Genotypic Features of Lentivirus Transgenic Mice\ † ‡

Marc-Olivier Sauvain,1 Alexander P. Dorr,1‡ Brian Stevenson,2 Alexandra Quazzola,1 Félix Naef,1 Maciej Wiznerowicz,1 Frédéric Schütz,2 Victor Jongeneel,2 Denis Duboule,1,3 François Spitz,5§ and Didier Trono1*  

School of Life Sciences and Frontiers in Genetics National Center for Competence in Research, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland;1 Ludwig Institute for Cancer Research and Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland;2 and Department of Zoology and Animal Biology, University of Geneva, Geneva, Switzerland3

Received 20 March 2008/Accepted 30 April 2008

Lentivector-mediated transgenesis is increasingly used, whether for basic studies as an alternative to pronuclear injection of naked DNA or to test candidate gene therapy vectors. In an effort to characterize the genotypic features of this approach, we first measured the frequency of germ line transmission of individual proviruses established by infection of fertilized mouse oocytes. Seventy integrants from 11 founder (G0) mice were passed to 111 first generation (G1) pups, for a total of 255 events corresponding to an average rate of transmission of 44%. This implies that integration had most often occurred at the one- or two-cell stage and that the degree of genotypic mosaicism in G0 mice obtained through this approach is generally minimal. Transmission analysis of eight individual proviruses in 13 G2 mice obtained by a G0-G1 cross revealed only 8% of proviral homozygosity, significantly below the 25% expected from purely Mendelian transmission, suggesting counter-selection due to interference with the functions of targeted loci. Mapping of 239 proviral integration sites in 49 founder animals revealed that about 60% resided within annotated genes, with a marked tendency for clustering in the middle of the transcribed region, and that integration was not influenced by the transcriptional orientation. Transcript levels of a set of arbitrarily chosen target genes were significantly higher in two-cell embryos than in embryonic stem cells or adult somatic cells, suggesting that, as previously noted in other settings, lentiviral vectors integrate preferentially into regions of the genome that are transcriptionally active or poised for activation.

Transgenic animals are essential research tools, whether they are used to address basic biological questions or to develop preclinical models of human diseases. Their generation through the injection of naked plasmid DNA into the male pronucleus of fertilized oocytes has been a standard practice for almost 3 decades (5), but the success of this procedure has been largely limited to mice. Recently, lentiviral vector-mediated transgenesis has emerged as an attractive alternative, as these human immunodeficiency virus (HIV)-derived gene transfer vehicles can mediate the efficient integration of their cargo into zygotes or early progenitors from a wide variety of species including mice, rats, pigs, cows, and chickens (4, 8, 13, 20, 23). In a typical procedure, the lentivirus particle is injected beneath the zona pellucida of a fertilized oocyte and penetrates this cell by fusion between the viral and plasma membranes. Its RNA genome then undergoes reverse transcription, yielding a double-stranded DNA copy that integrates into the host cell chromosomes.

The kinetics of lentivirus integration into the target genome defines the degree of genotypic mosaicism found in the resulting animal. Integration events occurring subsequent to the first cell division will, indeed, limit the presence of individual proviruses to only a subset of cells and, hence, condition their rate of transmission to the first generation (G1) progeny. Even though the phenotypic consequences of this mosaicism can usually be minimized through the use of vector doses high enough to induce multiple proviral copies per embryo, ensuring that all cells harbor at least one integrant, individual proviruses are subjected to different influences conditioned by their site of integration into the host genome (8). Furthermore, the integration site of individual proviruses will influence their potential to exert cis-acting effects on the host genome, a process known as insertional mutagenesis. Large-scale analyses of retroviral integration sites in somatic cells have revealed that, while vectors derived from both murine leukemia virus (MLV) and HIV favor active genes, the former tend to integrate in and around promoters, whereas the latter rather target transcribed regions (14, 19, 30). Noteworthy, when MLV vectors are used to infect early mouse embryos, they are rapidly silenced during development, in contrast to their lentiviral counterparts (9). It is not known whether this partly reflects differences in integration site selection or results solely from the sequence-specific recruitment of epigenetic repressors, such as recently demonstrated for the primer binding site-dependent KAP1/TRIM28-mediated silencing of MLV vectors in embryonic stem cells (28, 29).

