JOURNAL OF VIROLOGY, Feb. 1990, p. 907–912
0022-538X/90/020907-06$02.00/0

Vol. 64, No. 2

Transcriptionally Active Genome Regions Are Preferred Targets for Retrovirus Integration

ULRICH SCHERDIN,† KATHERINE RHODES, AND MICHAEL BREINDL*

Department of Biology and Molecular Biology Institute, San Diego State University, San Diego, California 92182-0057

Received 8 May 1989/Accepted 2 October 1989

We have analyzed the transcriptional activity of cellular target sequences for Moloney murine leukemia virus integration in mouse fibroblasts. At least five of the nine random, unselected integration target sequences studied showed direct evidence for transcriptional activity by hybridization to nuclear run-on transcripts prepared from uninfected cells. At least four of the sequences contained multiple recognition sites for several restriction enzymes that cut preferentially in CpG-rich islands, indicating integration into 5′ or 3′ ends or flanking regions of genes. Assuming that only a minor fraction (<20%) of the genome is transcribed in mammalian cells, we calculated the probability that this association of retroviral integration sites with transcribed sequences is due to chance to be very low (1.6×10^{-2}). Thus, our results strongly suggest that transcriptionally active genome regions are preferred targets for retrovirus integration.

An essential step in the replication of retroviruses is the integration of a DNA copy of the viral RNA genome (the provirus) into the DNA of the infected host cell. Although viral DNA synthesis by reverse transcriptase and the structure of proviral DNA have been studied in detail (5, 10, 27, 42, 45, 47), no clear rules have emerged regarding the sites of integration of retroviruses into the cellular genome. Many studies using restriction and DNA sequence analysis have shown that provirus integration can occur at multiple sites without apparent sequence specificity (5, 10, 27, 42, 45, 47), and it was generally assumed that integration occurs essentially at random. On the other hand, retroviruses and other insertion elements show a remarkable preference for integration into 5′ noncoding or intron sequences of cellular genes (1, 6, 7, 9, 12, 14–19, 22, 26, 30, 33, 35, 37, 40, 44, 48), and it has been proposed that factors other than nucleotide sequence may influence the integration process (4, 31, 43, 46). Most of these studies were performed with cells selected for phenotypic changes (e.g., retrovirus-induced transformation) and therefore did not allow any conclusions regarding a possible preference of the integration process. However, recent analyses of unselected integration events also indicate that provirus insertion is much less random than previously thought. Studies by Vijaya et al. (46) and us (31) have shown that proviral integration sites always map within a few hundred base pairs of a DNase-hypersensitive chromatin site. Shih et al. (36) have discovered highly preferred integration target sequences in the avian genome. Because the close association between retroviral integration sites and DNase-hypersensitive sites suggests that actively transcribed genome regions are preferred targets for provirus insertion, we have analyzed the transcriptional activity of cellular target sequences for provirus integration.

Random integration sites for Moloney murine leukemia virus (M-MuLV) in NIH 3T3 mouse fibroblasts were molecularly cloned as follows. 3T3 cells were infected with M-MuLV^{-sup}, a virus construct that contains a bacterial suppressor tRNA gene in its long terminal repeats (LTRs) and allows selective cloning of proviral-host DNA junction fragments (29, 31). Genomic DNA from M-MuLV^{-sup}-infected cells was digested with *Eco*RI or *Hin*dIII and ligated to *Eco*RI-digested bacteriophage lambda Charon 4A or *Hin*dIII-digested lambda NM788 arms. The presence of the bacterial suppressor tRNA gene in the LTRs of M-MuLV^{-sup} allows the direct selection of phages containing proviral insertions in Su$^-$ bacterial host strains. *Eco*RI does not cleave within the M-MuLV proviral DNA and therefore produces fragments containing a complete proviral genome plus flanking cellular sequences at either end. *Hin*dIII cleaves M-MuLV DNA once and produces fragments containing either the 5′ or the 3′ portion of the proviral DNA plus 5′- or 3′-flanking cellular sequences extending to the next *Hin*dIII site in the cellular DNA (31). Unique probes specific for individual integration sites were prepared as described previously (31). Briefly, the *Eco*RI or *Hin*dIII inserts from recombinant phages growing in Su$^-$ bacteria were recloned in pUC plasmids. The proviral and cellular fragments in each insert were identified by double digestion with *Eco*RI and *Hin*dIII, respectively, and *Bam*HI, which generates indicative internal fragments. 5′-Flanking sequences were then recovered and subcloned by using a *Bam*HI site close to the end of the 5′ LTR, and 3′-flanking sequences were recovered and subcloned by using a *Sma*I site in the 3′ LTR. Unique subfragments were then identified by appropriate restriction digests and hybridization to Southern blots of restriction enzyme-digested genomic mouse DNA.

