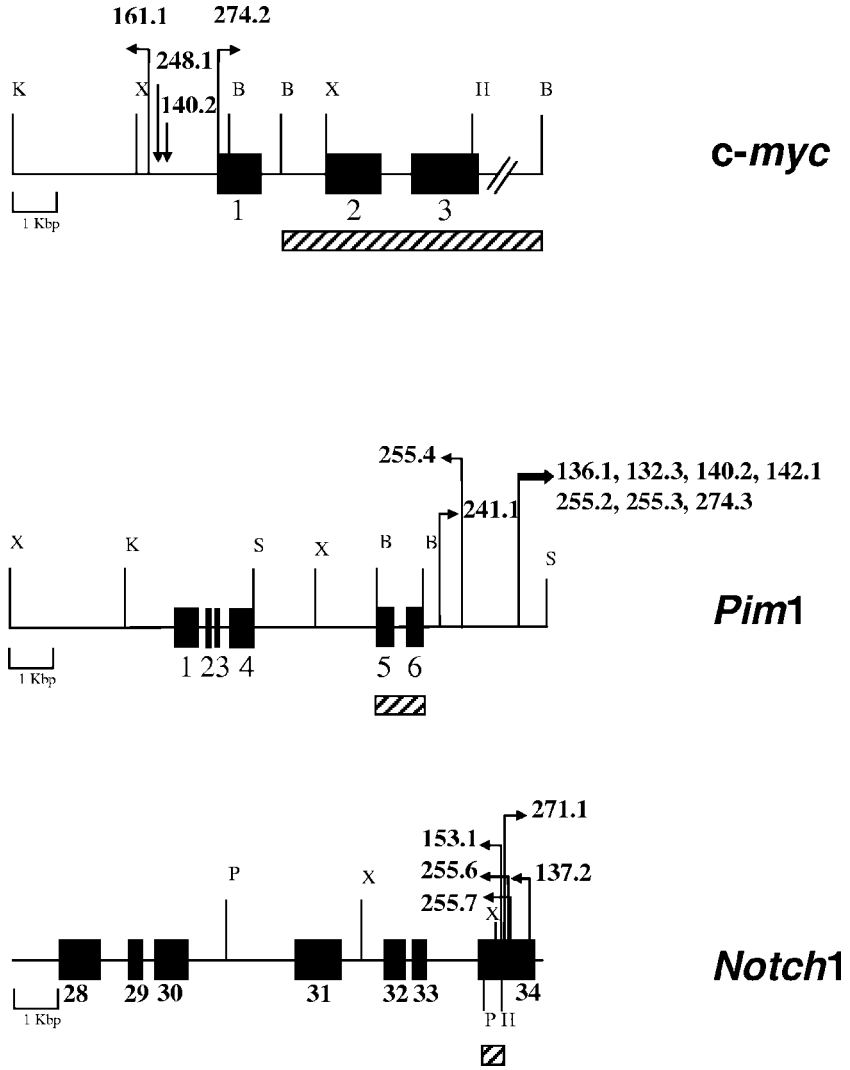

FIG. 1. Positions of RadLV integrations in c-*myc*, *Pim1*, and *Notch1* loci. The probes used are represented by dashed boxes. Arrows above the maps indicate the positions and the orientations of the RadLV integrations. Tumor 140.2 was rearranged for both c-*myc* and *Pim1*. Dark boxes represent exons. B, BamHI; H, HindIII; K, KpnI; P, PstI; S, SacII, X, XbaI.

products were eluted from agarose gel, and cloned in pEGFP-N1 and pEGFP-C2 (BD Pharmingen, Mississauga, Ontario, Canada). NIH 3T3 cells were seeded at 2×10^5 cells per plates and transfected with the different constructs mixed with Polyfect reagent (QIAGEN) according to the manufacturer's protocol. Fluorescent cells were analyzed 24 h post transfection by confocal microscopy.

Sequence analysis. DNA sequences were performed at Canadian Molecular Research Services Inc. (www.cmrsinc.com; Montreal, Quebec, Canada). Sequence analysis was done with the BLAST program of both databases Ensembl and NCBI. Sequence alignment were performed by CLUSTAL.

RESULTS

Integration pattern in RadLV-induced tumors. In an attempt to find new RadLV common integration sites, we used the CpG island technique described in Li et al. (39) to rapidly clone and sequence cellular DNA flanking the proviruses. We examined 72 tumors generated by inoculating newborn C57/Bl and SIM.R mice intraperitoneally with the RadLV/VL3 molecular clone. We found RadLV integrations in the *Pim1* and c-*myc* genes in 30% and 15% of the tumors analyzed, respectively (Fig. 1). Similar results have already been reported by Tremblay et al. (70). In our tumors, integrations in *Pim1* were all clustered in the 3′ region of the gene, while in the case of c-*myc*, integrations occurred upstream of exon 1, in the promoter region of the gene. This integration pattern is similar to that observed in T-lymphoma-inducing retroviruses (9, 10, 65; reviewed in reference 12).

We also found RadLV integrations in the *Notch1* gene in 16% of the tumors analyzed. This gene encodes a transmembrane receptor that is important in hematopoiesis for stem cell commitment to develop into functional T cells and particularly for the assembly of pre-T-cell receptor complexes in immature thymocytes (20, 47). Activating mutations in the heterodimerization domain and/or in the terminal PEST domain of *NOTCH1* seem to be implicated in 50% of human acute T-lymphoblastic leukemias (74). However, retroviral targeting of *Notch1* has not often been reported. Common retroviral integrations in *Notch1* have been reported in two cases, one being

in T-cell lymphomas induced in BALB/c mice by a Moloney murine leukemia virus mutant carrying specific deletions in the U3 long terminal repeat region (76), and the other in tumors from transgenic mice infected with Moloney murine leukemia virus (15, 24). Other single hits in *Notch1* have since then been reported by high-throughput analysis (for a summary, see the Mouse Retroviral Tagged Cancer Gene Database).

For RadLV/VL3, integrations were clustered in the last exon of the gene. This corresponds to the type II integrations described by Hoemann et al. (24) although, contrary to their study, we found proviruses in both orientations. RadLV/VL3 integrations occur, in terms of protein, before the OPA motif (tumors 153.1 and 271.1), within the OPA motif (tumor 255.6), after the PEST motif (tumor 255.7), or 400 nucleotides downstream of the stop codon in the 3′ untranslated region (tumor 137.2). *Notch1* activation in leukemia has been reported to be often associated with c-*myc* (15, 24) but none of our *Notch1*-rearranged tumors was also rearranged for c-*myc* (results not shown).

***Kis2*, a new common integration site.** One of the RadLV/VL3 integration site corresponded in the database to a single SL3-3 integration site (67), suggesting that this region could be a common integration site. Southern blot and PCR analyses revealed retroviral integration in this specific locus in 11% of the tumors analyzed (8 tumors out of 72). This new common integration site is situated on the X chromosome, more precisely in a region corresponding to bp 45091000 (XA4 region) according to the mouse genome sequence database (www.ensembl.org). We named this site *Kis2* (Kaplan Integration Site 2). Five of the tumors are almost clonal since the rearranged allele appears slightly weaker than the normal one (Fig. 2A and not shown). Note that the normal allele is absent in tumor 2 (T2) because this mouse was a male. DNA rearrangement in the three other tumors (T4, T7, and T8) was not detectable in Southern blots, but retroviral integration was easily detectable by PCR.

Southern blot analysis of DNA from RadLV/VL3-induced tumors with a U3 long terminal repeat probe or a T-cell receptor (TcR) probe has already been performed and these studies show that RadLV tumors are clonal cell populations (2, 35, 70). However, these studies did not allow to determine if the total cell population was homogenous. Our PCR-based strategy certainly allows to clone not only retroviral integrations that are present in the total tumor cell population but also in more minor cell population.

Figure 2B shows the gene and expressed sequence tag (EST) map of the *Kis2* region. We limited the region on one side by the glypican 3 gene (*Gpc3*), and by the *Phf6* gene on the other side. *Gpc3* encodes a heparan sulfate proteoglycan which is involved in developmental morphogenesis (13), while *Phf6* encodes a protein implicated in a mental disorder called Borjeman-Forsmann-Lehman syndrome (BFLS) (43). A third gene, *Rs17*, is present and encodes a ribosomal protein. Three ESTs (AI464896 and unigene clusters Mm. 277876 and Mm. 23411) have also been reported in this region (Fig. 2B). All RadLV/VL3 integrations were found clustered in a 4-kbp region between the *Rs17* gene and the Mm.277876 EST cluster. SL3-3 and Moloney murine leukemia virus integration sites are also present in this region.

Most of the time, retroviruses were integrated in an antisense orientation with respect to the transcription direction of the Mm.277876 EST cluster (Fig. 2B, tumors T1, T2, T3, T4, T7, and T8), as well as SL3-3. However, in the case of tumors T5 and T6, the retrovirus is integrated in the sense of transcription and closer to Mm.277876 (0.5 kbp). Interestingly, this region was also reported as a common integration site (called *XPCL*1) for Moloney nurine leukemia virus in $p27^{-/-}$ transgenic mice (27). Moloney integrations are localized, although not oriented, to the same area as RadLV integrations (Fig. 2B), and overexpression of EST AI464896 was reported in the Moloney-rearranged tumors. This EST is located 4 kbp downstream of the RadLV and Moloney integrations, and 0.5 kbp downstream of the Mm.277876 cluster (Fig. 2B). However, EST AI464896 expression was not altered in our RadLV-induced tumors (not shown) suggesting that another EST or gene was targeted in our tumors.