The present work presents a genotypic characterization of transgenic mice obtained from oocytes injected with lentiviral vectors. It demonstrates that integration occurs rapidly follow-
ing this procedure, so that the degree of mosaicism is low in the resulting animals, and that proviruses are located preferentially in genes likely to be active at the time of infection.

MATERIALS AND METHODS

Lentiviral vector-mediated transgenesis. We generated transgenic mice by perivitelline injection of fertilized oocytes, as described previously (13). Briefly, the hybrid strain B6D2G1 derived from C57BL/6JxDBA2J mice (Charles River, France) was used as an egg donor. Superovulation was induced by intraperitoneal injection of five B6D2G1 females with 10 IU of pregnant mare serum (Sigma, Buchs, Switzerland). Forty-six hours later, a second intraperitoneal injection was performed with 10 IU of human chorionic gonadotropin (Sigma) before mice were mated with B6D2G1 males. Sixteen hours later, oocytes were harvested, injected in the perivitelline space with a highly concentrated vector stock (5 × 10^9 to 1 × 10^10 HeLa-transducing units/ml), and kept in culture overnight at 37°C and 5% CO_2 in KSOM medium (MR-121; Specialty Media). Embryos that had reached the two-cell stage were then placed in the ampulla of foster NMRI mothers (eight embryos per ampulla). All vectors used in this study were previously described (27). Genotyping of the offspring was done by PCR using vector-specific primers (forward primer, 5′-TATGTTGCTCCTTTTACGTATGTG-3′; reverse primer, 5′-CGACACACACACCGAATTGT-3′).

LAM-PCR. DNA of transgenic mice was purified by standard phenol-chloroform extraction and ethanol precipitation, and DNA of NIH 3T3 fibroblasts grown in Dulbecco’s modified Eagle’s medium–10% fetal calf serum (Invitrogen, Paisley, United Kingdom) was purified with a DNeasy extraction kit according to the manufacturer’s recommendation (Qiagen, Hilden, Germany). Linear amplification-mediated (LAM)-PCR was carried out as described previously (18) with minor modifications. Five nanograms of genomic DNA was mixed in a 50-μl reaction volume with 1 μl Taq DNA polymerase buffer (Qiagen), a 200 μM concentration of the deoxyribonucleoside triphosphates (dNTPs) (Sigma, Buchs, Switzerland), a 2.5 nM concentration of oligonucleotides AD30 and AD31 (for all primers, see Table S3 in the supplemental material), and 2.5 U of Taq polymerase (Qiagen). Primers were extended after initial denaturation for 5 min at 94°C (hot start), followed by 50 cycles of 94°C for 1 min, 60°C for 45 s, and 72°C for 1.5 min, with a final extension at 72°C for 10 min. Dynal kilobaseBINDER beads (Invitrogen) were washed two times with 0.1% bovine serum albumin in phosphate-buffered saline, resuspended in 50 μl of binding buffer provided with the kit, and added to the PCR at a final concentration of 20 μl of beads per PCR. Extension products were captured for 1 h at room temperature under agitation and washed with water. Second strands were synthesized in 20-μl reaction mixtures containing 1× hexanucleotide buffer (Roche, Rotkreuz, Switzerland), 30 μM dNTPs, and 2 U of Klenow (Roche) for 1 h at 37°C. Beads were washed with water and digested in a 20-μl reaction containing 4 U of Tsp509I (New England Biolabs, Allschwil, Switzerland) for 1 h at 65°C. Beads were washed with water and ligated in a 10-μl reaction mixture containing 1× Fast-Link buffer (Epicenter, Dottikon, Switzerland), 1 mM ATP, 2 μl of AD25/AD26 linker cassette, and 2 U of Fast-Link ligase for 5 min at room temperature. Beads were washed with water and denatured for 10 min at room temperature with 5 μl of 0.1 M NaOH. Two microliters of this LAM product was exponentially amplified in 50-μl reaction mixtures containing 1× Taq DNA polymerase buffer (Qiagen), 200 μM dNTPs (Sigma), 500 nM concentrations of oligonucleotides AD27 and AD32, and 5 U of Taq polymerase (Qiagen). Primers were extended after initial denaturation for 3 min at 94°C (hot start) followed by 35 cycles of 94°C for 45 min,
Cloning, sequencing, and annotation of LAM-PCR products. PCRs were pu-
fified with a Nucleobond II kit (Macherey-Nagel, Switzerland) and eluted in 30 l of distilled 
water in a 20-l reaction volume at 37°C. Fifteen microliters of this eluate was digested 
overnight with 5 U of NarI (NEB) and 0.5 l of Tris-HCl, pH 8.0. Fifteen microliters of the final PCR product was mixed with 0.5 l of ROX-

