

Small Tumor Virus Genomes Are Integrated near Nuclear Matrix Attachment Regions in Transformed Cells

KATHERINE A. SHERA,¹ CHRISTOPHER A. SHERA,^{2,3} AND JAMES K. McDOUGALL^{1*}

Cancer Biology Program, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109-1024¹; Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114²; and Department of Otolaryngology, Harvard Medical School, Boston, Massachusetts 02115³

Received 5 July 2001/Accepted 21 September 2001

More than 15% of human cancers have a viral etiology. In benign lesions induced by the small DNA tumor viruses, viral genomes are typically maintained extrachromosomally. Malignant progression is often associated with viral integration into host cell chromatin. To study the role of viral integration in tumorigenesis, we analyzed the positions of integrated viral genomes in tumors and tumor cell lines induced by the small oncogenic viruses, including the high-risk human papillomaviruses, hepatitis B virus, simian virus 40, and human T-cell leukemia virus type 1. We show that viral integrations in tumor cells lie near cellular sequences identified as nuclear matrix attachment regions (MARs), while integrations in nonneoplastic cells show no significant correlation with these regions. In mammalian cells, the nuclear matrix functions in gene expression and DNA replication. MARs play varied but poorly understood roles in eukaryotic gene expression. Our results suggest that integrated tumor virus genomes are subject to MAR-mediated transcriptional regulation, providing insight into mechanisms of viral carcinogenesis. Furthermore, the viral oncoproteins serve as invaluable tools for the study of mechanisms controlling cellular growth. Similarly, our demonstration that integrated viral genomes may be subject to MAR-mediated transcriptional effects should facilitate elucidation of fundamental mechanisms regulating eukaryotic gene expression.

Tumor virus infection represents an early and necessary event in the genesis of a variety of human cancers. However, viral infection alone is insufficient to produce a fully malignant phenotype. Progression to malignancy has been associated with the integration of viral DNA into the host genome (19, 56), but the mechanisms driving this association remain unclear. To investigate the hypothesis that malignant progression is influenced by the position of viral integration in host chromatin, we analyzed integrated oncogenic virus genomes in human tumors. The “high-risk” human papillomaviruses, HPV type 16 (HPV16) and HPV type 18 (HPV18) and hepatitis B virus (HBV) are the etiologic agents of human cervical and hepatocellular carcinoma, respectively, and as such contribute significantly to the worldwide cancer burden. Human T-cell leukemia virus type 1 (HTLV-1), a retrovirus, is associated with both adult T-cell leukemia (ATL) and HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP), a nonneoplastic condition. Simian virus 40 (SV40) has served as a model for the study of basic mechanisms of viral carcinogenesis. However, the association of this tumor virus with human cancers is controversial. Preliminary reports link SV40 to mesothelioma, osteosarcoma, and brain tumors (7).

In mammalian cells, the nuclear matrix functions in gene expression and DNA replication (14). Nuclear matrix attachment regions (MARs) are DNA sequences that associate with the nuclear matrix or scaffold (12). MARs organize the eukaryotic genome into topologically independent transcriptional do-

main (38), augment the activity of nearby enhancers (20–22, 28), and can protect DNA transgenes from cellular silencing effects through local modifications of chromatin structure (30, 49). The association of MARs with eukaryotic transcriptional regulation suggests that these sequence elements may influence the expression of integrated tumor virus genomes.

To study the role of viral integration in tumorigenesis *in vivo*, we initially focused on tumors induced by HPV18. In HPV16-associated neoplasia, viral DNA may persist in mixed episomal and integrated forms (15, 33, 37, 50). In contrast, HPV18 genomes are reported to be exclusively integrated in HPV18-related malignancies (15, 37, 50). Tumors induced by this virus therefore represent an appropriate model system in which to study the role of integration in malignant progression. We developed an Alu-PCR technique to isolate cellular sequences flanking HPV18 integrants in cervical carcinoma.

To extend our analysis, we obtained cellular sequences flanking HPV16, HBV, SV40, and HTLV-1 integrations in tumor cells and cell lines from the public sequence databases. We determined the genomic position of each viral integration by BLAST similarity search (1) of the human genome and then defined the matrix association potential (MAP) of the integration site and surrounding 100 kb, using a mathematical algorithm that statistically weights various sequence motifs associated with cellular MARs, including *ori* signals, AT and TG richness, kinked and curved DNA, and topoisomerase II binding sites (48). We verified the algorithm's ability to accurately predict the positions of MARs in DNA sequences in separate experiments. We identified high-scoring regions (HSRs), DNA regions demonstrating a significant potential for attachment to the nuclear matrix, and determined the distance in kilobase pairs, D , from the viral integration point to the closest HSR. D_{up} and D_{down} represent the distances from the upstream or

* Corresponding author. Mailing address: Cancer Biology Program, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., C1-105, Seattle, WA 98109-1024. Phone: (206) 667-4498. Fax: (206) 667-5815. E-mail: jmcDougla@fhcrc.org.

downstream viral insertion point, IP_{up} or IP_{down} , to the closest HSR upstream or downstream, respectively.

