Distribution of targets for avian retrovirus DNA integration in vivo

Elizabeth S. Withers-Ward,1 Yoshihiro Kitamura,2 Joanne P. Barnes, and John M. Coffin3
Tufts University School of Medicine, Department of Molecular Biology and Microbiology, Boston, Massachusetts 02111 USA

The targeting of DNA integration in retrovirus-infected cells is a central yet very poorly understood aspect of the biology of the virus. To investigate this problem, we have assessed the use of specific sites for integration targets of avian leukosis virus (ALV) DNA within defined regions of turkey embryo fibroblast (TEF) cellular DNA. For this purpose, we developed an assay of sufficient sensitivity and specificity to allow detection and location of single integration events in a population of 5 million cells. Targets selected for study were either regions cloned by virtue of a previous integration event or clones chosen at random from cellular DNA. By use of this approach, we found that all genomic regions tested contained integration targets, with a frequency that varied from ~0.2 to 4 times that expected for random integration. Within regions, the frequency of use of specific sites varied considerably, with some sites used up to 280 times random frequency. When one region was introduced into cells at moderately high copy number by transfection, it provided integration targets in a pattern very much like that seen with the same sequence in vitro. On the basis of our sampling, we conclude that most or all regions of the TEF genome are accessible to ALV retroviral integration. As with integration in vitro, integration specificity seems to be determined largely by local structural features rather than accessibility of specific regions.

[Key Words: Retrovirus integration; ALV; integration targets; PCR]

Received December 24, 1993; revised version accepted May 2, 1994.

Integration of the provirus into the host genome is an intrinsic and essential feature of the retroviral life cycle. Despite considerable recent advances in understanding the biochemical mechanism of integration, the specificity of integration target site selection has not been clarified (for review, see Varmus and Brown 1989; Sandmeyer et al. 1990; Craigie 1992). Virus-encoded proteins such as integrase [IN], chromosomal DNA structure, host proteins, and nuclear functions may all be essential components of retroviral target site selection. In vivo studies with murine leukemia virus [MLV] and avian leukosis-sarcoma virus [ALV] indicate that retroviral integration is not sequence specific and that avian and murine retroviruses are able to integrate their DNA into many sites within the host DNA [Varmus and Swanstrom 1984]. Other in vivo studies have suggested a specificity in target site selection for structural features such as actively transcribed regions of chromatin [Mooslehner et al. 1990; Scherdin et al. 1990], DNase I hypersensitive regions [Robinson and Gagnon 1986; Vijaya et al. 1986; Goodenow and Hayward 1987; Rohdewold et al. 1987], or for certain specific sites within the host DNA of unknown structure [Shih et al. 1988]. However, these studies are based on small samples of cloned proviruses or cells that may be heavily biased by selection effects.

More information regarding insertion specificity has been obtained with the retrovirus-related transposable elements in yeast. The Ty3 element inserts into positions immediately upstream of the transcriptional start site of tRNA genes [Brodeur et al. 1983; Chalker and Sandmeyer 1990, 1992, 1993], probably reflecting a direct interaction of the integration and transcriptional machinery. Insertions of the retrotransposable element Ty1 are also found preferentially in regions associated with tRNA genes [Ji et al. 1993], although with less positional specificity. In studies performed to date in retrovirus systems, even strong targeting of this sort could have been completely overlooked because of the small samples of events analyzed.

The development of in vitro integration systems has provided a powerful tool with which to study the molecular basis of the integration process [Brown et al. 1987; Fujiwara and Mizuuchi 1988; Fujiwara and Craigie 1989; Bushman et al. 1990; Craigie et al. 1990; Ellison et al. 1990; Katz et al. 1990; Lee and Coffin 1990, 1991; Engelman et al. 1991]. When used in conjunction with PCR-based assays for mapping integration sites, they make possible a detailed analysis of target site usage within large populations of integration events. When we
used such an assay to study ALV target site selection in vitro, we found that target-site selection is highly non-random in naked DNA and is determined by local DNA structures rather than sequence or overall structure [Kitamura et al. 1992]. The targeting could be influenced by changes in structure. For example, strong targets for integration were created by methylation of runs of alternating CpG dinucleotide targets. Also, Pryciak and co-workers (Pryciak and Varmus 1992; Pryciak et al. 1992) have used a similar system to demonstrate that nucleosome-associated regions of minichromosomes are efficiently used as integration targets and that in the presence of phased nucleosomes, integration strongly alters specificity to clusters at 10-bp intervals. Taken together, these results argue strongly against a bias toward integration into transcriptionally active regions due to either hypomethylation of cytosines or a reduced concentration of nucleosomes.

In this report we extend the PCR-based approach to examine integration within chromosomal targets in vivo. These studies were initiated to test and extend our previous work on the specificity of ALV integration into cell DNA [Shih et al. 1988]. For this purpose, we developed a system that detects single integration events in a population of cells at any defined genomic site with single nucleotide resolution. Using this approach we have found that all regions of the genome examined were accessible to retroviral integration. Localized preferences were seen within regions, with some sites being used up to 280 times greater than random. However, we were unable to detect repeated integration into sites defined previously as hot spots (Shih et al. 1988). As with integration in vitro, integration specificity appeared to be determined by local structural features of the cellular DNA target.

Results

**In vivo integration site distribution assay**

The first question we addressed was whether integration sites are concentrated in specific cellular DNA regions, or distributed more or less at random throughout the genome. The strategy was to compare the frequency and distribution of integrations into regions identified previously as integration targets (cloned by virtue of an adjacent provirus) with regions cloned at random from the genome. For this purpose, we needed an assay that would allow us to rapidly screen a large number of recombination events, that was sensitive and specific enough to detect single-copy DNA within a background of many integrated proviruses, and that provided an accurate means of mapping events within a given region. To this end we modified the PCR-based assay, used previously for in vitro studies [Kitamura et al. 1992; Fig. 1]. The DNA product of integration, which could be either genomic or plasmid DNA, is amplified by use of primers derived from the viral LTR and cellular target sequences. These primer pairs permit amplification of virus–cell DNA junction fragments arising from integrations in a single orientation within −500 bp of the cellular primer. The PCR products are then used in a primer extension assay with a second, 32P-labeled, nested primer derived from cell sequences. The primer extension products are analyzed by electrophoresis in a sequencing gel alongside sequencing reactions of the same region of DNA.

To screen a large pool of independent integration events, turkey embryo fibroblast (TEF) cells were infected with CS8, a variant of ALV [Shih et al. 1988]. Four days after infection, high molecular weight DNA was isolated. The multiplicity of infection (m.o.i.) in integrated proviruses per cell was determined by Southern

---

**Figure 1.** The PCR-based integration site assay. Two oligonucleotide primers, one derived from viral U5 sequences and the other from TEF sequences (arrows), are used to amplify integration events within either preselected (left), or random single- or multicopy sequences (center), or within cloned TEF sequences supplied as the target in in vitro reactions (right). Following amplification, a second, nested, oligonucleotide primer also derived from TEF sequences is 5'-end-labeled with 32P and used in a primer extension reaction to specifically label PCR-amplified products. The primer extension products are analyzed by electrophoresis in a denaturing polyacrylamide gel. The length of each product defines the integration site within the region. The distribution of sizes of primer extended products in each lane provides a map or footprint of the distribution of integration events within the region examined.
Retrovirus integration specificity

blotting. In all experiments, estimates of the m.o.i. ranged from two to three integrated proviruses per cell. A pool of 10 μg of infected TEF DNA [from ~5×10^6 cells] would therefore contain ~1–1.5×10^7 integration events. If integration occurred at random throughout the genome [size 2×10^9 bp] we would expect to see about two to three integrations, in each orientation, within the 500-bp stretch of DNA analyzed in each reaction. Thus, in all of the experiments described below, random frequency of integration, as defined here, corresponds to about two to three bands per lane.