Probes specific for nine different proviral integration sites were used. Their derivations and locations with respect to proviral integration sites are shown in Fig. 3. Note that probes H5 and H6 and probes E7A and E7B are adjacent DNA fragments derived from the same integration site. All probes used were unique, i.e., hybridized in Southern blots with a single genomic *Hin*dIII or *Eco*RI fragment (data not shown) except for clone E7; all subfragments prepared from this clone contained either highly repetitive sequences (E7A) or middle repetitive sequences (E7B). Plasmid clones containing the individual probes were digested with appropriate restriction enzymes, and Southern blots of the digested DNAs were hybridized to radiolabeled RNA transcribed in vitro (nuclear run-on transcripts) prepared from uninfected

* Corresponding author.
† Present address: Molecular Biology Institute, University of Hamburg Medical School, 2000 Hamburg, Federal Republic of Germany.

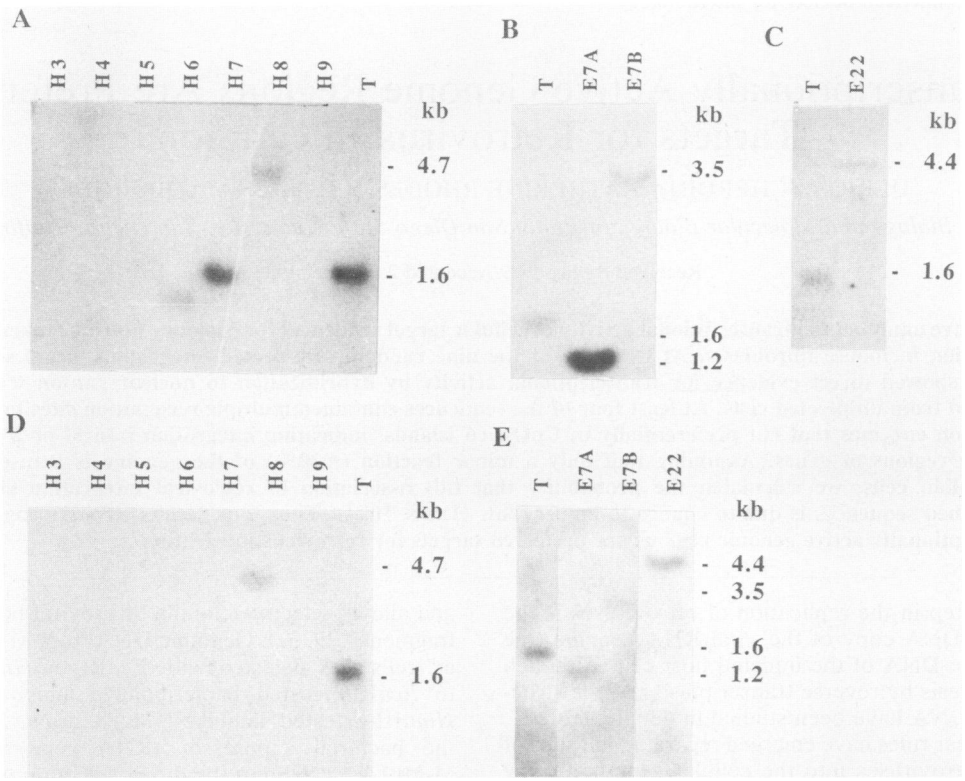

FIG. 1. Screening of retroviral integration target sequences for transcriptional activity. Plasmid clones containing the unique DNA fragments derived from nine M-MuLV integration target sequences (see Fig. 3) were digested with appropriate restriction enzymes to cut out the inserts (except for clones H8 and E22, which were linearized to produce plasmid-containing fragments of 4.7 and 4.4 kb, respectively). The DNAs were then separated by agarose gel electrophoresis, blotted onto nitrocellulose filters, and hybridized to radiolabeled nuclear run-on transcripts from 3T3 mouse fibroblasts (A to C) or P19 EC cells (D and E) as described previously (13). A plasmid containing a 1.6-kb rat alpha-tubulin cDNA clone (23) was included as a positive control (T). Numbers above the lanes refer to the integration target sequences shown in Fig. 3. Clones H5 and H6 are adjacent fragments derived from the same integration target sequence (Fig. 3). The inserts from at least five of the nine integration target sequences analyzed (H6, H7, H8, E7, and E22) hybridized to nuclear run-on transcripts from 3T3 cells. Three of these clones (H8, E7, and E22) also hybridized to transcripts from P19 cells.