Thus, we decided to analyze the expression of all the genes and ESTs present in the entire 450-kbp region. We did not find any change of expression for the *GPC3* and *Rs17* genes or for Mm.23411 EST. Interestingly, we found that expression of EST from the Mm.277876 cluster was drastically increased in the *Kis2* rearranged tumors. Therefore, we believe that the RadLV/VL3 retrovirus, although integrated in the same region as the Moloney murine leukemia virus, is targeting another gene. Also, the expression of the *Phf6* gene was affected (see below).

A novel gene is overexpressed in *Kis2*-rearranged tumors. Each gene and EST present in the region of 450 kbp was tested by Northern or RT-PCR analysis to detect altered expression in the *Kis2*-rearranged tumors. Mm.277876 expression was drastically altered in all *Kis2*-rearranged tumors. We named the gene corresponding to the Mm.277876 EST cluster *Kis2*.

Two clearly different patterns of expression were obtained depending upon the orientation of the provirus. First, five transcripts of different sizes were detected in tumors T1, T2, and T3 (Fig. 3A) and second, the third RNA species was largely overexpressed in tumors T5 and T6, where the retrovirus is integrated in the sense orientation. Overexpression of *Kis2* transcripts was barely detectable in tumors T4, T7, and T8 due to the nonclonality of these tumors. In the nonrearranged tumors, the level of expression of *Kis2* transcripts remained very low. Interestingly, the BL/VL3 cell line which is a lymphoid cell line established from a tumor induced by the passaged RadLV, showed low but significant quantities of the three longest transcripts. Nevertheless, no integration could be found in the immediate 6-kbp region of the *Kis2* locus in this cell line.

Using RACE and circularized RNA RT-PCR techniques, we could isolate and further characterize the complete cDNAs corresponding to the five *Kis2* RNA species (*Kis2*-t1 to -t5) overexpressed in the tumors (Fig. 3A). Complete nucleotide sequences of the five cDNAs are accessible in GenBank under accession numbers AY940614 to AY940618.

The five RNA species are mainly nonpolyadenylated, as they are detectable only on Northern blot performed with total RNA. However it is possible to amplify them when performing RT-PCR primed with an oligo(dT), suggesting that at least a little part of these RNAs is polyadenylated. Furthermore, the ESTs from cluster Mm.277876 were obtained from an oligodT-

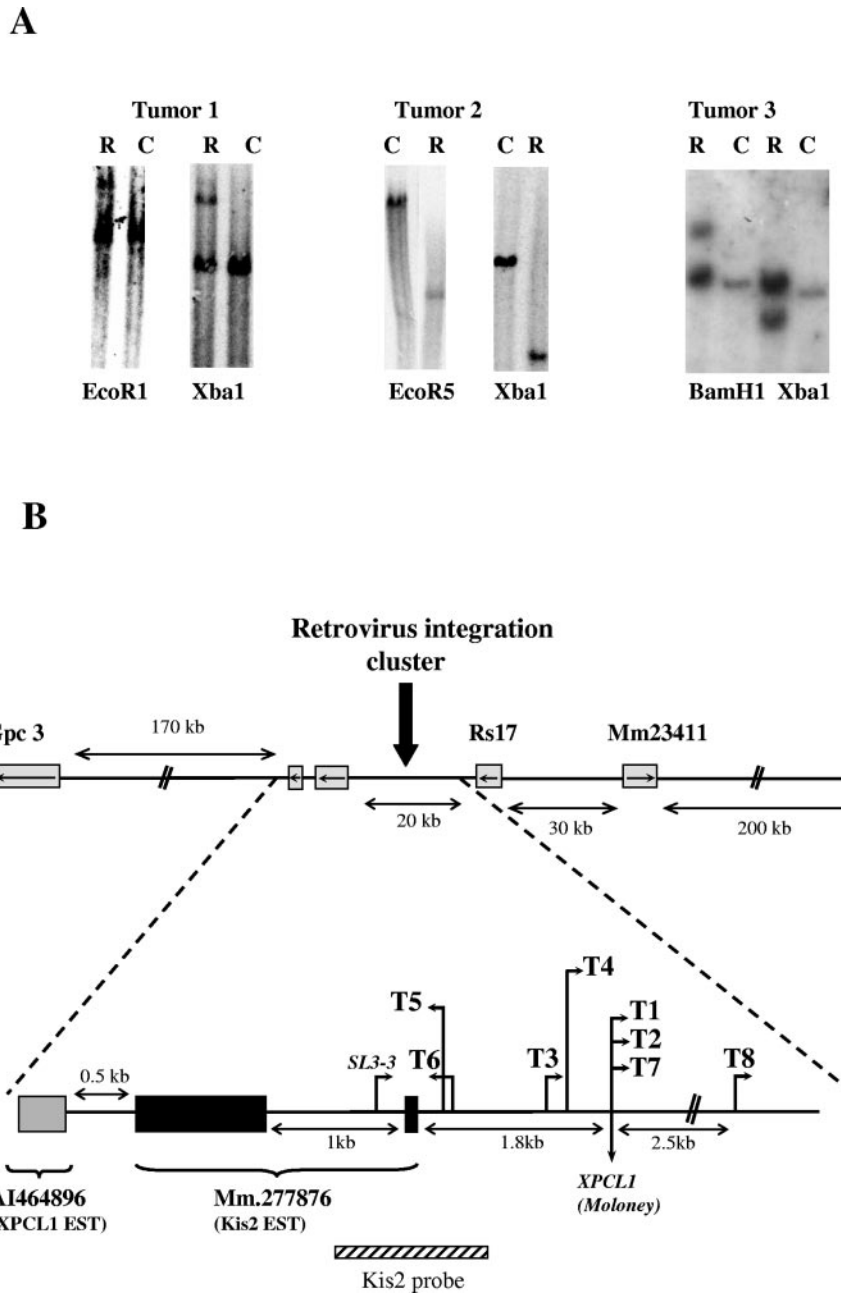

FIG. 2. Southern blot analysis of some RadLV-induced leukemias. (A) Control tissue (C) and tumor DNAs (R) were digested with restriction enzymes as indicated and hybridized with the *Kis2* probe. Note the absence of the normal allele in rearranged tumor 2 from a male mouse. (B) Genomic map of the *Kis2* locus. The RadLV integrations and orientations are represented by arrows (T1 to T7). SL3 and Moloney murine leukemia virus (*XPCL1*) integrations are also represented. Genes and EST clusters of the region are shown by gray boxes. The arrow in the box indicates the orientation of transcription.

primed library and they contained a 3′ end identical to that of those RNAs identified in our 3′ RACE study. The canonical AAUAAA polyadenylation signal could not be found in the 3′ end of the transcripts. But some genes are known to have alternative signal sequence and we found, at the 3′ end, the signal AAAGAA which has been involved in polyadenylation of some non-AAUAAA-containing mRNAs (72). This signal could represent a putative polyA signal of the *Kis2* gene.

The longest *Kis2* RNA species (t1), which initiates at TSS2 (Transcription Start Site 2), is 2 kb in length and gives rise to species t2, t4, and t5 by alternative splicing (Fig. 3A). Splicing sites are canonical and respect the GT/AG rules. The third RNA species (t3) is initiated from a different transcription start site (TSS3), located 0.6 kbp downstream of TSS2. These two TSSs are themselves different from that of EST BB658090 of the Mm.277876 cluster registered in the database, located 0.4 kbp upstream the TSS2 (TSS1) (Fig. 3A). Thus, The *Kis2* gene contains at least three different start sites, each one being