To test the sensitivity and specificity of the assay, reconstitution experiments were performed with primers derived from an integration region (number 13) described previously (Shih et al. 1988). Each reaction contained 10 μg of DNA from uninfected cells to which we added either 10, 1, or 0 copies of plasmid DNA containing cloned sequences of a provirus integrated within region 13. As seen in Figure 2A (lanes 25–34) we could readily detect a single molecule of proviral DNA integrated at this site in the presence of DNA from 5×10^6 cells at the frequency [7 of 10] expected for a Poisson distribution. No bands were seen in reactions containing only uninfected cell DNA (lanes 35–39). In other experiments [not shown] we have found identical sensitivity of detection of a cloned DNA control whether added to infected or uninfected cell DNA. Thus, the assay is capable of detecting a single integration event against a background of at least a 10^7-fold excess of sequences complementary to the DNA target in vitro. Despite the different sequence context, the two clones yielded identical patterns in the region of common sequence. Many integrations were distributed throughout the region with an obvious preference for some sites over others. One of the more intense bands was at the site detected as a frequent site in vivo (see arrow labeled 120×).

Figure 2A (lanes 5, 6) shows the integration site distribution when plasmid DNAs containing either 3 kb [lane 5] or 450 bp [lane 6] of region 13 sequences were used as the DNA target in vitro. Despite the different sequence context, the two clones yielded identical patterns in the region of common sequence. Many integrations were distributed throughout the region with an obvious preference for some sites over others. One of the more intense bands was at the site detected as a frequent site in vivo (see arrow labeled 120×).

Figure 2B–D shows the integration site distribution around preselected regions 29, 39, and 51 with their controls. All of these regions showed an integration site distribution with localized integration hot spots (arrows) similar in appearance to that seen with region 13. Frequent use of the same sites was seen in duplicate experiments with DNA from independent infections at a similar m.o.i. [data not shown]. The overall number of integrations within region 29 [Fig. 2B] was greater than the other three regions with a total of 42 events detected in 10 reactions [see Table 1 for an overall comparison to the other regions]. Again, we observed a localized preference for some sites over others [arrows], with one site used 280-fold more frequently than random.

To our surprise, no integrations were detected within regions 13 and 51 at the sites originally designated as preferred by screening integration site libraries (Shih et al. 1988). The control fragments used to standardize the system and test its sensitivity were derived from cloned proviruses integrated at these sites, and our ability to detect single copies of these fragments mixed with uninfected cell DNA in reconstruction experiments (Fig. 2A, lanes 25–34) argues strongly that we should have been able to detect integration at these sites if it had occurred.

Integration site distribution in random nonselected sites demonstrates a lack of regional inhibition in ALV target site selection

If integration targets displayed a strong regional bias, then we would have expected to see preferential use of
Figure 2. [See facing page for legend.]
target regions selected on the basis of prior integration events over regions obtained in the absence of selection. Therefore, we next examined the distribution of integration sites within sequences cloned at random from the cell genome. Oligonucleotides derived from these sequences were hybridized to TEF DNA to determine whether they represented single or multicopy TEF sequences (data not shown). All of the regions used in the analysis represented single copy sequences with the exception of one: region 18, which represented a multicopy sequence.

Primers derived from these regions were then used to detect integration events in the same pool of infected cell DNA as used in Figure 2 (Fig. 4). Figure 4A shows the result of analysis of region 15, a single-copy sequence. Lanes 5–21 and 30–34 show the distribution of 35 integration events within this region from a sampling of 26 reactions. The pattern observed was not obviously different from that seen with preselected regions, a distribution of integration sites within the region with some localized hotspots.

Figure 4B shows results of our analysis of region 18, present in ~50 copies in the genome, using DNA isolated after two independent infections. A large number of integration sites was observed, and the pattern of integrations resembled that seen in the in vitro reactions. The large number of integration events observed in a relatively small number of specific sites implies that many (or all) of the individual copies provide targets, and the pattern of use suggests that most of these are used in similar ways with minimal effects due to chromosomal location or structure. The repeated use of the same sites in experiments from independently infected cells (compare left and right panels) shows that the apparent site preference was not the result of reduplication of the same integration event by cell division.

Figure 5 is a graphic presentation summarizing all of the integration events detected within four preselected and two unselected regions. Integrations were detected throughout the sequence in each region with localized hyperreactive sites that were used at a frequency up to 280-fold greater than random. No obvious distinction could be made between sites selected on the basis of previous integration and those chosen at random. Table 1 gives a summary of results of the distribution analysis of all single-copy integration regions tested. We have detected integrations within all 12 of the regions examined thus far. This distribution suggests that there is little or no regional inhibition of ALV DNA integration and that all or most of the genome is accessible to the viral integration complex. When the integrations per base pair were calculated for each region (Table 1, column 6) we found a range of ~20-fold in frequency of usage. One of the target regions was used ~4-fold more frequently than random, and one ~5-fold less frequently. The majority of regions (8/11) were used between 0.6 and 1.6 times that expected for random. The slight bias in favor of targets selected for prior integration events (1.7 vs. 1.0) is obviously not statistically significant.

Target site preferences are conserved within multicopy regions

Finally, we were interested in assessing the effect of chromosomal context on target site selection. We did this by determining the integration site distribution within a preselected integration target when present at a relatively high copy number at different positions within the genome and comparing the distribution with that of the same DNA when used as a target in vitro. For these purposes, clones of QT6 cells containing varying numbers of copies of region 13 were isolated. Two clones