3T3 cells or P19 embryonal carcinoma cells as described previously (13). Because nuclear run-on transcripts are primary transcripts and hybridize equally well to exon and intron sequences (13), this procedure also gives a positive result when integration has occurred into intron sequences. A clone containing a rat alpha-tubulin cDNA (23) was used as a positive control. At least five of the nine random, unselected integration target sequences analyzed hybridized to run-on transcripts from 3T3 cells (clones H6, H7, H8, E7, and E22; Fig. 1A to C), indicating that integration in these cases has occurred into transcribed genome regions. Two of the transcriptionally active integration target sequences (H6 and H7) were expressed in 3T3 but not P19 cells (Fig. 1D and E), whereas the remaining three (H8, E7, and E22) were expressed in both. Note that clones H5 and H6 are adjacent fragments derived from the same integration target sequence. Whereas probe H6 reproducibly showed hybridization to nuclear run-on transcripts, probe H5 did not, indicating that integration in this case has occurred into a region at which transcription either initiates or terminates. The levels of expression of the individual clones were in the range of that of the alpha-tubulin gene and were similar in both cell types except for E7A. Transcripts hybridizing to this clone were approximately 10 times more abundant in 3T3 cells than in P19 cells (compare Fig. 1A and 1E). Because the run-on transcripts were prepared from uninfected cells, our results show that the integration target sequences are transcriptionally active before integration of the virus. The nature of the transcripts (e.g., size, polyadenylation, intracellular distribution, and abundance) and the expression patterns in additional cell types have not been studied. We also do not know whether the integration sites are located in exon or intron sequences. Our results do not exclude the possibility that the remaining integration target sequences are transcriptionally active as well but are transcribed at a frequency too low to be detected by the methods used.

Alternatively, integration in these cases may have occurred into flanking regions of transcriptionally active genes. To study this latter possibility in more detail, we analyzed the integration target sequences for the presence of CpG-rich islands. Increasing evidence indicates that CpG-rich islands are associated with the 5′ ends of all housekeeping genes and many tissue-specific genes and with the 3′ ends of some tissue-specific genes but are not present in CpG-depleted bulk DNA (2, 11). The association with CpG-rich islands has in fact been exploited to detect and molecularly clone potential gene sequences in mammalian DNA (8, 25, 28). The presence of a CpG-rich island in an integration target sequence would therefore strongly suggest integration in the vicinity of a gene even in the absence of direct proof for transcriptional activity. CpG-rich islands can be located by cleavage with restriction enzymes that have one or several