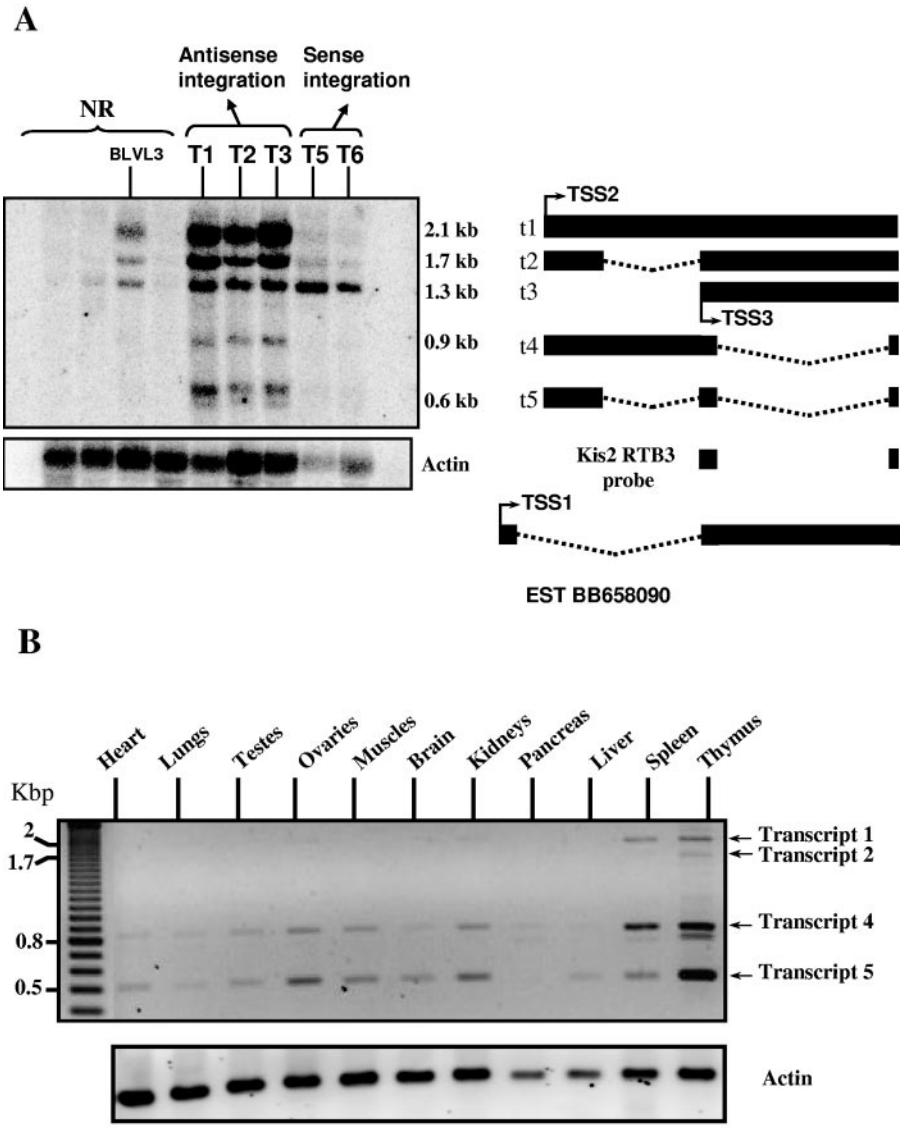

FIG. 3. (A) Northern blot analysis of the *Kis2* RNA species and schematic representation of their structure. Left side: total RNA of rearranged and nonrearranged (NR) tumor was hybridized with the *Kis2* RTB3 DNA probe (indicated in the right panel). T1, T2, and T3 correspond to tumors with an antisense retroviral integration, and T5 and T6 to tumors with a sense retroviral integration. BLVL3 shows RNA extracted from the BLVL3 tumor cell line. Right side: schematic representation of the five RNA species (t1 to t5) as deduced from sequencing. The splicing events are indicated. TSS: transcription start site. The *Kis2* EST BB658090 is also shown. (B) Semiquantitative analysis by nested RT-PCR of the expression level of *Kis2* transcripts in normal mouse tissues. The actin gene was used as an internal standard.

localized in a CpG island and sufficiently distant from the others suggesting three different promoters.

The coding potential of the five RNA species has been analyzed by computer and we could not find any open reading frame in their sequences. Thus, we concluded that these RNAs belong to the growing number of noncoding RNA group.

The *Kis2* EST unigene cluster (Mm.277876) contains only three ESTs: BB658090 and BB470007, which correspond to the 5′ and 3′ ends, respectively, of the same cDNA clone isolated from a 12-day embryo eyeball cDNA library, and BB749744, identical to BB470007 and isolated from a pool of tissues. Thus, this gene is very poorly represented in the EST database. In Northern blots and by RT-PCR, the signal corresponding to BB658090 remained undetectable in the *Kis2*-rearranged tumors as in normal tissues. However, we were able to amplify the transcript by nested RT-PCR (data not shown), confirming the rarity of this transcript.

In normal mouse tissues, *Kis2* RNAs were weakly detectable by Northern blot in the thymus only (data not shown), while detection for others tissues was only possible by nested RT-PCR (Fig. 3B), suggesting again a very weak transcription of those RNAs in a normal state. Transcripts 4 and 5 are detectable in almost all the tissues tested but with different intensity, the strongest signal being in thymus. Transcripts 1 and 2 are clearly seen in the thymus and spleen only. In general, in our RT-PCR analysis, we found that the intensity of transcripts 4 and 5 was always higher than that of transcripts 1 and 2. This observation was made in the normal tissue analysis (Fig. 3B)

but also in the *Kis2*-rearranged tumors (result not shown). However, when the same tumors are analyzed by Northern blot for *Kis2* expression, we observed a stronger signal for the 2 longest transcripts when compared to the 2 smallest (Fig. 3A). Thus, it appears that the differences observed in signal intensity between the *Kis2* long and small transcripts in Fig. 3B is probably the result of a skewed PCR in favor of the smaller transcripts since the same pair of oligonucleotides was used for transcript 1, 2, 4, and 5 amplification. Specific detection of transcript 3 in normal tissues was not possible by RT-PCR because its sequence is entirely included in that of transcripts 1 and 2 (Fig. 3A).

Until now, no orthologues of the *Kis2* gene have been found, as it is absent from EST database of other species. Interestingly, a comparison of genomic DNA between mouse, rat, and human revealed some highly conserved regions. In particular, striking sequence conservations are observed in the six regions corresponding to TSSs, splice sites, and a 250-bp sequence located in the last quarter of the *Kis2* transcription unit (Fig. 4, region 5). This suggests that the *Kis2* gene might be present in other species such as rat and human, even though ESTs have not been reported yet.

***Phf6* is also overexpressed in *Kis2*-rearranged tumors.** The *Phf6* gene is located about 250 kbp away from the RadLV integrations (Fig. 2B). Nevertheless, we verified the expression of all the genes and ESTs around the integration site. To our surprise, we observed an important overexpression of this gene specifically in *Kis2*-rearranged tumors (Fig. 5A). Tumors with strong overexpression of *Phf6* (T1, T2, T3, T5, and T6) were also those which strongly expressed *Kis2* RNAs, and they corresponded to the most clonal *Kis2*-rearranged tumors. Tumor T4 was partially rearranged for *Kis2* since the rearrangement was detected by PCR only. Interestingly, it expressed intermediate levels of *Phf6* (Fig. 5A). Two tumors showed barely detectable levels of *Phf6* although we were unable to detect a provirus in the 6-kbp region that encompass the *Kis2* integration locus. However, we cannot exclude the possibility that a retroviral integration has occurred outside this region.

The mouse *Phf6* gene has 11 exons and encodes a 365-amino-acid protein, which contains two Plant Homology Domain (PHD) fingers. *Phf6* was unknown until it was recently implicated in the Börjeson-Forssman-Lehmann X-linked mental retardation syndrome (BFLS; OMIM 301900) in humans (43). To isolate complete cDNAs of the *Phf6* gene, we constructed a cDNA library from a rearranged tumor and we obtained clones corresponding to three different transcripts (labeled t2, t3, and t7 in Fig. 5B and 6).

The longest (*Phf6*-t3) has a 4-kb 3′ untranslated region and is the most abundant in our tumors. This is also the major transcript found in the database as it constitutes 96% of ESTs in *Phf6* cluster Mm.26870 (400 total ESTs). The ESTs corresponding to this transcript are described in a broad variety of embryonic and adult tissues, with reports of its presence in brain and activated spleen being the most frequent.

The second transcript in size (*Phf6*-t2) also has 11 exons, but lacks the long 3′ untranslated region (Fig. 6). It is weakly expressed in our tumors, as only a very weak signal is observed on poly(A) RNA Northern blots (not shown). Furthermore, it represents only 2% of the ESTs present in the Mm.26870 cluster (see also the legend to Fig. 6). Both *Phf6*-t2 and *Phf6*-t3 encode an identical 365 amino acid protein. We could detect expression of t2 and t3 by nested RT-PCR (Fig. 5B) in almost all the normal tissues, with the weakest expression being in muscle. The primers used in this assay did not allow us to distinguish *Phf6* transcript 2 from transcript 3. However, we obtained identical results with primers specific to transcript 3 only (reverse primer located at the 3′end of the 3′ untranslated region; data not shown). Thus, the signal observed in Fig. 5B corresponds almost entirely to transcript 3.

Interestingly, we also found a third smaller transcript (*Phf6*-t7) in our cDNA library, lacking exons 2 to 5 inclusively (Fig. 6). This species has never been described before, and it gives rise to a protein lacking the first PHD domain. Its detection required nested RT-PCR and it seemed to be expressed mainly in normal thymus (Fig. 5B).

Analysis of all the ESTs from the *Phf6* cluster revealed four other alternative transcripts in the mouse, although they are rarely represented in the database (*Phf6*-t1, *Phf6*-t4, *Phf6*-t5, and *Phf6*-t6, Fig. 6). *Phf6*-t1 is composed of exon 1 and a part of intron 1. It contains an open reading frame totally different from that of PHF6, since the normal ATG start codon is in exon 2. *Phf6*-t4 lacks exons 2 and 3 and codes for a protein that still contains the two plant homology domains but has lost the first 80 amino acids at the N-terminal end. This transcript was detected in thymus, lungs, ovaries, kidneys, liver, and spleen (Fig. 5B). *Phf6*-t5 has lost exons 6 to 10 and gives rise to a protein lacking the second PHD domain. *Phf6*-t6 lacks exons 9 and 10, resulting in a protein 41 amino acids shorter at the C-terminal end that has conserved the two PHD domains.