Figure 2. Distribution of integration events within selected single-copy integration regions. TEFs were infected with CS8 to generate a large pool of integration events. High molecular weight DNA was isolated and the m.o.i determined by Southern blot analysis. Ten micrograms of DNA (~5 × 10^6 cells) from the single pool was used in each PCR reaction. (A) The integration site distribution within region 13. Lanes 23–39 are the control reactions. Each reaction contained DNA from uninfected TEF to which either 10 copies (lanes 23, 24), 1 copy (lanes 25–34), or 0 copies (lanes 35–39) of a cloned proviral insertion within the region were added to each reaction. Lanes 7–18 show reactions that contained DNA isolated from CS8-infected cells. The bands selected for further amplification (Fig. 3) are indicated by dots. Lanes 5 and 6 show the products of in vitro integration reactions using pPre13 or pK7333 as targets. A map of each plasmid is shown at the bottom. The downward arrow indicates the control integration site. Additional controls are shown in lanes 40–42. Lanes 40 and 41 contain the zero time controls for the in vitro reactions. We believe the faint band at position 297 in the in vivo reaction containing pPre13 to be a result of contamination with the control plasmid used in the in vivo study. In all, we have run at least 50 negative control reactions with uninfected cell DNA and have never seen such a band. Lane 42 contains the primer-extension reaction mixture to which no genomic DNA was added. Lanes 1–4 and 19–22 contain sequencing reactions of region 13 by use of the region 13A-specific primer. (B) The integration site distribution within region 29. Lanes 15–21 show reactions containing DNA from uninfected TEF to which either 10 copies (lanes 15, 16) or 0 copies (lanes 17–21) of control plasmid were added. Lanes 1–10 show reactions that contained DNA isolated from CS8-infected cells. Lanes 11–14 contained sequencing reactions of region 29 with the internal region 29-specific primer. (C) The integration site distribution within region 39. Lanes 19–25 contained DNA from uninfected TEF cells to which either 10 copies (lanes 19, 20), or 0 copies (lanes 21–25) of a cloned proviral insertion within region 39 were added. Lanes 5–14 show reactions that contained DNA isolated from CS8-infected cells. Lanes 1–4 and 15–18 are sequencing reactions of region 39 using the internal region 39-specific primer. (D) The integration site distribution within region 51. Lanes 24–28 show the control reactions containing DNA from uninfected TEF cells to which either 10 copies (lanes 24, 25) or 0 copies (lanes 26–28) of control sequences of a cloned proviral insertion within the region were added. Lanes 5–19 show reactions containing DNA from CS8 infected cells. Lanes 1–4 and 20–23 are sequencing reactions of region 51 with the internal region 51 specific primer. (A–D) Arrows indicate the sites of integration used at significantly greater than random frequency, along with the estimated frequency relative to random.
were selected for further study. Clone 1 contained ~30 copies of region 13 in tandem array at three different integration sites. Clone 10 contained two or three copies of region 13 sequences in tandem array at a single integration site. Although QT6 cells contain sequences related to region 13, they are not related closely enough to be amplified with primers derived from the TEF sequence [data not shown].

To determine the use of the exogenously acquired sequences as targets, QT6 clones 1 and 10 were infected with CS8 at an m.o.i. identical to that used previously [Fig. 6]. No bands were detected in control reactions that contained DNA from uninfected or infected QT6 cells [lanes 28–32], or from uninfected QT6 clones [lanes 1–9]. In contrast, we detected integration events in DNA from the two clones containing integrated copies of region 13 after infection with CS8, with ~10 times as many in clone 1 as in clone 10 cells [lanes 14–18, 19–23]. Thus, the region 13 sequences introduced by transfection were used as targets in rough proportion to their copy number.

The large number of bands in the lanes containing DNA from infected clone 1 cells but not in uninfected or infected cells without the insert allowed us to perform an additional control for the possibility that illegitimate recombination during the PCR reaction might contribute some signal to our assay resembling that derived from authentic integration events. In such a case, we would expect to see bands when DNA from uninfected cells containing the targets was simply mixed, prior to PCR, with DNA from infected cells lacking such targets. To test this possibility, equal amounts of DNA from CS8-infected QT6 cells and uninfected clone 1 cells were mixed and used in the integration site distribution assay. No bands were detectable [Fig. 6A, lanes 33–37]. This result confirms the specificity of the assay for integration events.

The integration site distribution of QT6 clone 1 showed many more integrations than clone 10 and a strong localized preference for some sites over others. In some cases, sites that were used frequently when present at low copy number were also used frequently when present at high copy number. We also observed that few or no integrations were detected within some regions when they were present at low copy number, yet the same sequences were used frequently as targets when present at high copy number. Comparison of the patterns obtained with the two clones shows that although the frequency of targets was greatly increased, the number of sites used was not, giving a pattern similar to that seen with an endogenous repeated DNA target such as region 18 [Fig. 5C]. This result implies that simply increasing the copy number of the target does not lead to a corresponding increase in the number of different sites that can be used. Furthermore, dilution of DNA from infected clone 1 ~30-fold prior to the PCR reaction yielded a pattern resembling that of clone 10 [data not shown]. These results suggest a lack of effect of chromosomal position of the target sequences on target site selection.

Figure 6B shows the results of a similar analysis of integration site distribution within infected clone 1 and clone 10 DNA. In this case, the PCR reaction products were analyzed in the primer extension assay with two different primers that allowed us to map integrations as close as 60 nucleotides from the PCR primer or as far as 600 nucleotides away from the PCR primer. The lack of larger products suggested that the upper limit of the assay was ~600 nucleotides from the PCR primer, probably reflecting the maximum length fragment that could be efficiently amplified from a single molecule under our conditions.

The large number of integrations into the many copies of region 13 in clone 1 cells allowed us to compare directly the use of the same region as target in vivo and in vitro. Comparison of the in vivo integration site distribution in lanes 20–24 with the products of an in vitro reaction shown in lane 25 demonstrates that the patterns of integration site distribution shared similarities when identical sequences were supplied as target DNA. Inte-
Retrovirus integration specificity

Table 1. Summary of integrations detected within preselected and randomly selected regions

| Region | PCR reactions | Number of sites used | Total integrations | Integrations/base ×10⁻³ | Frequency relative to random |
|--------|---------------|----------------------|--------------------|-------------------------|-----------------------------|
| 13     | 62            | 65                   | 86                 | 2.8                     | 1.0                         |
| 29     | 22            | 59                   | 125                | 11.4                    | 3.8                         |
| 39     | 20            | 27                   | 31                 | 3.1                     | 1.0                         |
| 51     | 24            | 26                   | 32                 | 2.7                     | 0.8                         |
| 1      | 21            | 14                   | 16                 | 1.5                     | 0.6                         |
| 2      | 17            | 16                   | 17                 | 2.0                     | 0.6                         |
| 6      | 20            | 8                    | 17                 | 0.1                     | 0.2                         |
| 12     | 24            | 31                   | 41                 | 3.4                     | 1.2                         |
| 15     | 31            | 33                   | 41                 | 2.6                     | 0.8                         |
| 16     | 20            | 43                   | 50                 | 5.0                     | 1.6                         |
| 17     | 15            | 36                   | 46                 | 6.1                     | 2.0                         |
| 11     | 257           | 335                  | 460                | 3.7                     | 1.2 ± 0.9                   |

aCalculated by dividing the total integrations by the total number of bases screened within that region (the total number of PCR reactions × 500).

bCalculated by dividing the total number of integrations by the total number of PCR reactions × 1.5 (the number of integrations we would expect in a 500-bp region if integration was random.) The average of the frequency relative to random for the preselected regions is 1.7 ± 1.2. The average of the frequency relative to random for the unselected regions is 1.0 ± 0.6.

Comparison of the distribution of integrations within preselected and randomly selected single-copy target sites

Our study has allowed us to look at the distribution of integration targets within the cell genome without requiring prior selection by molecular cloning. On the basis of our sampling of the avian genome, our results suggest that target sites for the majority of ALV integration events are distributed with surprising uniformity, and we could see no evidence for regional inhibition of integration.