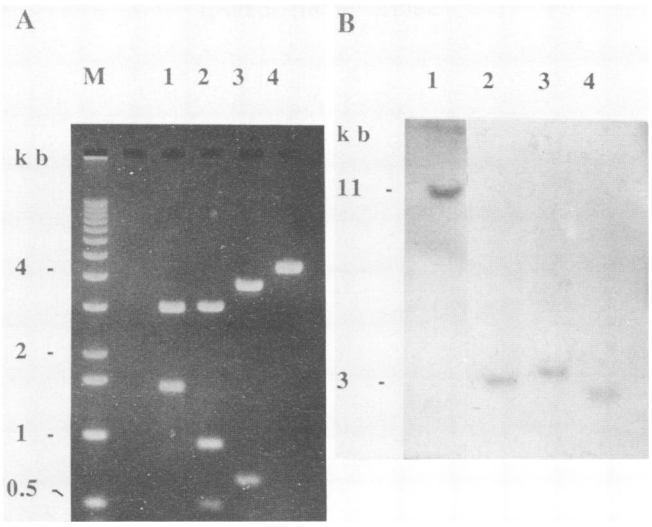

FIG. 2. Analysis of CpG-rich sequences in retroviral integration target sequences by restriction and Southern blot analysis. (A) Subclone of clone E5 containing cellular sequences located 5′ of the proviral integration site (the 1.5-kb *Eco*RI-*Pst*I fragment shown in Fig. 3) was linearized with *Eco*RI to produce a 4.2-kb fragment (lane 4) and further digested by the rare cutting C-G restriction enzymes *Sac*II (lane 3), *Sma*I (lane 2), and *Bss*HII (lane 1). The clone contains at least one restriction site each for *Sac*II and *Bss*HII and two sites for *Sma*I. Because these enzymes do not cut the vector DNA, the recognition sites must be in the insert sequences. Lane M, DNA size marker. (B) To detect possible CpG-rich sequences located near provirus integration sites but not included in the molecular clones isolated (i.e., in the portions of the integration target sequences shown in Fig. 3), genomic DNA from uninfected 3T3 mouse fibroblasts was digested with *Hin*dIII (lane 1) or double digested with *Hin*dIII and *Sma*I (lane 2), *Sac*II (lane 3), or *Bss*HII (lane 4). The DNAs were separated by gel electrophoresis, Southern blotted, and hybridized with ^{32}P-labeled H8 DNA (see Fig. 3). *Hin*dIII cleavage of genomic DNA from uninfected 3T3 cells generates an 11-kb integration target fragment (see Fig. 3) that is cleaved by *Sma*I, *Sac*II, and *Bss*HII into 2.8-, 3.2-, and 3.3-kb subfragments, respectively, showing that the genomic DNA contains a cluster of restriction sites for these enzymes in close proximity to the integration site outside of the cloned H8 DNA. In similar experiments, CpG-rich sequences were mapped by using clones H5 and H7 as hybridization probes. No CpG-rich sequences were detected in the remaining clones (data not shown).

CpG dinucleotides in their recognition sequences. Therefore, we mapped restriction sites for *Sac*II, *Sma*I, and *Bss*HII in the individual integration target sequences. Two examples are shown in Fig. 2; the restriction sites for the enzymes used and their approximate locations relative to the proviral integration sites are shown in Fig. 3. We found restriction sites for *Sac*II, *Sma*I, and *Bss*HII in four of the nine clones analyzed (H5/6, H7, H8, E5; H9 contained only a *Sma*I site). Three of these (H6, H7, and H8) also hybridized to run-on transcripts (Fig. 1), whereas one clone (E5) did not. In all cases, the sites were clustered in a region of 1 to 2 kilobase pairs (kb) and not distributed randomly over the DNA fragment analyzed. It has been estimated that recognition sites for such enzymes can be expected about once every 1,200 base pairs in CpG-rich islands but with an approximately 50- to 100-fold-lower frequency in bulk DNA (25). Thus, our results indicate that the clones contain CpG-rich islands, i.e., are likely to be derived from the 5′ or 3′ ends or flanking regions of genes. DNA sequence analyses will have to verify the presence of CpG-rich islands in the vicinity of the M-MuLV integration sites analyzed.

A summary of the results obtained so far is shown in Fig. 3. At least five of the integration target sequences showed direct evidence for transcriptional activity by hybridization to nuclear run-on transcripts. Thus, at least five of the nine random, unselected integrations of M-MuLV in mouse fibroblasts studied have occurred into transcriptionally active genome regions. This is a minimal number because the remaining integration sites may also be located in transcribed regions of the genome but were not detected by the methods used. Furthermore, one additional sequence also appears to be derived from a transcribed genome region because it contains multiple recognition sites for restriction enzymes that cut preferentially in CpG-rich islands. On the basis of saturation hybridization experiments with total cell RNA and nonrepetitive DNA, it has been estimated that about 2% of the mammalian genome is being expressed and that heterogenous nuclear RNA in mouse L cells is about fourfold more complex than mRNA (24). Thus, only a minor fraction of less than 10% of the genome is transcribed in mammalian cells. If we assume that a similar fraction of the genome comprises flanking, regulatory sequences of genes, the total transcriptionally active portion of the genome is less than 20%. The probability that five of nine random retrovirus integrations have occurred into the 20% transcriptionally active fraction of the genome by chance can be calculated to be

$$\left(\frac{9!}{5!\,4!}\right)\left(\frac{1}{5}\right)^5\left(\frac{4}{5}\right)^4$$