In humans, the PHF6 protein localizes to the nucleolus, and each of the four nuclear localization signals (NLSs) identified seems to be important (43) (Fig. 6). Using enhanced green fluorescent protein (EGFP) fusion protein constructs, we found the same localization in mouse NIH 3T3 cells. Protein from *Phf6*-t2 did localize to the nucleolus like the human homologue (Fig. 7). Interestingly, protein from transcript 7 of *Phf6* (labeled PHF6 t7 on Fig. 7), which lost all four NLSs, still localized to the nucleus but no longer to the nucleolus (Fig. 7, middle panel). These results were the same whether the EGFP was in the N- or C-terminal position in the fusion protein constructs. As NLSs are absent from the PHF6 t7 protein, this suggests that other unidentified NLSs are present in the protein or that the protein interacts with other nuclear proteins to be imported.

DISCUSSION

In this article, we report the identification of *Kis2*, a novel RadLV common integration site on the X chromosome. Before the mouse genome sequencing was completed, searching for genes located in a region was painstaking and time-consuming, and as a consequence, common integration sites were often associated with a unique gene. We analyzed the expression of all the genes present at the *Kis2* locus and, surprisingly, we found that two genes were overexpressed in the rearranged tumors: first, *Kis2*, which is located close to the retrovirus integration site, and second, the *Phf6* gene, situated about 250 kbp away. Thus, this article is also the first report on a retroviral integration that affects the expression of two different distant genes. Indeed, it is now possible to investigate the effect

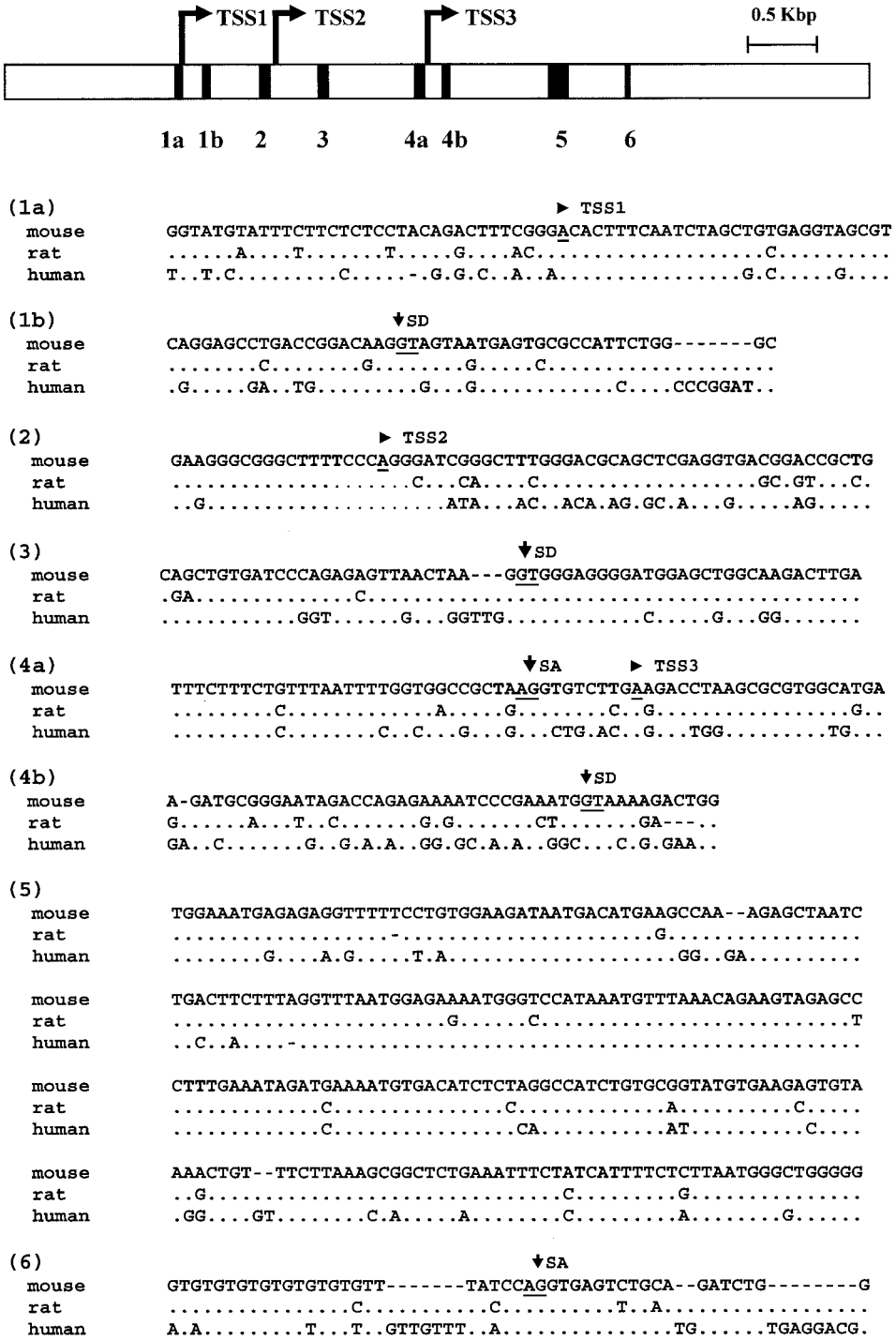

FIG. 4. Analysis of *Kis2* sequence conservation among species. Six regions of the *Kis2* locus were found to be highly conserved among rat, human, and mouse as indicated on the map at the top and were identified as 1 to 6. The three transcription start sites (TSSs) (see Fig. 3), the splice donor site (SD), and the splice acceptor site (SA) are indicated above the sequence and the nucleotides are underlined.

of an integrated retrovirus on all the genes from a given region since almost the entire mouse genome is now available and well annotated.

First of all, our results show that genes situated quite far from the integration site could be activated as well. This has been previously described in the case of Moloney murine leukemia virus when integrated at the *Mlvi*4 locus, 300 kbp from the c-*myc* locus (37). This fact was rarely described in the literature until recently, mainly due to the lack of the appropriate genetic information. Recent high-throughput studies using the mouse genome sequence database have reported common retroviral integrations in regions located much further

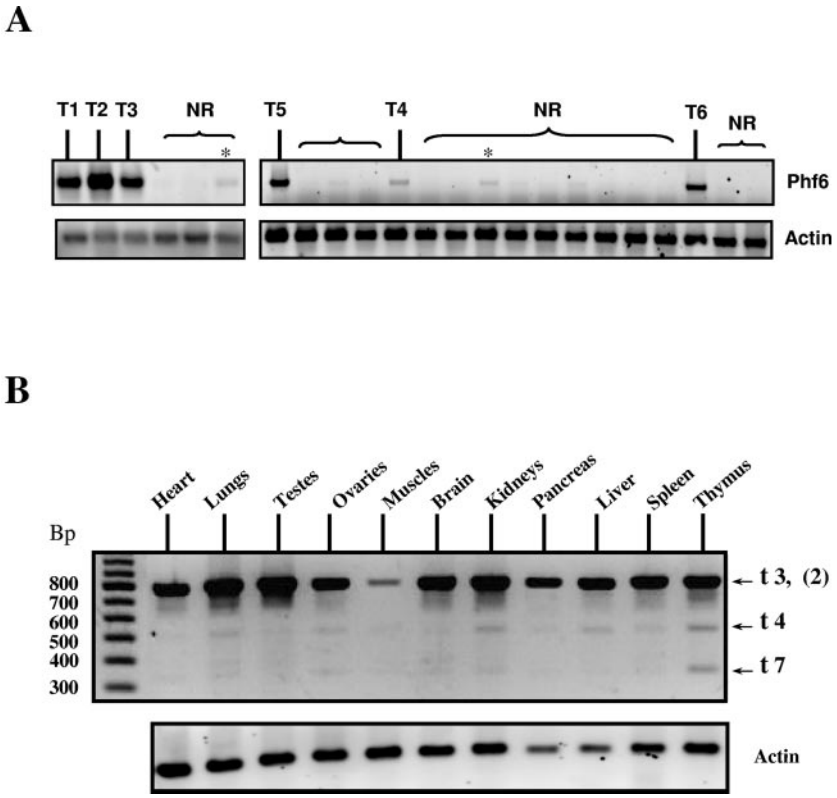

FIG. 5. Analysis of *Phf6* gene expression in tumors and normal mouse tissues. (A) *Kis2*-rearranged (T1 to T6) and nonrearranged (NR) tumors by RT-PCR. (B) Nested RT-PCR on a panel of normal tissues. Primers used allowed amplification of transcripts 2 and 3, as well as transcripts 4 and 7, according to the expected sizes (800 bp for transcripts 2 and 3, 550 bp for transcript 4, and 340 bp for transcript 7). Two tumors (*) were not rearranged for *Kis2* but expressed low levels of *Phf6*. The actin gene was used as an internal standard.

away from the target gene (11, 32, 68). It is clear that identification of target genes located further away from the common retroviral integration sites will now be possible.

How can retroviral integrations in regions located far away from a target gene act so efficiently on its expression? Specific chromatin structures could be involved so that distant regions are brought closer. Alternatively, key regulatory elements of gene expression could be present at the integration site and be affected by the presence of the retrovirus. So retroviruses could be tools not only to find oncogenes, but also to find their regulatory elements, as already mentioned by Kim et al. (32).