The average frequency of integration into all target sequences examined was almost exactly that expected from a random distribution. Of the 12 regions examined in this study, 9 were used at a frequency ~0.6–1.6 times that expected on a random basis. We do not consider these numbers to be significantly different from random,

Discussion

The studies described here were designed to provide a global look at retroviral integration into single-copy cell DNA in vivo. For this purpose, we developed a system that achieves the levels of sensitivity and specificity required to detect single integration events within a population of cells in any region of the genome with single nucleotide resolution. The system should provide us with a valuable tool for examining the effects of chromatin structure and cellular functions on retroviral integration target site selection in vivo.

The assay is based on our previous experience with integration in vitro (Kitamura et al. 1992) and uses primers derived from viral LTR and cellular sequences to amplify single-copy integration events in a single orientation to one side of the cellular primer. Amplified products are used for extension of a 32P-labeled nested primer derived from cellular sequences. The extension products are then analyzed in sequencing gels to assign precisely the position of integration sites within the region. Reconstitution experiments demonstrated the sensitivity and selectivity of the assay for single integration events within a background of genomic DNA. First, when a single copy of plasmid DNA containing cloned provirus and flanking DNA sequences was added to reactions with TEF DNA, we saw a clear band of the correct size and at the frequency expected for a Poisson distribution. Second, repeated negative controls demonstrated the selectivity of the primer pair for integration events and the absence of contaminating sequences that would result in false positives. Third, a control experiment allowed us to exclude the possibility of false positives arising by illegitimate recombination during PCR given the 10⁷-fold (or greater) excess of viral and cellular sequences related to the PCR primers. When equal amounts of DNA from infected QT6 cells and cells that were not infected, but contained ~30 copies of the target region, were mixed and used in our integration site distribution assay, no bands were detectable. The specificity was ultimately confirmed by reamplification and sequencing of many of the bands detected. Of the 23 bands analyzed from three target regions, all were found to be correct integration events in that LTR sequences lacking the 3' terminal 2 bases were joined to the sequence of the target region predicted from the size of the fragment.

Comparison of the distribution of integrations within preselected and randomly selected single-copy target sites

Our study has allowed us to look at the distribution of integration targets within the cell genome without requiring prior selection by molecular cloning. On the basis of our sampling of the avian genome, our results suggest that target sites for the majority of ALV integration events are distributed with surprising uniformity, and we could see no evidence for regional inhibition of integration.

The average frequency of integration into all target sequences examined was almost exactly that expected from a random distribution. Of the 12 regions examined in this study, 9 were used at a frequency ~0.6–1.6 times that expected on a random basis. We do not consider these numbers to be significantly different from random,
Figure 4. Integration within nonselected integration sites. Integration site distribution was examined within regions chosen at random to represent single- or multicopy sequences within the turkey genome. (A) The integration site distribution within the single-copy region 15. Lanes 22–25 contained the control reactions with DNA from uninfected TEF. Lanes 5–21 and 30–34 show reactions that contained DNA isolated from CS8-infected cells after independent infections. Lanes 1–4 and 26–29 show sequencing reactions of region 15 with the region 15 internal primer. (B) The integration site distribution within multicopy region 18. Lanes 1–5 show reactions containing DNA from uninfected TEF cells. Lanes 10–19 and lanes 24–33 are reactions that contained DNA isolated from two independent infections with CS8. Lanes 6–9 and 20–23 show sequencing reactions of region 18 with the region 18-specific internal primer.

given the relatively small sampling size and uncertainties in some of the other estimates used in the calculation. One of the regions, apparently present in single copy, attracted about fourfold more integration events, and one about fivefold less, suggesting a small preference for specific regions. As a further indication of the global nature of the genomic integration target, we observed only a slight, and not significant, bias toward regions preselected on the basis of a prior integration event.

Although we have sampled only a small fraction of the cell genome (~6000 of 2 x 10^9 bases) the presence of integration targets in 12 of 12 regions sampled leads us to conclude that the majority of the cellular genome is available for integration and the total number of potential integration sites must be very large. Within regions, the use of specific sites as integration targets was decidedly nonrandom because some target sites in single-copy DNA were used up to 280 times the frequency expected for a random distribution. Presumably, many sites are used at a much lower than random frequency as well, but these would not be seen. The frequent use of identical sites in experiments from independently infected cells excludes the formal possibility of reduplication of the same integration event by replication of the infected cell DNA.

These calculations presume that there is no significant bias introduced into the estimates by the PCR method. Because frequencies of use of specific sites were estimated by counting the frequency of appearance of bands of specific sizes in different reactions, it is not possible to overestimate the number of events. However, a systematic bias could arise from certain integration events failing to amplify as efficiently as others and being lost to analysis. We cannot absolutely rule out such bias for all possible integration sites. However, our ability to efficiently amplify the standards in all experiments (Fig. 2) and the consistent intensity of the bands of all sizes within a single experiment lead us to believe that we are
detecting all events. In a previous paper (Kitamura et al. 1992), we presented evidence that efficiency of amplification in PCR reactions like these is relatively independent of the size or composition of the sequence between the primers. It should be kept in mind that these calculations are subject to considerable statistical uncertainty because of the small number of events scored. We expect that frequent events at relatively hot spots are being missed for this reason.

Our present results are not entirely consistent with the previous analysis in our laboratory that led to the conclusion that a fraction of integrations is into a small number of sites that are used with a very strong preference (Shih et al. 1988). Two of the regions examined in the present study [numbers 13 and 51] contained sites identified previously as preferred. We expected to detect new integration events at these sites, because these would have led to amplified products identical in size to the control DNA molecules used [Fig. 2A,C]. Despite the strong prediction that we would do so, we saw no such products. In the previous study, integrations were detected within these regions by screening two different integration site libraries constructed by isolating DNA from independent infections of TEF with CS8 virus. The second integration site library was constructed in such a way as to rule out the possibility that multiple use of the same sites was attributable to the presence of contaminating phage. Thus, although we cannot absolutely exclude some sort of experimental error, we cannot identify any obvious way in which such an error could have occurred.

Remaining possibilities to account for the discrepancy revealed by the two systems are [1] that the sequences originally designated as preferred share some feature(s)
that allowed them to be selectively packaged by bacteriophage λ, or (2) that some unknown variation in the cell culture conditions or TEF cells used could account for the contradictory results. The former explanation is unsatisfactory because we would then expect to have seen preferential packaging of other integrations within these regions as well. Whatever the reason for the discrepancy, because we would have detected integration into the identical site as in the prior study (i.e., the 270-base control band, Fig. 2A), we therefore consider the present result to provide the more accurate picture of integration site distribution.

Integration into multicopy DNA targets

Because we could only collect small numbers of integration events by use of single-copy DNA targets, we also examined integration into regions of DNA repeated multiple times. The first of these was the randomly selected region 18, which attracted a number of integration events in rough proportion to its copy number (Fig. 4B). These integrations also demonstrated a clear pattern of preferred and avoided sites that was similar in samples from independently infected cells. This pattern of use implies that most or all copies of this region present similar targets to the integration machinery in spite of probable differences in structure and context.