or 1.6×10^{-2}. This is a conservative estimate because provirus integration may be confined to areas in the 5′ or 3′ ends or flanking regions of genes characterized by a DNase-hypersensitive chromatin structure; i.e., the actual target sequences available for integration may be much less than 20% of the genome. In any case, our results strongly suggest that unselected retrovirus integrations occur preferentially into transcriptionally active genome regions. A similar preference was reported in a study of several adenovirus recombination sites which showed that all insertions analyzed had occurred into transcriptionally active genome regions (34). Similarly, there is evidence that other cellular events involving DNA recombination such as immunoglobulin gene rearrangement (3) or mating-type switching in yeast cells (20) are enhanced by transcription. Although the enzymes catalyzing different recombination events certainly differ from the retroviral integrase, there may be common general features of the host cell DNA that render a sequence a preferred target for recombination. Transcriptional activity per se may be required to establish a chromatin configuration that allows access of the foreign DNA and of the recombination machinery to the cellular DNA and may thus favor recombination. It is noteworthy that in vitro integration reactions failed to reveal preferential integration target sequences (5), supporting the notion that not the DNA sequence but the topological state of the chromatin or DNA determines the accessibility of a target sequence for provirus insertion in vivo.

Three of the nine integration target sequences studied by us so far showed neither evidence for transcriptional activity nor the presence of a CpG-rich island in the vicinity of the integration site (H3, H4, and H9; Fig. 3). We cannot rule out the possibility that these integrations also have occurred into transcribed genome regions but the transcription rate is too