GeneScan standard (Applied Biosystems, Switzerland) and 8.5 l of distilld 
formamide (Applied Biosystems) and denatured for 2 min at 95°C. PCRs were 
separated on a 3100 sequencer (Applied Biosystems) and analyzed with the 
GeneScan software.

FIG. 3. Lentivirus transmission detected by ISS-PCR. (A) Scheme of primer sets used for ISS-PCR on tail DNA targeting the integrated 
provirus or the integration site in G1. Primer set A plus C surrounds the lentivirus integration site (locus), and primer set B plus C targets the 
lentiviral vector and the downstream genome sequence (junction). (B) PCR analysis of the offspring from mouse 2267 crossed with a 
wild-type (wt) animal. LV, lentivirus.

60°C for 45 min, and 72°C for 45 s, with a final extension at 72°C for 10 min. One 
microliter of this PCR product was amplified in a 50-l reaction mixture with 1× 
Taq DNA polymerase buffer (Qiagen), 200 μM dNTPs (Sigma), 500 nM AD33 
(oligonucleotide), 400 nM AD28, 100 nM AD62, and 5 U of Taq polymerase 
(Qiagen). One microliter of the final PCR product was mixed with 0.5 l of ROX-

GeneScan standard (Applied Biosystems, Switzerland) and 8.5 l of distilld 
formamide (Applied Biosystems) and denatured for 2 min at 95°C. PCRs were 
separated on a 3100 sequencer (Applied Biosystems) and analyzed with the 
GeneScan software.

Cloning, sequencing, and annotation of LAM-PCR products. PCRs were pu-
fified with a Nucleobond II kit (Macherey-Nagel, Switzerland) and eluted in 30 l of distilled 
water in a 20-l reaction volume at 37°C. Fifteen microliters of this eluate was digested 
overnight with 5 U of NarI (NEB) in a 20-l reaction volume at 37°C. Of this 
digest, 0.6 l was mixed in 96-well plates with 0.5 l of pCR4-TOPO (Invitrogen) 
and 0.4 l of salt solution in a final volume of 2.3 l. After incubation for 5 min 
at room temperature, reaction mixtures were heat shocked into DH5α cells 
provided with the kit (Invitrogen). Colonies were picked into 96-well plates and 
sequenced at GATC Biotech, Konstanz, Germany. Sequences were extracted 
from chromatograms using phred (3). Vector and HIV-contaminating sequences 
plus US and LC fragments were identified with BLAT (12) and removed prior to 
annotation. Mouse repeats in the cleaned sequences were identified and masked 
The location of each cleaned, masked sequence in the mouse genome (NCBI-b34 
assembly) was determined with the Megablast program (32) using a word length 
of 24. The annotation procedure was automated using a series of ad hoc Perl 
scripts. Gene-relative mappings were determined using the ENSEMBL database 
(http://oct2006.archive.ensembl.org/). We defined genes as transcription units 
annotated in the ENSEMBL database; data tables were assembled in Open-

Office, version 2.0, and statistical analysis was performed with R or PRISM 
(version 4.0) software.