A similar conclusion can be drawn from the distribution of targets when an initially single-copy region was introduced into cells at high copy number (Fig. 6). This experiment was done as a test of the feasibility of using deliberately introduced high copy sequences to study the effects of chromatin structure and activity on integration. The preliminary results suggested that usage of specific target sites in vivo may resemble that seen in vitro and that it might be possible to use this approach to detect the presence of tightly bound proteins or other structural modifications. Experiments to test these ideas directly are in progress.

Figure 6. (See facing page for B and legend.)
The basis of target site selection in vivo

The conclusion that we have drawn from our results—that there is a very large number of potential integration sites for ALV widely distributed in the cell genome—is rather different from those of prior studies suggesting a clustering of sites, particularly at features associated with transcriptional activity. We have not yet directly

![Diagram of PCR analysis](Image)

**Figure 6.** Integration into artificially introduced targets. PCR analysis of the integration site distribution within region 13 in QT6 clones infected with CS8. (A) Lanes 1–9 show control reactions that contained DNA isolated from uninfected QT6 clones 1 and 10, respectively. Lanes 28–32 show control reactions that contained DNA isolated from either uninfected or CS8 infected QT6 cells. The reaction in lane 28 contained 10 copies of pGem13. A band of the expected size (297 nucleotides) was amplified. Lanes 33–37 show additional control reactions that each contained 5 μg of DNA isolated from uninfected clone 1 cells mixed with 5 μg of DNA isolated from CS8-infected QT6 cells. Lanes 14–18 show the integration site distribution within the multicopy region 13 in clone 1. Lanes 19–23 show the integration site distribution within the low-copy region 13 in clone 10. Lanes 10–13 and 24–27 contained sequencing reactions of region 13 with the internal region 13-specific primer. Arrow at top indicates copies of standard. (B) A comparison of integration site distribution in CS8-infected clones 1 and 10 with two different region 13-specific primers. Reactions in lanes 5–14 contained the nested primer (primer A) located just upstream of the PCR primer and allowed precise mapping of integrations closer to it. The rest of the figure shows reactions using the nested primer (primer B) located 130 nucleotides more distant from the region-specific primer [lanes 20–24, clone 1; lanes 26-30, clone 10]. Lane 25 shows the reaction products when pKT333 [Fig. 3A] was supplied as target in an in vitro reaction. Lanes 31–39 show control reactions that contained DNA isolated from uninfected QT6 clones 1 and 10. Lanes 44–53 show control reactions that contained DNA isolated from either uninfected or CS8-infected QT6 cells. The reaction in lane 44 contained 10 copies of control plasmid. A band of the correct size (164 nucleotides) was amplified. Lanes 1–4, 16–19, and 40–43 contained sequencing reactions of region 13 with the region 13-specific internal primer A. Lane 15 contained the zero time control for the in vitro reaction. The location of the primers and control integration site is shown at the bottom. The arrows joined by a dashed line show identical bands detected with the two different primers. Arrow at top indicates copies of standards.
assessed the effect of such features on integration specificity. However, the distribution of events throughout the regions that we have analyzed implies that the majority of integrations cannot be concentrated into any specific type of chromosomal region. We believe that the difference between the present results and those of previous studies is that our assay allowed us to survey a large number of unselected integration events by direct analysis of infected cell DNA very soon after infection. In contrast, all of the systems described previously could have introduced a selection bias; thus, they may not reflect the true range of target-site selection. For example, many studies suggesting that integrations may cluster within nucleosome-free regions associated with DNase I hypersensitive sites [Hayward et al. 1981; Cullen et al. 1984; Shih et al. 1984; Goodenow and Hayward 1987; Robinson and Gagnon 1987] used insertions selected on the basis of an effect on c-myc expression and may reflect the selective growth advantage afforded by ALV integration at specific sites within the c-myc locus. Even studies that analyzed unselected proviruses within a few hundred base pairs of DNase hypersensitive sites in tumors and cell lines [Vijaya et al. 1986; Rohdewohld et al. 1987] may suffer from the same sort of problem, because many proviral insertions in tumors and in cell lines derived from them may be selected for relatively subtle effects on cell growth [Tsichlis and Lazo 1991]. Additionally, these conclusions were based on assumptions about the distribution of DNase hypersensitive sites within cell DNA, an issue not addressed in the studies.

Finally, selection schemes based on disruption of the target genes by integration have given much lower estimates of integration frequency into certain genes than expected by the investigators on the basis of a random distribution of sites [Frankel et al. 1985; King et al. 1985]. These calculations, however, required making a number of untested assumptions about the efficiency of inactivation by integration and the efficiency of the selection scheme used. By use of a gene fusion trap approach, Reddy et al. [1985] found that they could detect integration into active transcription units at the frequency expected from a random distribution of sites.

The strongest case for targeting of specific sites for integration among retrovirus-like elements is in the yeast retrotransposon Ty3, which inserts almost exclusively into positions within 1–4 bp of the transcription initiation site for transcription of tRNA genes [Chalker and Sandmeyer 1990, 1992]. This specificity apparently reflects a highly selective mechanism for target site selection, which involves direct interaction of the integration machinery with the polymerase III transcription apparatus [Chalker and Sandmeyer 1993]. To date, we have been unable to obtain evidence for a similar effect in retroviruses, but the system that we describe here will allow us to examine this issue directly.

Unlike Ty3 insertions, Ty1 insertions are not position specific with respect to tRNA genes. However, Ty1 insertions have been found preferentially in regions associated with tRNA genes [Natsoulis et al. 1989; Sandmeyer et al. 1990, Ji et al. 1993]. This result is quite different from the position-specific insertion seen with Ty3 and most likely reflects an entirely different mechanism of target site selection. However, although these results appear to support the hypothesis that Ty1 integration favors tRNA genes, they could also reflect an unwitting bias introduced by the numerous cycles of cell growth and genetic manipulations necessary to capture insertions. The assay that we describe here could be usefully applied to this system as well to sort out effects of integration specificity per se from effects attributable to selection during subsequent growth.

In contrast to the Ty elements, the results of our study suggest that accessibility to different regions of the genome for ALV integration is relatively unrestricted. We consider it likely that the same sort of distribution will be found for other retroviruses, but this remains to be tested directly. Within the limited group of regions that we surveyed, both preselected and randomly selected, we detected no strong avoidance or bias toward any sequences. The localized specific targeting that we detected most likely reflects a local DNA structure that promotes integration in some way. Our results paint a picture of integration target site selection that is very different from prevailing models, which argue that transcriptionally active regions and regions that are associated with DNase I hypersensitive sites are preferred by the integration machinery. Ongoing studies in our laboratory that address the effects of transcription and DNase I hypersensitive sites on target site selection directly will allow us to assess the role of these features further in retroviral integration.

Materials and methods

Cells and viruses

CS8, a replication-competent derivative of ALV has been described [Shih et al. 1988]. TEF and QT6 [Moscovici et al. 1977] cells were infected with CS8 virus as described previously [Shih et al. 1988], with the exception that the cells were grown for 4 days prior to isolation of DNA. High molecular weight DNA was isolated by standard procedures [Maniatis et al. 1989].