It is also interesting that not all the genes in the region are affected by the integration. This had already been reported in the case of SL3-3 when integrated in the *Ras* locus: only the expression of *Ras* is affected, whereas the expression of *unr* is not, despite its localization close to the integration site (48). This could result from the fact that some genes are located in an active area of the chromatin while others are not. It can also be illustrated by Moloney murine leukemia virus integrations at the *XPCL1* locus (Fig. 2B) which affect the expression of EST AI464896 (27), whereas RadLV did not. A response of a given promoter to a specific enhancer could also be involved so that different enhancers would affect gene expression differently. Indeed, long terminal repeat sequences have a great influence on the insertional activation, as recently shown by Johnson et al. (29). In light of this, it would be interesting to know if Moloney murine leukemia virus or SL3-3 enhancers do affect the expression of *Kis2* and *Phf6* when integrated at the *Kis2* locus. Also relevant to this question, it is interesting that antisense retroviral integrations affect transcription from TSS2 and TSS3 in a similar fashion, whereas integrations in the sense orientation affect expression from TSS3 to a much greater extent (Fig. 3A).

Kis2 is a novel gene that has never been described before. It is poorly represented in the mouse database. Indeed, we found that *Kis2* is weakly expressed in all the normal tissues, the thymus being the highest expressor (Fig. 3B). To date, no ESTs specific to *Kis2* have been described in other species. However, the *Kis2* transcription unit, or at least some of its features, seems to be conserved in rat and human, as alignment of the genomic regions reveals highly conserved areas in the sequence overlapping TSSs and splice sites.

The main characteristic of *Kis2* transcripts is that they contain only minute open reading frames, thus classifying them as noncoding RNAs. A few years ago, noncoding RNAs were considered noise in the transcriptional system. But recent studies from the human and mouse genome sequencing projects revealed that a real noncoding RNA dimension remains to be explored. The function of most of those noncoding RNAs is still unknown. However, a few of them have now been functionally characterized. Noncoding RNAs are implicated in mechanisms such as imprinting (H19) (57), dosage compensation transcription regulation (Xist [6], Tsix [38], and roX [49]), (SRA [36] and 7SK [33]), regulation of transcript translation by

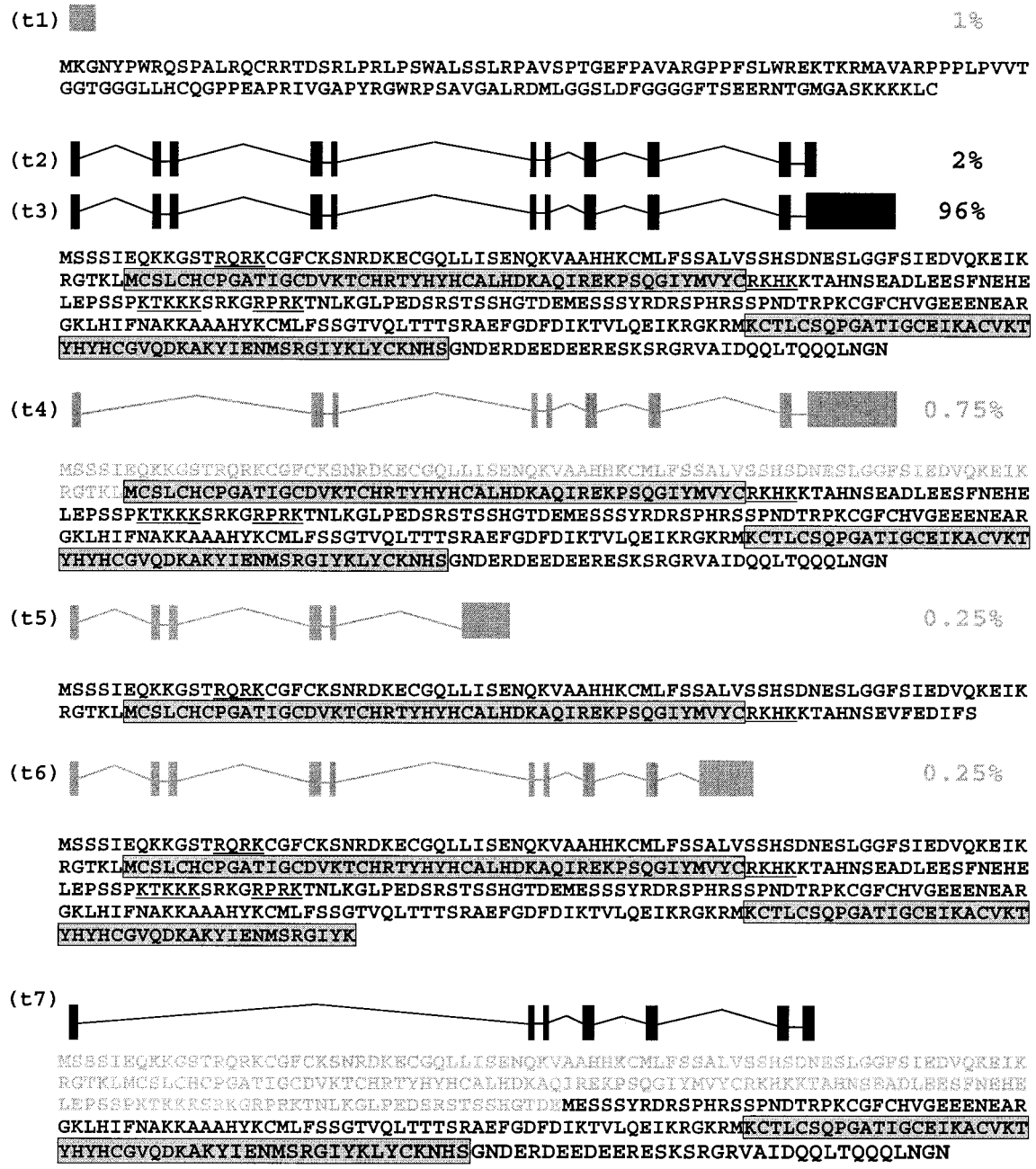

FIG. 6. Schematic representation of all *Phf6* transcripts. All the ESTs (400 total) of *Phf6* Unigen cluster Mm.26870 were analyzed using the NCBI and Ensembl BLAST programs. The relative abundance of each transcript is indicated at the right of each transcript in % except for transcript 7, which has not been reported yet. The tissue origin for each transcript was as follows: transcript 3 (t3) is the most abundant mRNA species of the cluster (96%) is expressed in a broad variety of tissues; transcript 2 (t2): embryonic stem cells (BB400028), adult thymus (BB033610), 13-day embryo lung (BB482014), 12-day embryo (BB092516), 8-cell embryo (BB728515), testes (AV272544), 7-day embryo (BB408501), and dendritic cell line (BY764534); transcript 1 (t1): dendritic cells (BY206667), bronchial arches (BX527136), 11-day embryo (AV123417), and adult thymus (BB033906); transcript 4 (t4): newborn skin (BB612388), newborn whole eye (CO429499), and 16-day embryo kidney (BY060450); transcript 5 (t5): B cells (BG145595); and transcript 6 (t6): 16-day newborn thymus (BB197834). Transcripts represented in black (t2, t3, and t7) correspond to those found in the cDNA tumor library. The amino acid sequence is indicated below each transcript. The PHD fingers are gray boxed, and the nuclear localization signals are underlined. The amino acid sequence in gray is not present in the protein and is indicated for comparison.

antisense effect involving a particular micro-RNA (lin-4 [54] and let-7 [62]), and finally oxidative stress (adapt33 [73] and gadd7 [25]).

The function of *Kis2* remains to be determined. One hypothesis about the role of noncoding transcripts argues against a function of the RNA itself but rather stipulates that only the transcription across the locus would be important, as this is the case for the β-globin locus control region (16). However, if this