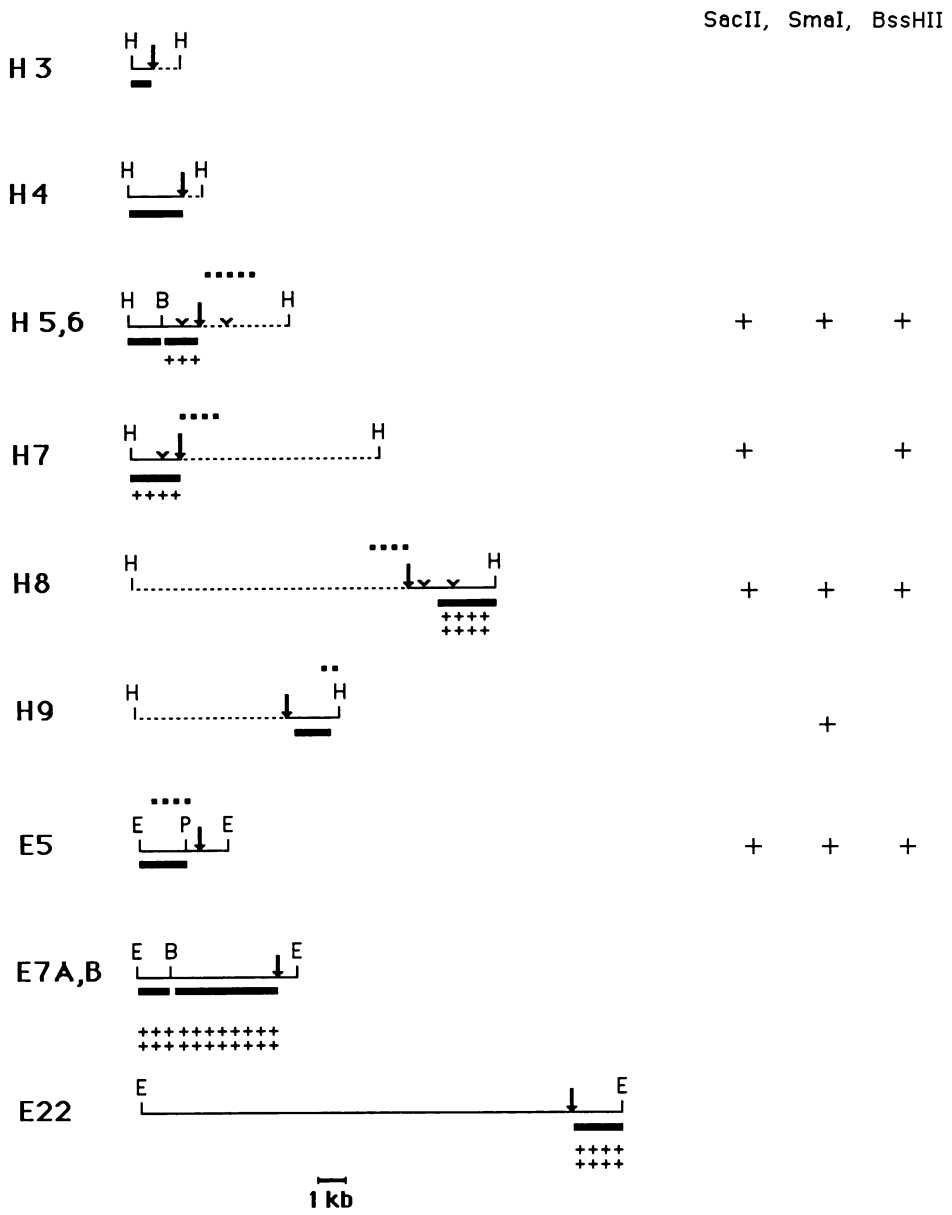

FIG. 3. Summary of analysis of unselected M-MuLV integration target sequences in 3T3 mouse fibroblasts. Clones H3 to H9 were isolated as *Hin*dIII fragments containing either a 5′ proviral moiety plus 5′-flanking cellular sequences (H3 to H7) or a 3′ proviral moiety plus 3′-flanking cellular sequences (H8 and H9). ---, Sequences outside the molecularly cloned portions of the integration target sequences; their sizes and the proviral insertion sites were deduced from restriction and Southern blotting experiments (31; data not shown). Clones E5, E7, and E22 were isolated as *Eco*RI fragments containing an entire proviral genome plus flanking cellular sequences from either end (see text). Restriction sites for *Bam*HI (B), *Eco*RI (E), *Hin*dIII (H), and *Pst*I (P) that were used to prepare the unique subfragments are indicated. Other symbols: ▬, unique probes used for Southern blot and run-on analyses; ↓, proviral integration site; ++++, sequence expressed in 3T3 cells; $\overset{++++}{++++}$, sequence transcribed in 3T3 and P19 cells; ▪▪▪▪, location of a CpG-rich sequence; V, position of a previously mapped DNase I-hypersensitive site (29) (DNase-hypersensitive sites in the remaining clones have not been mapped).

low to be detected by the methods used or that transcription initiates or terminates further upstream or downstream of the DNA fragments analyzed. It is also possible that integration in these cases has occurred into a region with an altered chromatin structure which is associated with cellular functions other than transcription, such as matrix- or topoisomerase-binding sites or origins of DNA replication. We have in fact obtained preliminary evidence that two of these clones (H3 and H4) contain a sequence capable of inducing DNA bending (D. Roche and M. Breindl, unpublished observation). Such sequences have been detected in the proximity of several origins of DNA replication, suggesting a function in the initiation of DNA replication (21, 32, 38, 39, 41, 49). Some of the highly preferred integration sites for Rous sarcoma virus studied by Shih et al. (36) appear to be associated with nuclear matrix-binding sites (J. Coffin, personal communication). Thus, retroviruses may be valuable tools for the identification and molecular cloning of DNA sequences with cellular functions other than transcription for which no screening procedures exist.

This work was supported by a grant from the San Diego State University Foundation. U.S. was supported by a postdoctoral fellowship from Deutsche Forschungsgemeinschaft.