ISS-PCR. Integration site-specific PCRs (ISS-PCRs) were exponentially am-
plicated in 25-μl reaction mixtures containing 100 ng of DNA, 1.5× Taq Polymer-

ase buffer (Qiagen), 200 μM dNTPs (Promega), 0.8 μg of bovine serum albumin 
(NEB), 25 mmol MgCl2 (Qiagen), 5 U of HotStart Taq polymerase (Qiagen), 
and 200 nM concentrations of each primer. The PCR started with an initial 
activation step of 95°C for 15 min (hot start), followed by 10 cycles of 94°C for 
45 s, 60°C (with a change of ~1°C per cycle) for 45 s, and 72°C for 1 min 
touchdown); this program was followed by 30 cycles of 94°C for 45 s, 50°C for 
45 s, and 72°C for 1 min, with a final extension at 72°C for 10 min. Five milliliters 
of loading buffer (6×) was added to each PCR, and 15 μl was run on a 0.8% 
agarose gel for about 30 min at 80V before analysis under UV light.

RNA expression measurements. NIH 3T3 fibroblasts were cultured in Dul-
becco’s modified Eagle’s medium–10% fetal calf serum (Invitrogen) at 37°C in 
5% CO2. Oct4-GFP (kind gift A. Smith, Edinburgh, United Kingdom) embryonic 
stem (ES) cells were cultured as described previously (31) in the presence of 500 
ng of puromycin (Sigma)/ml at 37°C in 10% CO2. RNA was isolated with Trizol 
(Invitrogen) and ethanol precipitation according to standard procedures; cDNA 
was synthesized with hexanucleotides according to the manufacturer’s recom-

mendation using a SuperScript III first-strand cDNA synthesis system (Invitro-
gen). Fifty unfertilized eggs or two-cell embryos were resuspended in 1 μl of 
RNase Out and 10 μl of resuspension buffer (SuperScript III CellsDirect cDNA 
synthesis system; Invitrogen) and denatured for 10 min at 70°C. DNase digestion 
was carried out according to the manufacturer’s instructions, and 2 μl of 50 ng/μl 
hexamer was added with 1 μl of dNTPs. Reaction mixtures were denatured at 
70°C for 5 min, and the remaining reagents were added as indicated in the 
protocol. Reaction mixtures were left for 10 min at 25°C before reverse tran-
scription. RNase H-digested cDNAs were used for quantitative PCR. The 
expression of each gene was assayed in triplicate in a total volume of 5 μl contain-
ing 1× Power Sybr (Applied Biosystems), a 200 nM concentration of each 
gene-specific primer pair (see Table S3 in the supplemental material), and 
diluted cDNA (cultured cells, 1:17; embryonic cDNA, 1:2.6). To verify specificity, 
each PCR was followed by a melting curve analysis, and samples lacking reverse 
transcriptase were run in parallel. The increase in fluorescence was analyzed with 
SDS software, version 2.2.2. (Applied Biosystems). For all amplification plots, 
the baseline data were set with the automatic cycle threshold function available 
with SDS, version 2.2.2, calculating the optimal baseline range and threshold 
values using the AutoCt algorithm (SDS version 2.2 user’s manual, Applied 
Biosystems, Foster City, CA). A mean quantity was calculated from triplicate 
PCRs for each sample, and this quantity was normalized to two similarly mea-
sured quantities of normalization genes as described previously (24). Normalized 
quantities were averaged for three replicates for each data point and are repre-
sented as the means ± standard deviations (SD). The highest normalized relative
quantity was arbitrarily given a value of 1.0. Relative changes were calculated from the quotient of the means of these normalized quantities and are reported as means ± SD.

RESULTS

Timing lentiviral vector integration in germ cell precursors.