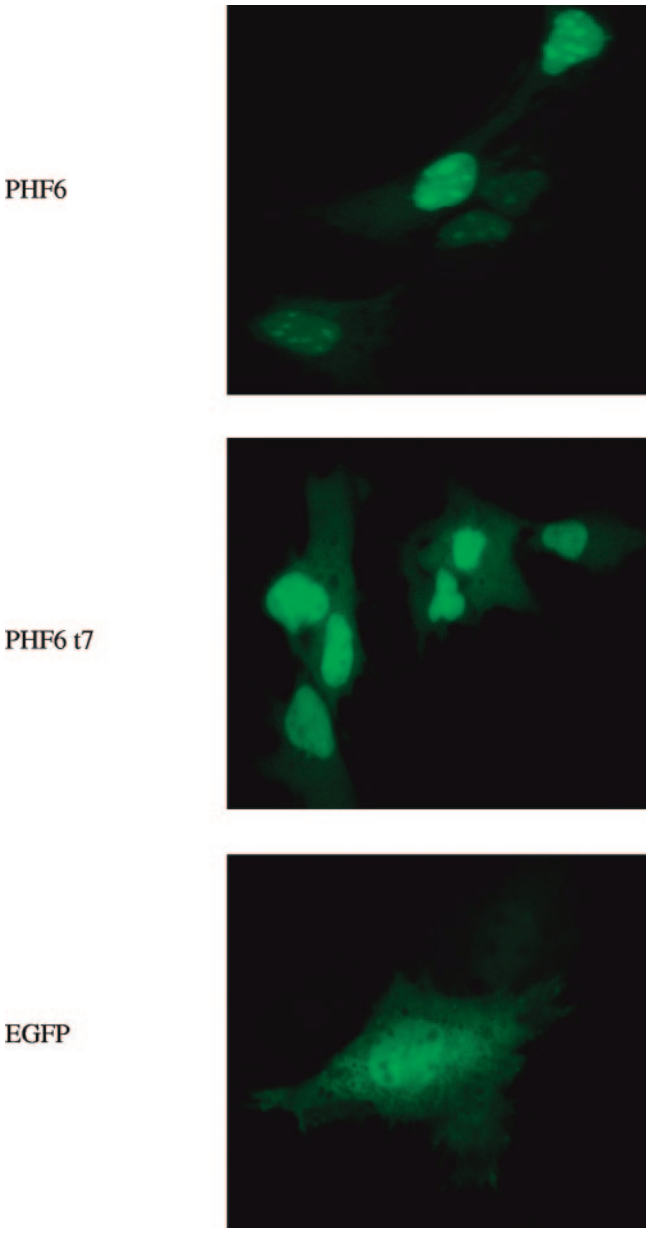

FIG. 7. Subcellular localization of PHF6 protein from transcripts 2 and 7. The normal PHF6 protein and the truncated form encoded by transcript 7 (PHF6 t7; see Fig. 6) were fused to the enhanced GFP at either the C- or N-terminal end using the pEGFP-C2 or pEGFP-N1 vector, respectively. The resulting constructs were transfected into NIH 3T3 cells, and the cells were examined by confocal microscopy. The normal PHF6 protein and the truncated form (PHF6 t7), both in pEGFP-N1, are presented in the top and central panels, respectively. As a control, the pEGFP plasmid is shown in the bottom panel.

is the case for the *Kis2* locus, it is difficult to understand the meaning of the three promoters and the six alternative transcripts. Such characteristics have been reported for others noncoding RNAs such as *Xist* and *Tsix*, for which no functions for the spliced species have been attributed yet (52, 53, 66). In the case of noncoding RNAs, the sequence could really be important either at the level of antisense functions or as a structure. Because of their sequence differences, the *Kis2* RNA species very likely have distinct properties, which argues in favor of a function for the *Kis2* RNA per se.

Another interesting particularity of *Kis2* RNAs is that they apparently lack a poly(A) tail, as we could not detect them on Northern blot performed with poly(A) RNA. The poly(A) tail is of great importance for mRNA, both at the level of transcription regulation and for stability. Sequence elements promoting mRNA instability have been described, such as those named AU-rich elements, which are located in the 3′ untranslated region of mRNA and control their deadenylation (75). Thus, some RNAs could be present as polyadenylated or deadenylated forms as a result of poly(A) tail length regulation, and are qualified as bimorphic RNAs. Alternatively, some mRNAs have sequences in their 3′untranslated region named poly(A)-limiting elements, which limit the length of the poly(A) tail (17, 19). Thus, those RNAs are generated with a poly(A) tail containing less than 20 A's, which leads eventually to transcript instability in response to some regulatory factors. The same situation could occur for the *Kis2* mRNAs: They could have a poly(A) tail too short to be successfully purified by oligo(dT) affinity chromatography, but still long enough to be amplified by oligo(dT)-primed RT-PCR. Such poly(A)-limiting element consensus sequences are not present in the *Kis2* 3′ region, although other sequences could be responsible for limiting the poly(A) length, as this is also the case for immunoglobulin M mRNA (58). Finally, the lack of a poly(A) tail of normal length in *Kis2* RNAs could be in part responsible for their lower representation in the databases since most cDNA libraries are primed by oligo(dT).

The *Phf6* gene was discovered recently in human and was implicated in the Börjeson-Forssman-Lehmann syndrome (43). BFLS is characterized by moderate to severe mental retardation, epilepsy, hypogonadism, hypometabolism, obesity with marked gynecomastia, swelling of subcutaneous tissue of the face, narrow palpebral fissures, and large but not deformed ears (5). The *Phf6* gene encodes a 365-amino-acid protein, which has two plant homology domain fingers. Several PHF6 mutations associated with BFLS have been found (43): some are missense mutations, with one of them reported to affect a cysteine in the first plant homology domain, and others are truncation mutations. One of those truncation mutations occurs in the beginning of the protein, resulting in its complete loss, while another one, which appears to be recurrent in BFLS (44), occurs at the end of the protein (R342X).

The multiple symptoms which characterize this syndrome can be correlated with the expression of the gene in a broad variety of tissues, and with the existence of several alternatively spliced transcripts. Although *Phf6* expression is detected in most human and mouse tissues, expression is low and only detected by nested RT-PCR. However, in the *Kis2*-rearranged tumors, its expression was easily detectable by Northern blot and RT-PCR analysis (not shown and Fig. 5). In consequence, this article is the first report suggesting an involvement of *Phf6* in the development of leukemia.

PHD finger composition is close to the RING finger and LIM domain (1), which both function in association with two Zn^{2+} atoms. Although the PHD finger function is still unknown, it is mainly found in proteins involved in transcriptional regulation (1). Two groups can be distinguished: the first one including transcriptional activators, repressors and cofactors

(28), and the second one including proteins of chromatin modulating complexes such as acetyltransferases or acetyltransferase-containing complexes such as p300 or CBP (4). To corroborate this, the PHD finger of the transcription factor SPBP is involved in chromatin-mediated transcriptional regulation as a protein-protein interaction domain (46). Furthermore, naturally occurring single amino acid substitutions in the PHD finger of the ATRX and AIRE1 proteins predispose individuals to α-thalassemia, developmental defects and autoimmune diseases (3, 14). Interestingly, ALL1, the human homologue of *Drosophila* trithorax, and other PHD finger-containing genes are also frequently disrupted by chromosomal rearrangements that occur in acute lymphoid and myeloid leukemias (8, 18, 51, 56, 69).

As in human, the mouse PHF6 protein was found to localize to the nucleolus. This organelle is mainly known as a site dedicated to ribosome biogenesis. Other events take place in the nucleolus, such as for example, the maturation of polymerase III transcripts (7). The nucleolus also seems to behave as a sequestering compartment for regulatory complexes. For example, the $p19^{ARF}$ tumor suppressor, which is implicated in p53 regulation, exclusively localizes to the nucleolus. Cytoplasmic p53 degradation is mediated by the oncoprotein Mdm2 and $p19^{ARF}$ acts by complexing and sequestering mdm2 in the nucleolus, thus preventing its nuclear export and p53 degradation (55). Nucleolus also hosts viral protein, such as HIV-Rev (26). According to these observations, the nucleolus still appears as a "dark box", now including *Phf6*.

However, it is likely that *Phf6* serves fundamental functions in the cell since it is highly conserved among species. Several transcripts are present in the database, but the major one is transcript 3 (Fig. 6). It differs from the others by its long 3′ untranslated region, which probably serves some regulatory or stability goals. Transcripts 4 and 6 give rise to a protein which retains the two PHD fingers, while the protein encoded by transcript 5 has lost the second one. Interestingly, we found another alternative transcript (transcript 7, Fig. 6) which has never been described before. It appears very weakly expressed, mainly in the thymus. This splicing eliminates exons 2 to 5, giving rise to a protein initiated at an ATG codon further downstream and thus lacks the first PHD domain and all the identified nuclear/nucleolar localization signals. Furthermore, the corresponding protein no longer localizes to the nucleolus, but still remains in the nucleus.

Finally, it is tempting to make a link between the *Kis2* and *Phf6* genes which are both expressed simultaneously in *Kis2*-rearranged tumors. Because of the noncoding nature of *Kis2* RNAs, one could imagine a trans-action of these RNAs on the *Phf6* locus, which would lead to its overexpression. Otherwise, because of the relative proximity of *Kis2* and *Phf6* on the chromatin, a *cis*-action of the transcription across the *Kis2* locus on the *Phf6* expression, mainly through a chromatin structure remodeling could be imagined.

There is no doubt that the *Kis2* locus is involved in thymomas occurrence since it is targeted by three different retroviruses that induce T-cell leukemia: Moloney, SL3-3 and RadLV. Since there is some variability in tumor clonality, some being almost clonal and some not at all, it is tempting to speculate that the role of the *Kis2* locus is more oriented towards tumor progression rather than tumor initiation. We must now further characterize *Kis2* RNAs and PHF6 protein and determine if either the *Kis2* or *Phf6* gene or both have oncogenic potential.

ACKNOWLEDGMENTS

We thank Elsy Edouard and Laurent Poliquin for helpful discussions.

This work was supported by grant MOP-37994 from the Canadian Institutes of Health Research.