Plasmids

pGem-13, 29, 39 and 51 have been described previously [Shih et al. 1988]. They each contain a ScaI fragment of a proviral integration within regions 13, 29, 39, and 51 cloned into pGem-1. pPre-13 contains the 3.0-kb preintegration ScaI fragment of region 13 cloned into pGem-1 [Shih et al. 1988]. pKT333 contains the HindIII fragment of pPre-13 (nucleotides 464–920) cloned into pUC19. pKT334 contains the HindIII–RsaI fragment of pPre-13 (nucleotides 464–680) cloned into pUC19. pME18S/HygB carries a multicloning site and the gene conferring ampicillin and hygromycin resistance (courtesy of Kazuo Maruyama). pHyg/Pre13 was derived by replacing the 2.7-kb stuffer region of pME18S/HygB with a 4.4-kb BamHI–FspI fragment containing the 3.0-kb ScaI fragment of region 13 from pPre 13.

Determination of multiplicity of infection with CS8 virus

Serial dilutions of undigested genomic DNA isolated from CS8-
infected TEF cells were analyzed by electrophoresis in an agarose gel next to serial dilutions of DNA isolated from chicken embryo fibroblast (CEF) cells, which carry a single copy of an endogenous virus, ev-1 (Astrin 1978). The unblots were probed with 32P-labeled Rous sarcoma virus (RSV) env sequences. These blots were stripped and reprobed with 32P-labeled chicken fibronectin sequences, which allowed us to normalize the amounts of integrated virus DNA loaded in each lane. By comparing the intensity of bands representing integrated CS8 chicken fibronectin sequences, which allowed us to normalize the amounts of integrated virus DNA loaded in each lane. By comparing the intensity of bands representing integrated CS8 chicken fibronectin sequences, which allowed us to normalize the amounts of integrated virus DNA loaded in each lane. By comparing the intensity of bands representing integrated CS8 chicken fibronectin sequences, which allowed us to normalize the amounts of integrated virus DNA loaded in each lane.

These blots were stripped and reprobed with 32P-labeled Rous sarcoma virus (RSV) endogenous virus, embryo fibroblast (CEF) cells, which carry a single copy of an endogenous virus, embryo fibroblast (CEF) cells, which carry a single copy of an endogenous virus. The unblots were stripped and reprobed with 32P-labeled Rous sarcoma virus (RSV) endogenous virus, embryo fibroblast (CEF) cells, which carry a single copy of an endogenous virus. The unblots were stripped and reprobed with 32P-labeled Rous sarcoma virus (RSV) endogenous virus, embryo fibroblast (CEF) cells, which carry a single copy of an endogenous virus. The unblots were stripped and reprobed with 32P-labeled Rous sarcoma virus (RSV) endogenous virus, embryo fibroblast (CEF) cells, which carry a single copy of an endogenous virus.

**Table 2. Oligonucleotides used in PCR and primer extension**

| Region | PCR primer | Internal primer |
|--------|------------|----------------|
| U5     | CGGAATTTCAAGATGGAACACCTGAAA | CGGGATCCAGTCATTACTTAAAGCTGTTG |
| 13     | CGGGATCCGAATTCGCTTCAGAGCTACCTA | CGGGATCCAGTCAACAGCCTGTTG |
| 51     | CGGGATCCCAATACGGGATACGAAATCACCAC | CGGGATCCAGTCAACAGCCTGTTG |
| 29     | CGGGATCTGCTGCTGATGCTTTTCA | TCTTGGTGGTGGTTTTTTTTTTC |
| 39     | CGGGATCTGCTGCTGATGCTTTTCA | TCTTGGTGGTGGTTTTTTTTTTC |
| 1      | CGGGATCTGCTGCTGATGCTTTTCA | TCTTGGTGGTGGTTTTTTTTTTC |
| 2      | CGGGATCTGCTGCTGATGCTTTTCA | TCTTGGTGGTGGTTTTTTTTTTC |
| 4      | CGGGATCTGCTGCTGATGCTTTTCA | TCTTGGTGGTGGTTTTTTTTTTC |
| 6      | CGGGATCTGCTGCTGATGCTTTTCA | TCTTGGTGGTGGTTTTTTTTTTC |
| 8      | CGGGATCTGCTGCTGATGCTTTTCA | TCTTGGTGGTGGTTTTTTTTTTC |
| 12     | CGGGATCTGCTGCTGATGCTTTTCA | TCTTGGTGGTGGTTTTTTTTTTC |
| 15     | CGGGATCTGCTGCTGATGCTTTTCA | TCTTGGTGGTGGTTTTTTTTTTC |
| 16     | CGGGATCTGCTGCTGATGCTTTTCA | TCTTGGTGGTGGTTTTTTTTTTC |
| 17     | CGGGATCTGCTGCTGATGCTTTTCA | TCTTGGTGGTGGTTTTTTTTTTC |
| 18     | CGGGATCTGCTGCTGATGCTTTTCA | TCTTGGTGGTGGTTTTTTTTTTC |

The forward or virus-specific primer [U5] used in the analysis of each site has an EcoRI restriction site at its 5' end. The reverse or region-specific primers have a BamHI restriction site at their 5' ends.

**Isolation of HygB' clones containing multicopy region I3 sequences**

QT6 cells [2×10⁶ cells per 6 cm plate] were transfected by the lipofection method (Gibco BRL) with 5 μg of pHyg/Pre13. Five micrograms of pHyg/Pre13 and 50 μg of lipofectin were mixed in 2 ml of serum-free Dulbecco's minimal essential medium (DMEM) and incubated with QT6 cells for 6 hr at 37°C. HygB' clones were isolated after selection in hygromycin B [200 μg/ml], expanded, and high molecular weight DNA was isolated by standard procedures (Maniatis et al. 1989). Southern blot analysis after digestion with BamHI, which cuts once within pKT335, allowed us to determine the number of integrations per clone and the approximate copy number of region I3 sequences. Two clones were selected for further analysis. Clone 1 contained three integration sites with multiple copies in tandem array at each site for an estimated total of 30 copies of region I3 sequences per cell. Clone 10 contained two or three tandem copies at a single integration site.

**Oligonucleotides**

The oligonucleotides used in this study were selected for PCR and primer extension with the PRIMER version 0.5 program (Lincoln et al. 1991). This program automatically selects primer pairs compatible with specific reaction component concentrations and annealing temperatures. The annealing temperatures are computed on the basis of analysis of the base composition and the energy of base stacking of the target sequences. Primers were synthesized with an Applied Biosystems 380B DNA synthesizer. They are designated by the integration target region or viral sequences they are derived from and are listed in Table 2. The virus-specific primer represented bases 61-82 from the U5 region of Prague (Pr)–RSV (Weiss et al. 1985) with the addition of an EcoRI recognition site at the 5' end.