The degree of mosaicism of mice obtained by lentivector-mediated transgenesis reflects the timing of integration in primary blastomeres. As a first assessment of this parameter, we injected an enhanced green fluorescent protein (eGFP)-expressing vector (Fig. 1A, pWPTS) into the perivitelline space of fertilized mouse oocytes and let the resulting embryos develop in vitro to the blastocyst stage. Upon fluorescence microscopy examination, both the inner cell mass and the trophectoderm revealed uniformly high levels of eGFP expression (Fig. 1B). While this result indicated that transcription proceeded efficiently from the EF1α/H9251 (locus PCR) and the other across the junction between the 3′ end of the provirus and the host DNA (junction PCR) (Fig. 3B). The combined results of these two PCRs indicated unequivocally whether a provirus was present in a mouse.

These techniques in hand, we generated 11 transgenic founders through as many independent sessions of injection; five carried the pLV-tTRKRAB-red vector, and six carried the pRRL-GFP vector (Fig. 1A). These animals were crossed with a wild-type, animal, yielding a total of 111 pups. Seventy proviral insertions were mapped in the G0 mice, and 255 were mapped in the G1 mice. Overall, the average rate of transmission of individual proviruses was 44% (Fig. 4A). When mice were grouped according to the vector used, germ line transmission rates were 48% for pRRL-GFP and 34% for pLV-tTRKRAB-red (Fig. 4B). Of note, none of the vectors induced a drop in the size of the progeny in these or in other experiments, indicating that, as heterozygous provirus, none of the vectors exerted major phenotoxic effects that could have invalidated our genotypic analyses.

We also characterized the distribution of lentiviral integrants in G2 animals obtained by G1-G0 backcrossing of the 2267 lineage by ISS-PCR. Eight integrants could be success-

TABLE 1. Integration sites of G0 mouse 2267 outside annotated genes

<table>
<thead>
<tr>
<th>Integrant identifier</th>
<th>Chromosome</th>
<th>Position (nt)</th>
<th>Strand</th>
<th>Distance to nearest gene (nt)</th>
<th>Sense of transcription</th>
<th>ENSEMBL identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>2267-132D</td>
<td>1</td>
<td>23905241</td>
<td>+</td>
<td>101,830</td>
<td>Upstream</td>
<td>ENSMUSG00000026156</td>
</tr>
<tr>
<td>2267-132R3</td>
<td>X</td>
<td>46760215</td>
<td>+</td>
<td>33,008</td>
<td>Upstream</td>
<td>ENSMUSG000000055653</td>
</tr>
<tr>
<td>2267-132E</td>
<td>10</td>
<td>103196281</td>
<td>+</td>
<td>45,568</td>
<td>Downstream</td>
<td>ENSMUSG00000019892</td>
</tr>
<tr>
<td>2267-132R2</td>
<td>2</td>
<td>150665634</td>
<td>+</td>
<td>165,266</td>
<td>Upstream</td>
<td>ENSMUSG00000048918</td>
</tr>
<tr>
<td>2267-132C</td>
<td>15</td>
<td>69704246</td>
<td>−</td>
<td>578,594</td>
<td>Downstream</td>
<td>ENSMUSG00000022332</td>
</tr>
<tr>
<td>2267-132D</td>
<td>1</td>
<td>23905241</td>
<td>+</td>
<td>101,830</td>
<td>Upstream</td>
<td>ENSMUSG00000026156</td>
</tr>
</tbody>
</table>

TABLE 2. Integration sites of G0 mouse 2267 inside annotated genes

<table>
<thead>
<tr>
<th>Integrant identifier</th>
<th>Chromosome</th>
<th>Position (nt)</th>
<th>Strand</th>
<th>Sense of TUa</th>
<th>ENSEMBL identifier</th>
<th>Geneb</th>
</tr>
</thead>
<tbody>
<tr>
<td>2267-132A</td>
<td>5</td>
<td>61423827</td>
<td>+</td>
<td>−</td>
<td>ENSMUSG00000073999</td>
<td>CenT1</td>
</tr>
<tr>
<td>2267-132E</td>
<td>1</td>
<td>8362904</td>
<td>+</td>
<td>−</td>
<td>ENSMUSG00000067894</td>
<td>No description</td>
</tr>
<tr>
<td>2267-132L</td>
<td>3</td>
<td>58744450</td>
<td>−</td>
<td>+</td>
<td>ENSMUSG00000056476</td>
<td>Med12L</td>
</tr>
<tr>
<td>2267-132G</td>
<td>4</td>
<td>8753066</td>
<td>−</td>
<td>+</td>
<td>ENSMUSG00000050506</td>
<td>No description</td>
</tr>
<tr>
<td>2267-132H</td>
<td>4</td>
<td>10990240</td>
<td>−</td>
<td>−</td>
<td>ENSMUSG00000050323</td>
<td>XP_131300</td>
</tr>
</tbody>
</table>