REFERENCES

1. **Aasland, R., T. J Gibson, and A. F. Stewart.** 1995. The PHD finger: implication for chromatin-mediated transcriptional regulation. Trends Biochem. Sci. **20:**56–59.
2. **Ben-David, Y., E. Yefenof, and M. Kotler.** 1987. Clonal analysis of radiation leukemia virus-induced leukemic and preleukemicmurine cells. Cancer Res. **47:**6590–6594.
3. **Bjorses, P., M. Halonen, J. J. Palvimo, M. Kolmer, J. Aaltonen, P. Ellonen, J. Perheentupa, I. Ulmanen, and L. Peltonen.** 2000. Mutations in the AIRE gene: effects on subcellular location and transactivation function of the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy protein. Am. J. Hum. Genet. **66:**378–392.
4. **Bordoli, L., S. Husser, U. Luthi, M. Netsch, H. Osmani, and R. Eckner.** 1992. Functional analysis of the p300 acetyltransferase domain: the PHD finger of p300 but not of CBP is dispensable for enzymatic activity. Nucleic Acids Res. **29:**4462–4471.
5. **Börjeson, M., H. Forssman, and O. Lehmann.** 1962. An X-linked, recessively inherited syndrome characterized by grave mental deficiency, epilepsy, and endocrine disorder. Acta Med. Scand. **171:**13–21.
6. **Brockdorff, N., A. ashworth, G. F. Kay, V. M. McCabe, D. P. Norris, P. J. Cooper, S. Swift, and S. Rastan.** 1992. The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell **71:**515–526.
7. **Carmo-Foncesca, M., L. Mendes-Soares, and I. Campos.** 2000. To be or not to be in the nucleolus. Nat. Cell. Biol. **2:**107–112.
8. **Chesi, M., E. Nardini, R. S. Lim, K. D. Smith, W. M. Kuehl, and P. L. Bergsagel.** 1998. The t(4;14) translocation in myeloma dysregulates both FGFR3 and a novel gene, MMSET, resulting in IgH/MMSET hybrid transcripts. Blood **92:**3025–3034.
9. **Corcoran, L. M., J. M. Adams, A. R. Dunn, and S. Cory.** 1984. Murine T lymphomas in which the cellular myc oncogene has been activated by retroviral insertion. Cell **37:**113–122.
10. **Cuypers, H. T., G. Selten, W. Quint, M. Zijlstra, E. R. Maandag, W. Boelens, P. van Wezenbeek, C. Melief, and A. Berns.** 1984. Murine leukemia virus-induced T-cell lynphomagenesis: integration of proviruses in a distinct chromosomal region. Cell **37:**141–150.
11. **Denicourt, C., C. A. Kozak, and E. Rassart.** 2003. Gris1, a new common integration site in Graffi murine leukemia virus-induced leukemias: overexpression of a truncated cyclin D2 due to alternative splicing. J. Virol. **77:**37–44.
12. **Dudley, J. P., J. A. Mertz, M. Lozano, and D. R. Broussard.** 2002. What retroviruses teach us about the involvement of c-Myc in leukemias and lymphomas. Leukemia **16:**1086–1098.
13. **Filmus, J., and S. B. Selleck.** 2001. Glypicans: proteoglycans with a surprise. J. Clin. Investig. **108:**497–501.
14. **Gibbons, R. J., S. Bachoo, D. J. Picketts, S. Aftimos, B. Asenbauer, J. Bergoffen, S. A. Berry, N. Dahl, A. Fryer, K. Keppler, K. Kurosawa, M. L. Levin, M. Masuno, G. Neri, M. E. Pierpont, S. F. Slaney, and D. R. Higgs.** 1997. Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain. Nat. Genet. **17:**146–148.
15. **Girard, L., Z. Hanna, N. Beaulieu, C. D. Hoemann, C. Simard, C. A. Kozak, and P. Jolicoeur.** 1996. Frequent provirus insertional mutagenesis of *Notch1* in thymomas of MMTVD/myc transgenic mice suggests a collaboration of c-myc and *Notch1* for oncogenesis. Genes Dev. **10:**1930–1944.
16. **Gribnaud, J., K. Diderich, S. Pruzina, R. Calzolari, and P. Fraser.** 2000. Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. Mol. Cell **5:**377–386.
17. **Gu, H., J. D. Gupta, and D. R. Schoenberg.** 1999. The poly(A)-limiting element is a conserved cis-acting sequence that regulates poly(A) tail length on nuclear pre-mRNAs. Proc. Natl. Acad. Sci. USA **96:**8943–8948.
18. **Gu, Y., T. Nakamura, H. Alder, R. Prasad, O. Canaani, G. Cimino, C. M. Croce, and E. Canaani.** 1992. The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to Drosophila trithorax, to the AF-4 gene. Cell **71:**701–708.
19. **Gu, H., and D. R. Schoenberg.** 2003. U2AF modulates poly(A) length control by the poly(A)-limiting element. Nucleic Acids Res. **31:**6264–6471.
20. **Hadland, B. K., S. S. Huppert, J. Kanungo, Y. Xue, R. Jiang, T. Gridley, R. A. Conlon, A. M. Cheng, R. Kopan, and G. D. Longmore.** 2004. A requirement for *Notch1* distinguishes 2 phases of definitive hematopoiesis during development. Blood **104:**3097–3105.