LITERATURE CITED

1. **Ben-David, Y., V. R. Prideaux, V. Chow, S. Benchimol, and A. Bernstein.** 1988. Inactivation of the p53 oncogene by internal deletion or retroviral integration in erythroleukemic cell lines induced by Friend leukemia virus. Oncogene Res. **3**:179–185.
2. **Bird, A. P.** 1986. CpG-rich islands and the function of DNA methylation. Nature (London) **321**:209–213.
3. **Blackwell, T. K., M. W. Moore, G. D. Yancopoulos, H. Suh, S. Lutzker, E. Selsing, and F. W. Alt.** 1986. Recombination between immunoglobulin variable region gene segments is enhanced by transcription. Nature (London) **324**:585–589.
4. **Breindl, M., K. Harbers, and R. Jaenisch.** 1984. Retrovirus-induced lethal mutation in collagen I gene of mice associated with an altered chromatin structure. Cell **38**:9–16.
5. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1987. Correct integration of retroviral DNA in vitro. Cell **49**:347–356.
6. **Corcoran, L. M., J. M. Adams, A. R. Dunn, and S. Cory.** 1984. Murine T lymphomas in which the cellular myc oncogene has been activated by retroviral insertion. Cell **37**:113–122.
7. **Eibel, H., and P. Philippsen.** 1984. Preferential integration of yeast transposable element Ty into a promoter region. Nature (London) **307**:386–388.
8. **Estivill, X., M. Farrall, P. J. Scambler, G. M. Bell, K. M. F. Hawley, N. J. Lench, G. P. Bates, H. C. Kruyer, P. A. Frederick, P. Stanier, E. K. Watson, R. Williamson, and B. J. Wainwright.** 1987. A candidate for the cystic fibrosis locus isolated by selection for methylation-free islands. Nature (London) **326**: 840–845.
9. **Frankel, W., T. A. Potter, N. Rosenberg, J. Lenz, and T. V. Rajan.** 1985. Retroviral insertional mutagenesis of a target allele in a heterozygous murine cell line. Proc. Natl. Acad. Sci. USA **82**:6600–6604.
10. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54**:497–504.
11. **Gardiner-Garden, M., and M. Frommer.** 1987. CpG islands in vertebrate genomes. J. Mol. Biol. **196**:216–282.
12. **Harbers, K., M. Kuehn, H. Delius, and R. Jaenisch.** 1984. Insertion of retrovirus into first intron of alpha 1(I) collagen gene leads to embryonic lethal mutation in mice. Proc. Natl. Acad. Sci. USA **81**:1504–1508.
13. **Hartung, S., R. Jaenisch, and M. Breindl.** 1986. Retrovirus insertion inactivates some al(I)collagen gene by blocking initiation of transcription. Nature (London) **320**:365–367.
14. **Hawley, R. G., M. J. Shulman, H. Murialdo, D. M. Gibson, and N. Hozumio.** 1982. Mutant immunoglobulin genes have repetitive DNA elements inserted into the intervening sequences. Proc. Natl. Acad. Sci. USA **79**:7425–7429.
15. **Hutchinson, K. W., N. G. Copeland, and N. Jenkins.** 1984. Dilute-coat-color locus of mice: nucleotide sequence analysis of th d^{+2J} and d^{+Ha} revertant alleles. Mol. Cell. Biol. **4**:2899–2904.
16. **Ikenaga, M., and K. Saigo.** 1982. Insertion of a movable genetic element, 297, into the T-A-T-A box for the H3 histone gene in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA **79**: 4143–4147.
17. **Inouye, S., S. Yuki, and K. Saigo.** 1984. Sequence-specific insertion of the Drosophila transposable element 17.6. Nature (London) **310**:332–333.
18. **Isfort, R., R. L. Witter, and H.-J. Kung.** 1987. c-myc activation in an unusual retrovirus-induced avian T-lymphoma resembling Marek's disease: proviral insertion 5′ of exon one enhances the expression of an intron promoter. Oncogene Res. **2**:81–94.
19. **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.
20. **Klar, A. J. S., J. N. Strathern, and J. A. Abraham.** 1984. Involvement of double-strand chromosomal breaks for mating-type switching in Saccharomyces cerevisiae. Cold Spring Harb. Symp. Quant. Biol **49**:77–88.
21. **Koo, H. S., H. M. Wu, and D. M. Crothers.** 1986. DNA bending at adenine-thymine tracts. Nature (London) **320**:501–506.
22. **Lavu, S., and E. P. Reddy.** 1986. Structural organization and nucleotide sequence of mouse c-myb oncogene: activation in ABPL tumors is due to viral integration in an intron which results in the deletion of the 5′ coding sequences. Nucleic Acids Res. **14**:5309–53.