a TU, transcription unit.
b From the Mouse Genome Informatics database.
fully amplified using this technique. In the G2 animals, integrants 2, 4, 5, 8, and 11 yielded a positive result by junction PCR and a negative result by locus PCR, indicating that the corresponding provirus was present in both alleles (Fig. 5A). Noteworthy, the frequency of homozygous integrants in G2 animals positively correlated with the number of proviral copies in G1 animals ($r^2 = 0.99$), reaching an overall value of 7.69% (8/104) in this particular case, which is significantly lower than expected from the 25% of a purely Mendelian

### TABLE 3. Integration sites of G0 mouse 2267 in repeats

<table>
<thead>
<tr>
<th>Integrant identifier</th>
<th>Class</th>
<th>Repeat</th>
</tr>
</thead>
<tbody>
<tr>
<td>2267-132R1</td>
<td>LTR/ERVL</td>
<td>MERVL$^b$</td>
</tr>
<tr>
<td>2267-132R2</td>
<td>LTR/ERVK</td>
<td>IAPEZ-int</td>
</tr>
<tr>
<td>2267-132R3</td>
<td>LINE/L1</td>
<td>Lx7</td>
</tr>
</tbody>
</table>

$^a$ ERVL/K, endogenous retrovirus repeat family.
$^b$ MERVL, mouse endogenous retrovirus repeat family.

---

**FIG. 4.** Frequency of G0-to-G1 transmission of lentiviral proviruses. (A) All integrants were pooled to obtain a global rate of transmission of lentiviral vector. Data are given for pooled integrants containing either pLV-tTRKRA8-red or pRRL-GFP from transgenic mice. (B) Point estimates represent the actual proportion of transmission for each integrant. Error bars indicate the 95% confidence interval obtained. The observations corresponding to different integrants in the same founder were pooled, assuming an equal probability of transmission for all integrants within the founder; the pooled estimates and their associated confidence intervals are shown for each family. The confidence intervals are narrower than those obtained for single integrants since they are based on larger sets of data. Finally, all the integrants of founders carrying the same promoter were pooled to obtain an average rate of transmission of individual integrants from this vector, with confidence interval.
transmission. However, there was no difference in the G1 transmission frequency of proviruses that did or did not reach homozygosity in G2 animals (Fig. 3), suggesting that these proviruses were not toxic in the heterozygous state.

Integration site selection. We then examined the in vivo integration site selection in lentivirus transgenic mice at a genomic level. We limited our mapping effort to LAM-PCR-generated amplicons containing at least 20 nucleotides (nt) of genomic sequence in addition to the 120 nt derived from the viral long terminal repeat (LTR) and linker cassettes, mapping successfully 239 sites in 49 G0 animals. We compared this integration data set to that of 108 lentiviral integrants obtained by transduction of murine 3T3 fibroblasts (see Tables S1 and S2 in the supplemental material). Proviruses were detectable in all chromosomes except Y (Fig. 6A). The likelihood of integration inside a gene was 1.75- and 1.88-fold higher in transgenic mice and 3T3 fibroblasts, respectively, than expected from the distribution of annotated genes in the mouse genome (mean value ± SD of 0.527 ± 0.255 for transgenic mice and 0.523 ± 0.274 for 3T3 fibroblasts). To compare these results to data previously obtained in human cells, we similarly analyzed 124 randomly picked, previously mapped HIV-1 integration sites in human SupT1 cells (19). In this collection as well, we observed the accumulation of lentiviral integrants in the middle of the genes (distance from transcription start/length of transcribed region, 0.500 ± 0.281).