21. **Hahn, W. C., and R. A. Weinberg.** 2002. Modelling the molecular circuitry of cancer. Nat. Rev. **2:**331–341.
22. **Hanahan, D., and R. A. Weinberg.** 2000. The hallmarks of cancer. Cell **100:**57–70.
23. **Hanna, Z., M. Janowski, P. Tremblay, X. Jiang, A. Milatovich, U. Francke, and P. Jolicoeur.** 1993. The Vin-1 gene, identified by provirus insertional mutagenesis, is the cyclin D2. Oncogene **8:**1661–1666.
24. **Hoemann, C. D., N. Beaulieu, L. Girard, N. Rebai, and P. Jolicoeur.** 2000. Two distinct *Notch1* mutant alleles are involved in the induction of T-cell leukemia in c-myc transgenic mice. Mol. Cell. Biol. **20:**3831–3842.
25. **Hollander, M. C., I. Alamo, and A. J. Fornace Jr.** 1996. A novel DNA damage-inducible transcript, gadd7, inhibits cell growth, but lacks a protein product. Nucleic Acids Res. **24:**1589–1593.
26. **Hope, T. J.** 1999. The ins and outs of HIV Rev. Arch. Biochem. Biophys. **365:**186–191.
27. **Hwang, H. C., C. P. Martins, Y. Bronkhorst, E. Randel, A. Berns, M. Fero, and B. E. Clurman.** 2002. Identification of oncogenes collaborating with p27Kip1 loss by insertional mutagenesis and high-throughput insertion site analysis. Proc. Natl. Acad. Sci. USA **99:**11293–11298.
28. **Ikegawa, S., M. Isomura, Y. Koshizuka, and Y. Nakamura.** 1999. Cloning and characterisation of ASH2L and Ash21, human and mouse homologs of the drosophila ash2 gene. Cytogenet. Cell. Genet. **84:**167–172.
29. **Johnson, C., P. A. Lobelle-Rich, A. Puetter, and L. S. Levy.** 2005. Substitution of feline leukemia virus long terminal repeat sequences into murine leukemia virus alters the pattern of insertional activation and identifies new common insertion sites. J. Virol. **79:**57–66.
30. **Joosten, M., Y. Vankan-Berkhoudt, M. Tas, M. Lunghi, Y. Jenniskens, E. Parganas, P. J. M. Valk, E. Van den Akker, and R. Delwel.** 2002. Large-scale identification of novel potential disease loci in mouse leukemia applying an improved strategy for cloning common virus integration site. Oncogene **21:**7247–7255.
31. **Kaplan, H. S.** 1967. On the natural history of the murine leukemias: presidential address. Cancer Res. **27:**1325–1340.
32. **Kim, R., A. Trubetskoy, T. Suzuki, N. A. Jenkins, N. G. Copeland, and J. Lenz.** 2003. Genome based identification of cancer genes by proviral tagging in mouse retrovirus-induced T-cell lymphomas. J. Virol. **77:**2056–2062.
33. **Krause, M. O.** 1996. Chromatin structure and function: the heretical path to an RNA transcription factor. Biochem. Cell. Biol. **74:**623–632.
34. **Kuhn, J., and S. Binder.** 2002. RT-PCR analysis of 5′ and 3′-end-ligated mRNAs identifies the extremities of *cox2* transcripts in pea mitochondria. Nucleic Acids Res. **30:**439–446.
35. **Lambert, J., D. Bergeron, C. A. Kozak, and E. Rassart.** 1999. Identification of a common site of provirus integration in radiation leukemia virus-induced T-cell lymphomas in mice. Virology **264:**181–186.
36. **Lanz, R. B., N. J. McKenna, S. A. Onate, U. Albrecht, J. Wong, S. Y. Tsai, M. J. Tsai, and B. W. O'Malley.** 1999. A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. Cell **97:**17–27.
37. **Lazo, P. A., J. S. Lee, and P. N. Tsichlis.** 1990. 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. USA **87:**170–173.
38. **Lee, J. T., S. L. Davidow, and D. Warshawsky.** 1999. Tsix, a gene antisense to Xist at the X-inactivation center. Nat. Genet. **21:**400–404.
39. **Li, J., H. Shen, K. L. Himmel, A. J. Dupuy, D. A. Largaespada, T. Nakamura, J. D. Jr Shaughnessy, N. A. Jenkins, and N. G. Copeland.** 1999. Leukaemia disease genes: large-scale cloning and pathway predictions. Nat. Genet. **23:** 348–353.
40. **Lieberman, M., A. Declève, E. P. Gellmann, and H. S. Kaplan.** 1977. Biological and serological characterization of the C-type RNA viruses isolated from C57BL/Ka strain of mice. II. Induction and propagation of the isolates in radiation-induced leukemogenesis and related viruses, p. 231–246. INSERM Symposium 4, Paris, France.
41. **Lieberman, M., A. Declève, P. Ricciardi-Castagnoli, J. Bonier, O. J. Finn, and H. S. Kaplan.** 1979. Establishment, characterization and virus expression of cell lines derived from radiation-and virus-induced lymphomas of C57BL/Ka mice. Int. J. Cancer **24:**168–177.
42. **Lieberman, M., O. Niwa, A. Declève, and H. S. Kaplan.** 1973. Continuous propagation of radiation leukemia virus on a C57BL mouse embryo fibroblast line with attenuation of leukemogenic activity Proc. Natl. Acad. Sci. USA **70:**1250–1253.
43. **Lower, K. M., G. Turner, B. A. Kerr, K. D. Mathews, M. A. Shaw, A. K. Gedeon, S. Schelley, H. E. Hoyme, S. M. White, M. B. Delatycki, A. K. Lampe, J. Clayton-Smith, H. Stewart, C. M. van Ravenswaay, B. B. de Vries, B. Cox, M. Grompe, S. Ross, P. Thomas, J. C. Mulley, and J. Gecz.** 2002. Mutations in PHF6 are associated with Börjeson-Forssman-Lehmann syndrome. Nat. Genet. **32:**661–665.
44. **Lower, K. M., G. Solders, M. L. Bondeson, J. Nelson, A. Brun, J. Crawford, G. Malm, M. Borjeson, G. Turner, M. Partington, and J. Gecz.** 2004. 1024C>T (R342X) is a recurrent PHF6 mutation also found in the original Börjeson-Forssman-Lehmann syndrome family. Eur. J. Hum. Genet. **12:** 787–789.
45. **Lund, A. H., G. Turner, A. Trubetskoy, E. Verhoeven, E. Wientjens, D. Hulsman, R. Russel, R. A. DePinho, J. Lenz, and M. Van Lohuizen.** 2002. Genome wide retroviral insertional tagging of genes involved in cancer in Cdkn2a-deficient mice. Nat. Genet. **32:**160–165.
46. **Lyngso, C., G. Bouteiller, C. K. Damgaard, D. Ryom, S. Sanchez-Munoz, P. L. Norby, B. J. Bonven, and P. Jorgensen.** 2000. Interaction between the transcritption factor SPBP and the positive cofactor RNF4: an interplay between protein binding zing fingers. J. Biol. Chem. **275:**26144–26149.
47. **Maillard, I., A. P. Weng, A. C. Carpenter, C. G. Rodriguez, H. Sai, L. Xu, D. Allman, J. C. Aster, and W. S. Pear.** 2004. Mastermind critically regulates Notch-mediated lymphoid cell fate decisions. Blood **104:**1696–1702.
48. **Martin-Hernandez, J., A. B. Sorensen, and F. S. Pedersen.** 2001. Murine leukemia virus proviral insertions between the N-*ras* and *unr* genes in B-cell lymphoma DNA affect the expression of N-*ras* only. J. Virol. **75:**11907–11912.
49. **Meller, V. H., K. H. Wu, G. Roman, M. I. Kuroda, and R. L. Davis.** 1997. roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system. Cell **88:**445–457.
50. **Mikkers, H., J. Allen, P. Knipscheer, L. Romeyn, A. Hart, E. Vink, and A. Berns.** 2002. High-throughput retroviral tagging to identify components of specific signaling pathways in cancer. Nat. Genet. **32:**153–159.
51. **Nakamura, T., H. Alder, Y. Gu, R. Prasad, O. Canaani, N. Kamada, R. P. Gale, B. Lange, W. M. Crist, P. C. Nowell, et al.** 1993. Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leukemia share sequence homology and/or common motifs. Proc. Natl. Acad. Sci. USA **90:**4631–4635.
52. **Nesterova, T. B., C. M. Johnston, R. Appanah, A. E. Newall, J. Godwin, M. Alexiou, and N. Brockdorff.** 2003. Skewing X chromosome choice by modulating sense transcription across the Xist locus. Genes Dev. **17:**2177–2190.
53. **Ogawa, Y., J. T. Lee.** 2003. Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. Mol. Cell **11:**731–743.
54. **Olsen, P. H., and V. Ambros.** 1999. The *lin-4* regulatory RNA controls developmental timing in Caenorabditis elegans by blocking LIN-14 protein synthsis after the initiation of translation. Dev. Biol. **216:**671–680.
55. **Olson, M. O., M. Dundr, and A. Szebeni.** 2000. The nucleolus: an old factory with unexpected capabilities. Trends Cell. Biol. **10:**189–196.
56. **Parry, P., M. Djabali, M. Bower, J. Khristich, M. Waterman, et al.** 1993. Structure and expression of the human trithorax-like gene 1 involved in acute leukemia. Proc. Natl. Acad. Sci. USA **90:**4738–4742.
57. **Peters, J.** 2000. Imprinting: silently crossing the boundary. Genome Biol. **1:**1028.
58. **Phillips, C., S. Jung, and S. I. Gunderson.** 2001. Regulation of nuclear poly(A) addition controls the expression of immunoglobulin M secretory mRNA. EMBO J. **20:**6443–6452.
59. **Poliquin, L., D. Bergeron, J. L. Fortier, Y. Paquette, R. Bergeron, and E. Rassart.** 1992. Determinants of thymotropism in Kaplan radiation leukemia virus and nucleotide sequence of its envelope region. J. Virol. **66:**5141–5146.
60. **Rassart, E., P. Sankar-Mistry, G. Lemay, L. Desgroseillers, and P. Jolicoeur.** 1983. New class of leukemogenic ecotropic recombinant murine leukemia virus isolated from radiation induced thymomas of C57BL/Ka mice. J. Virol. **45:**565–575.
61. **Rassart, E., M. Shang, Y. Boie, and P. Jolicoeur.** 1986. Studies on emerging radiation leukemia virus variants in C57BL/Ka mice. J. Virol. **58:**96–106.
62. **Reinhart, B. J., F. J. Slack, M. Basson, A. E. Pasquinelli, J. C. Bettinger, A. E. Rougvie, H. R. Horvitz, and G. Ruvkun.** 2000. The 21-nucleotide *let-7* RNA regulates developmental timing in Caenorhabditis elegans. Nature **403:**901–906.
63. **Sambrook, J., and D. W. Russell.** 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
64. **Selten, G., H. T. Cuypers, and A. Berns.** 1985. Proviral activatoin of the putative oncogene *Pim*-1 in MuLV induced T-cell lymphomas. EMBO J. **4:**1793–1798.
65. **Selten, G., H. T. Cuypers, M. Zijlstra, C. Melief, and A. Berns.** 1984. Involvement of c-Myc in MuLV-induced T cell lymphomas in mice: frequency and mechanisms of activation. EMBO J. **3:**3215–3222.
66. **Shibata, S., and J. T. Lee.** 2003. Characterisation and quantitation of differential Tsix transcripts: implication for *Tsix* function. Hum. Mol. Genet. **12:**125–136.
67. **SØrensen, A. B., M. Duch, H. W. Amtoft, P. J. Ørgensen, and F. S. Pedersen.** 1996. Sequence tags of provirus integration sites in DNAs of tumors induced by the murine retrovirus SL3-3. J. Virol. **70:**4063–4070.
68. **Suzuki, T., H. Shen, K. Akabi, H. C. Morse III, J. D. Malley, D. Q. Naiman, N. A. Jenkins, and N. G. Copeland.** 2002. New genes involved in cancer identified by retroviral tagging. Nat. Genet. **32:**166–174.
69. **Tkachuk, D. C., S. Kohler, amd M. L. Cleary.** 1992. Involvement of a homologue of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias. Cell **71:**691–700.
70. **Tremblay, P. J., C. A. Kozak, and P. Jolicoeur.** 1992. Identification of a novel gene, *Vin-1*, in murine leukemia virus-induced T-cell leukemias by provirus insertional mutagenesis. J. Virol. **66:**1344–1353.
71. **Van Lohuizen, M., S. Verbeek, P. Krimperfort, J. Domen, C. Saris, T. Radaszkiewicz, and A. Berns.** 1989. Predisposition to lymphomagenesis in

pim-1 transgenic mice: cooperation with c-myc and N-myc in murine leukemia virus-induced tumors. Cell **56:**673–682.

72. **Wallace, A. M., D. Brinda, S. E. Ravnik, V. Tonk, N. A. Jenkins, D. J. Gilbert, N. G. Copeland, and C. C. MacDonald.** 1999. Two distinct forms of the 64,000 protein of the cleavage stimulation factor are expressed in mouse male germ cells. Proc. Natl. Acad. Sci. USA **96:**6763–6768.
73. **Wang, Y., D. R. Crawford, and K. J. A. Davies.** 1996. *adapt33*, a novel oxidant-inducible RNA from hamster HA-1 cells. Arch. Biochem. Biophys. **332:**255–260.
74. **Weng, A. P., A. A. Ferrando, W. Lee, J. P. Morris 4th, L. B. Silverman, C. Sanchez-Irizarry, S. C. Blacklow, A. T. Look, and J. C. Aster.** 2004. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science **306:**269–271.
75. **Xu, N., C. Y. Chen, and A. B. Shyu.** 1997. Modulation of the fate of cytoplasmic mRNA by AU-rich elements: key sequence features controlling mRNA deadenylation and decay. Mol. Cell. Biol. **17:**4611–4621.
76. **Yanagawa, S., Lee, J. S., Kakimi, K., Matsuda, Y., Honjo, T., and A. Ishimoto.** 2000. Identification of *Notch1* as a frequent target for provirus insertional mutagenesis in T-cell lymphomas induced by leukemogenic mutants of mouse mammary tumor virus. J. Virol. **74:**9786–9791.