23. **Lemischka, T. R., S. Farmer, V. R. Racaniello, and P. A. Sharp.** 1981. Nucleotide sequence and evolution of a mammalian alpha tubulin messenger RNA. J. Mol. Biol. **151**:101–120.
24. **Lewin, B.** 1975. Units of transcription and translation: sequence components of heterogeneous nuclear RNA and messenger RNA. Cell **4**:77–93.
25. **Lindsay, S., and A. P. Bird.** 1987. Use of restriction enzymes to detect potential gene sequences in mammalian DNA. Nature (London) **327**:336–338.
26. **McGinnis, W., A. W. Shermoen, and S. K. Beckendorf.** 1983. A transposable element inserted just 5′ to a Drosophila glue protein gene alters gene expression and chromatin structure. Cell **34**:75–84.
27. **Panganiban, A. T.** 1985. Retroviral DNA integration. Cell **42**:5–6.
28. **Rappold, G. A., L. Stubbs, S. Labeit, R. B. Crkvenjakov, and H. Lehrach.** 1987. Identification of a testis-specific gene from the mouse t-complex next to a CpG-rich island. EMBO J. **6**: 1975–1980.
29. **Reik, W., H. Weiher, and R. Jaenisch.** 1985. Replication-competent Moloney murine leukemia virus carrying a bacterial suppressor tRNA gene: selective cloning of proviral and flanking host sequences. Proc. Natl. Acad. Sci. USA **82**:1141–1145.
30. **Robinson, H. L., and C. G. Gagnon.** 1986. Patterns of proviral insertion and deletion in avian leukosis virus-induced lymphomas. J. Virol. **57**:28–36.
31. **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.
32. **Ryder, K., S. Silver, A. L. DeLucia, E. Fanning, and P. Tegtmeyer.** 1986. An altered DNA conformation in origin region I is a determinant for the binding of SV40 large T antigen. Cell **44**:719–725.
33. **Schubach, W., and M. Groudine.** 1984. Alteration of c-myc chromatin structure by avian leukosis virus integration. Nature (London) **307**:702–708.
34. **Schulz, M., U. Freiser-Rabien, R. Jessberger, and W. Doerfler.** 1987. Transcriptional activities of mammalian genomes at sites of recombination with foreign DNA. J. Virol. **61**:344–353.
35. **Selten, G., H. T. Cuypers, M. Zijlstra, C. Melief, and A. Berns.** 1984. Involvement of c-myc in MuLV-induced T cell lymphomas in mice: frequency and mechanisms of activation. EMBO J. **3**:3215–3222.
36. **Shih, C. C., J. P. Stoye, and J. M. Coffin.** 1988. Highly preferred targets for retrovirus integration. Cell **53**:531–537.
37. **Shih, C.-K., M. Lineal, 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. USA **81**:4697–4701.
38. **Snyder, M., A. R. Buchman, and R. W. Davis.** 1986. Bent DNA at yeast autonomously replicating sequence. Nature (London) **324**:87–89.
39. **Stenzel, T. T., P. Patel, and D. Bastian.** 1987. The integration host factor of Escherichia coli bands to bent DNA at the origin of replication of the plasmid pSC101. Cell **49**:709–717.
40. **Swift, R. A., C. Boerkel, A. Ridgway, D. J. Fujita, J. B. Dodgson, and H.-J. Kung.** 1987. B-lymphoma induction by reticuloendotheliosis virus: characterization of a mutated chicken syncytial virus provirus involved in c-myc activation. J. Virol. **61**:2084–2090.
41. **Trifonov, E. N.** 1986. Curved DNA. Crit. Rev. Biochem. **19**: 89–106.
42. **Varmus, H. E.** 1982. Form and function of retroviral proviruses.

Science **216**:812–820.
43. **Varmus, H. E., T. Padgett, S. Heasley, G. Simon, and J. M. Bishop.** 1977. Cellular functions are required for the synthesis and integration of avian sarcoma virus-specific DNA. Cell **11**:307–319.
44. **Varmus, H. E., N. Quintrell, and S. Ortitz.** 1981. Retroviruses as mutagens: insertion and excision of a nontransforming provirus alter expression of a resident transforming provirus. Cell **25**: 23–36.
45. **Varmus, H. E., and R. Swanstrom.** 1985. Replication of retroviruses, p. 369–512. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
46. **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.
47. **Weinberg, R. A.** 1980. Integrated genomes of animal viruses. Annu. Rev. Biochem. **49**:197–226.
48. **Wolf, D., and V. Rotter.** 1984. Inactivation of p53 gene expression by an insertion of Moloney murine leukemia virus-like sequences. Mol. Cell. Biol. **4**:1402–1410.
49. **Zahn, K., and F. R. Blattner.** 1985. Sequence-induced DNA curvature at the bacteriophage lambda origin of replication. Nature (London) **317**:451–453.