One characteristic feature of mammalian genomes is the high content of repetitive elements. The mouse genome harbors three major classes of repeats: long interspersed nuclear elements (LINEs), short interspersed nuclear elements, and LTRs (11). We found that the percentage of integration into repeats closely corresponded to their relative abundance in the mouse genome (36.0% versus 38.0%) (26). Integration into repeats was slightly lower in 3T3 fibroblasts (30.8%), but the difference between these cells and cells of transgenic animals was not statistically significant. Comparison of the fraction of lentiviral vectors integrated into different repeat classes

FIG. 5. Transmission of lentiviral vectors to G2 mice. (A) G2 animals obtained from crossing G0 animal 2267 with three G1 animals. G2 animals were analyzed by the same PCR as described in the legend of Fig. 3. The asterisks mark gene-specific PCRs that failed to amplify the wild-type locus (mice carrying lentiviral vectors in both alleles). Every PCR was run with DNA from the founder animal amplifying all integration sites and a negative control from a nontransgenic animal (not shown). (B) The frequency (Freq.) of homozygous transmission of lentivirus integration sites 2, 4, 5, 8, and 11 in G2 animals correlates with the copy number of lentiviral vectors detected in the G1 animals. LV, lentivirus.
showed a slight preference for LTR, LINE/L1, and satellite repeats in the transgenic mice, possibly contributing to this marginal difference (Fig. 6E).

Integration appears to favor genes expressed during early embryogenesis. In order to test whether integration favored genes transcriptionally active during the preimplantation stages of the embryo, as previously observed in adult cells, we used quantitative PCR to compare the expression patterns of 10 randomly selected lentivirus target genes in two-cell embryos, ES cells, and 3T3 fibroblasts (Fig. 7A). Expression of these genes in two-cell embryos was from 2 to 42 times higher than that of housekeeping genes taken as controls (left panel). In ES cells and 3T3 fibroblasts, their expression was generally lower than in two-cell embryos (Fig. 7A), but this in part reflected higher average levels of RNA transcripts in the latter cells (9.5-fold difference; one-way analysis of variance, \( P < 0.0001 \)) (Fig. 7B). Additionally, we tested the impact of maternal RNA expression levels by comparing the expression of eight lentivirus target genes in unfertilized oocytes with expression in two-cell embryos. Two of the genes (5830435K17Rik and Bbs4) had similar expression levels in both settings, suggesting that their transcripts might be predominantly of maternal origin (Fig. 7C). In contrast, expression levels of the remaining six target genes were higher in two-cell embryos than in unfertilized oocytes.

**DISCUSSION**

PCR-based analyses of the kinetics of lentivirus/HIV integration previously revealed that, in human T-lymphoid cell lines, integration starts 4 h after infection and reaches its maximum after 16 h (22). Due to the low abundance of material at hand following infection of zygotes, we could not define this parameter directly in this setting. Instead, we examined the rates of transmission of individual proviruses from G0 transgenic mice to their G1 progeny. The overall rate of transmission of 44% for individual proviruses suggests that they were most often established at the one-cell stage after the S phase or at the two-cell stage before the S phase, since integration prior to S phase would transmit the provirus to both daughter cells, while integration after the S phase would transmit it to only one daughter cell. Considering the approximately 12-h division time of early blastomeres typical of our experimental conditions, the kinetics of lentivirus integration are thus not greatly delayed in these cells compared to T-lymphoid cells. A consequence is that the degree of mosaicism for individual integrants is minimal in transgenic mice obtained through this technique. This contrasts with data obtained in transgenic rats similarly generated, where a far higher degree of mosaicism was observed (23). While species-specific differences are possible, only a side-by-side comparison of similar vector preparations used to infect mouse and rat oocytes in parallel would properly address this question.