**PCR assay for detecting integrations in vivo**

The PCR reactions for analysis of integration site distribution within a given region were prepared as follows. Genomic DNA [1 μg/μl] isolated from CS8-infected TEF cells was boiled for 5 min and then placed in an 80°C heating block. Ten microliters of the forward or virus-specific primer [U5] used in the analysis of each site has an EcoRI restriction site at its 5' end. The reverse or region-specific primers have a BamHI restriction site at their 5' ends.
of the DNA was added to a PCR mixture (overlaid with 25 μl of mineral oil) that had been prewarmed for 5 min in an 80°C heating block to give final reaction concentrations of 10 mM Tris (pH 8.3), 5 mM MgCl₂, 50 mM KCl, 0.01% gelatin, 3.2 mM dNTPs, 0.6 μM each primer (virus and region specific), and 2.5 units of Taq polymerase (AmpliTaq; Cetus-Perkin Elmer) in a final reaction volume of 35 μl. These reactions were then transferred to a thermal cycling machine that was prewarmed to 80°C. Control reactions, which contained 10 μg of DNA isolated from uninfected TEF and varying numbers of copies of a clone of a proviral insertion within the site to be analyzed, were prepared in a similar manner. The reaction mixtures were then transferred directly to a thermal cycling machine that had been preheated to 80°C and, after 5 min at 94°C, were amplified for 29 cycles by thermal cycling at 94°C for 1 min and 68°C for 2 min. The final cycle was 94°C for 1 min, 68°C for 5 min. For PCR analysis of the QT6 clones 1 and 10, the same procedure was followed. In this case, the genomic DNA added to each reaction was isolated from either CS8-infected or uninfected QT6, clone 1, or clone 10 cells.

PCR assay for detecting integrations in vitro

In vitro integration reactions were performed with cytoplasmic extracts of QT6 cells infected 16 hr previously with CS8 virus as described (Kitamura et al. 1992). Five microliters of each reaction was amplified by use of the PCR assay described by Kitamura et al. (1992) with the virus- and region 13-specific primers. Briefly, 5 μl of the in vitro integration reaction was added to a PCR reaction mixture (overlaid with 100 μl of mineral oil) to give final reaction concentrations of 25 mM Tris (pH 9.5), 2 mM MgCl₂, 50 mM KCl, 1 mM 2-mercaptoethanol, 200 μM dNTPs, 1.0 μM each primer (virus and region 13 specific) and 2.5 units of Taq polymerase (AmpliTaq; Cetus-Perkin Elmer), and 1 unit of PerfectMatch polymerase enhancer (Stratagene) in a final reaction volume of 100 μl. The reactions were subjected to 30 rounds of thermal cycling (94°C for 1 min, 55°C for 2 min, 74°C for 3 min).

Primer extension assay for mapping integration site distribution

Ten microliters of each PCR reaction (in vivo and in vitro) was digested with EcoRI (NEB) in a reaction volume of 50 μl for at least 12 hr at 37°C. This step was included to cleave near the end of the virus-specific primer and thus ensure that all of the PCR-amplified products at a given site were of the same length. Reactions were then extracted with phenol/chloroform, and the nucleic acid was precipitated with ethanol. The precipitated nucleic acids were heat denatured and annealed with ~0.1 pmole of the internally nested 5'−32P-labeled primer (~107 cpm/pmol) in 10 μl of 40 mM Tris (pH 7.5), 20 mM MgCl₂, and 50 mM NaCl. Extension was carried out at 42°C for 30 min. Samples were analyzed by electrophoresis in a 6% polyacrylamide denaturing gel under standard conditions. Sequencing ladders derived from each region were run in parallel to provide size standards.

Sequencing PCR-amplified products

Bands were also eluted by boiling the excised gel slice in 100 μl of water for 30 min. The sample was then cooled to room temperature and 1 μl of glycerogen (20 mg/ml, Boehringer Mannheim) or 25 μg of rRNA was added as carrier, and the nucleic acid was precipitated with sodium acetate and ethanol overnight at −20°C. The eluted fragment was resuspended in 5 μl of water and reamplified by PCR by use of conditions described previously. The reamplified PCR product was separated on an 8% polyacrylamide gel, eluted by the crush and soak method (Maniatis et al. 1989), precipitated, and used in an asymmetric PCR reaction to produce single-stranded DNA for direct sequencing.

Gel-purified PCR products were used in another round of PCR under conditions identical to those described previously, with the exception that the concentration of one of the primers was decreased 50-fold. These conditions favor the amplification of a single strand of DNA template, in this case, one that is complementary to the primer used for sequencing. The single-stranded PCR product was then applied to a Centricon 30 column (Perkin-Elmer Cetus Corp.) or a QIAQuick-spin column (Quiagen Inc.), which are designed for the purification of single- or double-stranded PCR products away from primers, nucleotides, and polymerases. The eluate was then dried down in a speed-vac and used directly in sequencing reactions.

Acknowledgments

We thank Kazuo Maruyama for pME18SHygB and Swee-Kee Wong for expert technical contributions. This work was supported by a grant from the National Cancer Institute to J.M.C. Y.K. was a fellow of the Leukemia Society of America. J.M.C. is a research professor of the American Cancer Society.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Astrin, S. 1978. Endogenous viral genes of white leghorn chicken: A common site of residence as well as sites associated with specific phenotypic expression of viral gene expression. Proc. Natl. Acad. Sci. 75: 5941–5945.

Brodie, G.M., S.B. Sandmeyer, and M.V. Olson. 1983. Consistent association between sigma elements and tRNA genes in yeast. Proc. Natl. Acad. Sci. 80: 3292–3296.

Brown, P.O., B. Bowerman, H.E. Varmus, and J.M. Bishop. 1987. Correct integration of retroviral DNA in vitro. Cell 49: 347–356.

—. 1989. Retroviral integration: Structure of the initial covalent product and its precursor, and a role for the IN protein. Proc. Natl. Acad. Sci. 86: 2525–2529.

Bushman, F.D., T. Fujiwara, and R. Craigie. 1990. Retroviral DNA integration directed by HIV integration protein in vitro. Science 249: 1555–1558.

Chalker, D.L. and S.B. Sandmeyer. 1990. Transfer RNA genes are genomic targets for de novo transposition of Ty3. Genetics 126: 837–850.

—. 1992. Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes & Dev. 6: 117–128.

—. 1993. Sites of RNA polymerase III transcription initiation and Ty3 integration at the U6 gene are positioned by the TATA box. Proc. Natl. Acad. Sci. 90: 4927–4931.

Craigie, R. 1992. Hotspots and warm spots: Integration specificity of retroelements. Trends Genet. 8: 187–190.

Craigie, R., T. Fujiwara, and F. Bushman. 1990. The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. Cell 62: 829–837.

Cullen, B.R., P.T. Lomedico, and G. Ju. 1984. Transcriptional interference in avian retroviruses: Implications for the promoter insertion model of leukemogenesis. Nature 307: 241–244.
Ellison, V., H. Abrams, T.Y. Rose, J. Lifson, and P. Brown. 1990. Human immunodeficiency virus integration in a cell-free system. J. Virol. 64: 2711–2715.

Engelman, A., K. Mizuuchi, and R. Craigie. 1991. HIV-1 DNA integration: Mechanism of viral DNA cleavage and DNA strand transfer. Cell 67: 1211–1222.