We also determined the genomic distribution of the proviral integrants generated by lentivector-mediated transgenesis. Following LAM-PCR-mediated amplification and sequencing of the host genome-provirus junctions, we could successfully locate more than 85% of the integration sites using a Perl script, which automatically removes cloning vector sequences, masks repetitive elements, and aligns the remaining polynucleotidic stretches with the mouse genome. This stepwise procedure enabled us to assign LAM-PCR-generated sequences to the mouse genome even when part of the ampli-con was constituted by repetitive sequences, a point verified by ISS-PCR. This latter method, which dually targets the provi-
rus-genome junction and the targeted locus, was previously used to trace retrovirally marked repopulating hematopoietic stem cells. Here, it allowed us to identify unequivocally transgenic G2 mice that were homozygous for specific lentiviral integrants following G0-G1 crosses. The observed frequency of homozygosity was below the 25% expected from a purely Mendelian mode of transmission. Considering the rapid kinetics of integration, it is unlikely that these non-Mendelian transmission rates reflect mosaicism in the germ line. Rather, some of the homozygous G2 mice must be nonviable due to the target gene inactivation potential of retrovirus integrants. However, the normal rates of G1 representation of proviruses that did not achieve homozygosity in G2 suggest an absence of toxicity in the heterozygous state. It remains that our series is very small and that a much larger study would certainly be needed to probe this issue and to confirm that homozygous lentiviral integrants are significantly counter-selected due to gene inactivation.

It was previously noted that the lentiviruses HIV, simian immunodeficiency virus, equine infectious anemia virus, and feline immunodeficiency virus and the vectors derived thereof exhibit similar integration patterns in human, monkey, and murine cells (1, 2, 6, 7, 10). Here, we extend the analysis to transgenic mice generated by infection of fertilized oocytes. We found that, in this case, HIV-derived lentiviral integrants favor genes, where they integrate with a tendency for the middle of the transcribed region, a trend already noted in human cells. One explanation could be that genes form a loop, with the transcriptional start and termination sites serving as bridging points and the tip of the loop protruding to the outside, which makes it more accessible for integration. Alternatively, occupancy of the transcriptional start and termination regions by regulatory proteins might interfere with this process. However, the accumulation of MLV integrants within or near promoters makes either one of these simple models unlikely. Instead, it is tempting to postulate that retroviral preintegration complexes interact with the RNA polymerase II holoenzyme and associated proteins and/or recognize specific histone modifications. Histone 3 dimethylation at lysine 4 correlates with active transcription and accumulates in the middle of genes, as do lentiviral integrants (17). In contrast, histone 3 monomethylation at lysine 4 and diacetylation at lysine 14 cluster at the transcriptional start sites, reminiscent of the MLV integration pattern. Thus, chromatin modifications or factors catalyzing these modifications might act as key players in the viral DNA-host genome interface, as recently suggested (25).

We observed a subtle difference in the integration patterns of lentiviral vectors in 3T3 fibroblasts and transgenic mice, with a slightly higher frequency of integration into repeats in the latter setting. Although this difference was statistically not significant, it might reflect the elevated transcriptional activity of repetitive elements in preimplantation embryos (15, 16, 21). In support of this model, we found that genes targeted by the lentiviral integrants in transgenic mice generally had higher levels of expression in two-cell embryos than in adult tissues. However, the constantly evolving picture of expressed regions of the genome, now known to encompass far more than conventional genes, as well as evidence indicating that genes that are poised for activation may share chromatin marks with actively transcribed genes, call for caution in establishing overly strict correlations, based on monomethylation at lysine 4, between “gene expression” and retroviral integration site selection.
ACKNOWLEDGMENTS

We thank S. Verp, M. Arcangeli, S. Liao, and J. Jakobsson (EPFL) for technical assistance or the gift of reagents; M. Docquier, C. De- lucing, and P. Docquier (WCR Frontiers in Genomics) for help with the genomics analyses; C. Prinz and M. Schmidt (University of Freiburg, Germany) for teaching us LAM-PCR; and M. Delorenzi (Swiss Institute of Bioinformatics) for advice on the statistics.

This work was supported by the European Union as part of the 6th Framework Program CONSERT project and by the Swiss National Science Foundation.

The authors declare that they have no conflict of interests.

REFERENCES