Frankel, W., T.A. Potter, N. Rosenberg, I. Lenz, and T.V. Rajan. 1985. Retroviral insertional mutagenesis of a target allele in a heterozygous murine ceml line. Proc. Natl. Acad. Sci. 82: 6600–6604.

Fujiwara, T. and R. Craigie. 1989. Integration of mini-retroviral DNA: A cell-free reaction for biochemical analysis of retroviral integration. Proc. Natl. Acad. Sci. 86: 3065–3069.

Fujiwara, T. and K. Mizuuchi. 1988. Retroviral DNA integration: Structure of an integration intermediate. Cell 54: 497–504.

Goodenow, M.M. and W.S. Hayward. 1987. 5′ Long terminal repeats of myc-associated proviruses appear structurally intact but are functionally impaired in tumors induced by avian leukosis viruses. J. Virol. 61: 2489–2498.

Hayward, W.S., B.G. Neel, and S.M. Astrin. 1981. Activation of a cellular one gene by promoter insertion in ALV-induced lymphomas. Nature 290: 475–480.

Ji, H., D.P. Moore, M.A. Blomberg, L.T. Braiterman, D.F. Voytas, G. Natsoulis, and J.D. Boeke. 1993. Hotspots for unselected Ty1 transposition events on yeast chromosome III are near rDNA genes and LTR sequences. Cell 73: 1007–1018.

Katz, R.A., G. Merkel, J. Kulkosky, J. Leis, and A.M. Skalka. 1990. The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro. Cell 63: 87–95.

Katzman, M., R.A. Katz, A.M. Skalka, and J. Leis. 1989. The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. J. Virol. 63: 5319–5327.

King, W., M.D. Patel, L.I. Lobel, S.P. Goff, and M.C. Nguyen-Huu. 1985. Insertion mutagenesis of embryonal carcinoma cells by retroviruses. Science 228: 554–558.

Kitamura, Y., Y.M.H. Lee, and J.M. Coffin. 1992. Nonrandom integration of retroviral DNA in vitro: Effect of CpG methylation. Proc. Natl. Acad. Sci. 89: 5532–5536.

Lee, Y.M.H. and J.M. Coffin. 1990. Efficient autocleavage of avian retroviral DNA in vivo. J. Virol. 64: 5958–5965.

———. 1991. Relationship of avian retrovirus DNA synthesis to integration in vitro. Mol. Cell Biol. 11: 1419–1430.

Lincoln, S.E., M.J. Daly, and E.S. Lander. 1991. Primer: A computer program for automatically selecting PCR primers. Massachusetts Institute of Technology, Cambridge, MA.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1989. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Mooslechner, K., U. Karls, and K. Harbers. 1990. Retroviral integration sites in transgenic Mov mice frequently map in the vicinity of transcribed DNA regions. J. Virol. 64: 3056–3058.

Moscovici, C., M.G. Moscovici, H. Jimenez, M.M.C. Lai, M.J. Hayman, and P.K. Vogt. 1977. Continuous tissue culture cell lines derived from chemically induced tumors of Japanese quail. Cell 11: 95–103.

Natsoulis, G., W. Thomas, M.C. Rognmann, F. Winston, and J.D. Boeke. 1989. Ty1 transposition in Saccharomyces cerevisiae is non random. Genetics 123: 269–279.

Pryciak, P.M. and H.E. Varmus. 1992. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. Cell 69: 769–780.

Pryciak, P.M., H.-P. Mueller, and H.E. Varmus. 1992a. Simian virus 40 minichromosomes as targets for retroviral DNA integration in vivo. Proc. Natl. Acad. Sci. 89: 9237–9241.

Pryciak, P.M., A. Sil, and H.E. Varmus. 1992b. Retroviral integration into minichromosomes in vitro. EMBO J. 11: 291–303.

Reddy, S., J.V. DeGregori, H.V. Melchner, and H.E. Ruley. 1991. Retrovirus promoter-trap vector to induce lacZ gene fusions in mammalian cells. J. Virol. 65: 1507–1515.

Robinson, H.L. and G.C. Gagnon. 1986. Patterns of proviral insertion in avian leukaosis virus-induced lymphomas. J. Virol. 57: 28–36.

Rohdewohld, H., H. Weiher, W. Reik, R. Jaenisch, and M. Breindl. 1987. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. J. Virol. 61: 336–343.

Roth, M.J., P.L. Schwartzberg, and S.P. Goff. 1989. Structure of the termini of DNA intermediates in the integration of retroviral DNA: Dependence on IN function and terminal DNA sequence. Cell 58: 47–54.

Sandmeyer, S.B., L.J. Hansen, and D.L. Chalker. 1990. Integration specificity of retrotransposons and retroviruses. Annu. Rev. Genet. 24: 491–518.

Scherdin, U., K. Rhodes, and M. Breindl. 1990. Transcriptionally active genome regions are preferred targets for retrovirus integration. J. Virol. 64: 907–912.

Shih, C.K., M. Linial, M.M. Goodenow, and W.S. Hayward. 1984. Nucleotide sequence 5′ of the chicken c-myc coding region: Localization of a noncoding exon that is absent from myc transcripts in most avian leukosis virus-induced lymphomas. Proc. Natl. Acad. Sci. 81: 4697–4701.

Shih, C.-C., J.P. Stoye, and J.M. Coffin. 1988. Highly preferred targets for retrovirus integration. Cell 53: 531–537.

Stoye, J.P., W.N. Frankel, and J.M. Coffin. 1991. DNA hybridization in dried gels with fragmented probes: An improvement over blotting techniques. Technique 3: 123–128.

Tsichlis, P.N. and P.A. Lazo. 1991. Virus-host interactions and the pathogenesis of murine and human oncogenic retroviruses. Curr. Top. Microbiol. Immunol. 171: 95–179.

Varmus, H.E. and P. Brown. 1989. Retroviruses. In Mobile DNA (ed. D.E. Berg and M.M. Howe), pp. 53–108. American Society for Microbiology, Washington, D.C.

Varmus, H.E. and R. Swanstrom. 1984. Replication of retroviruses. In RNA tumor viruses, 1st ed. [ed. R. Weiss, N. Teich, H. Varmus, and J. Coffin], pp. 369–512. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Vijaya, S., D.L. Steffen, and H.L. Robinson. 1986. Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. J. Virol. 60: 683–692.

Wallace, R.B. and C.G. Miyanda. 1987. Oligonucleotide probes for the screening of recombinant DNA libraries. Methods Enzymol. 152: 440–442.

Weiss, R., N. Teich, H. Varmus, and J. Coffin. 1985. RNA tumor viruses. Supplement and Appendixes, Part 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
Distribution of targets for avian retrovirus DNA integration in vivo.

E S Withers-Ward, Y Kitamura, J P Barnes, et al.

Genes Dev. 1994, 8:
Access the most recent version at doi:10.1101/gad.8.12.1473

References
This article cites 43 articles, 25 of which can be accessed free at:
http://genesdev.cshlp.org/content/8/12/1473.full.html#ref-list-1

License

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.