We calculated the corresponding mean distances to the closest HSR, d_{up} and d_{down} , for viral integrations in both transformed and nonneoplastic cells. In some instances, viral integrations in transformed cells disrupted genes involved in the control of cellular growth, conferring a proliferative advantage to the host cell irrespective of viral gene expression; independent means were calculated for these integrations and for those that occurred outside known genes. The statistical significance of the observed mean distances was determined using Monte Carlo simulations to test the hypothesis that viral integration occurred randomly with respect to HSRs. We show that tumor virus integrations in transformed cells and cell lines are adjacent to cellular sequences identified as MARs. Viral integrations in nonneoplastic cells, and in tumor cells in which the integration disrupted genes involved in cellular growth control, show no association with these regions. MAR-mediated effects on the expression of integrated viral oncogenes might confer a proliferative advantage to a cell and its progeny, promoting tumor development.

MATERIALS AND METHODS

Tissue preparation and HPV detection. Cervical tumors were obtained as part of a related project using protocols consistent with National Institutes of Health human studies guidelines. DNA was extracted by standard methods (44). HPV18-positive tumors were identified as previously described (2).

Isolation of HPV18 flanking sequences. PCR amplification was performed with a sense or antisense biotinylated primer specific for HPV18-E6/E7, together with a sense or antisense primer complementary to the most highly conserved residues of the Alu repeat subfamilies Sb, Sb-1, Sb-2, Sc, Sp, Sq, Sx, and Alu-J (27). 3' mismatches were incorporated to prevent amplification of 7SL RNA sequences. After purification of the PCR products over streptavidin-coated beads, a second PCR with an internal HPV primer was performed. An Alu-HPV18 hybrid plasmid was constructed from Alu consensus sequence pPD39 (4) and plasmid pHPV18 for use as a positive control. Normal genomic DNA and DNA extracted from an HPV16-positive cervical carcinoma served as negative controls. Cellular flanking sequences were reproducibly amplified in separate experiments. PCR conditions were as follows: 30 ng of tumor DNA or 30 pg of hybrid plasmid was amplified in a 50- μ l PCR mixture containing 1 \times Expand buffer (Roche, Indianapolis, Ind.), 2.6 U of Expand High Fidelity enzyme (Roche), 2.0 mM $MgCl_2$, a 200 μ M concentration of each deoxynucleoside triphosphate (Amersham Pharmacia Biotech, Piscataway, N.J.), and a 25 nM concentration of either R18Bio or F18Bio with either AluF or AluR. Thermocycling parameters were as follows: 3 min at 95°C; 30 s at 95°C, 15 s at 61°C, and 75 s at 72°C for a total of 34 cycles and a final 10 min at 72°C. The amplified biotinylated fragments were purified on Dynabeads-M280 Streptavidin (Dyna, Oslo, Norway). One microliter of the purified products was amplified in a second-round PCR with primers F18Int or R18Int and AluF or AluR. Second-round PCR conditions were identical to the first, but the cycles were limited to 25. A 5- μ l volume of each second-round PCR product was electrophoresed in 1% agarose, transferred to a positively charged nylon membrane, and hybridized with digoxigenin-11-dideoxy-UTP-labeled (Roche) HPV18-specific oligonucleotide probe 18E6-3. Positive fragments were cloned into pCR 2.1-Topo vector (Invitrogen, Carlsbad, Calif.). Automated sequencing was performed using Big-Dye Terminator cycle sequencing (ABI Prism, Foster City, Calif.). DNA sequence analysis was performed with Wisconsin Package Version 10.1 (Genetics Computer Group, Madison, Wis.). PCR primer and probe sequences follow: R18Bio, 5'-Bio-TACTGTGTGTTTCTCTGCGTC-3'; F18Bio, 5'-Bio-CCCTACAA GCTACCTGATCTG-3'; AluF, 5'-CCTGTAATCCAGCACTTTG-3'; AluR, 5'-CCCAAAGTGCTGGGATTAC-3'; F18Int, 5'-ACAGTATACCCCATGCTGC-3'; R18Int, 5'-TTCAACGGTTTCTGGCAC-3'; and 18E6-3, 5'-CAGACTCTGTG TATGGAGACACATTGGA AAAA CTA AACTAA-3'.

Other viral flanking sequences. Cellular sequences flanking single viral integrants, isolated by variety of methods from tumors, transformed cell lines, and nonneoplastic T cells, were obtained from the scientific literature and public sequence databases. The genomic position of each viral integration was deter-

mined by BLAST similarity search (version 2.1.2) of cellular flanking sequence against the nucleotide and high-throughput genome sequence databases. The maximum E value permissible for genomic localization of a viral integration was 10^{-25} ; the majority demonstrated perfect or near-perfect identity. No insertion point could be defined for HBV flanking sequences AJ222811 or AJ000498 or the SV40 integration in FRA7H (AF017104); these sequences were excluded from the analysis.

MAP analysis and proximity to viral insertions. MAP was defined for the 100 kb of DNA sequence surrounding the integration site, when available, using Mar-Finder (<http://www.futuresoft.org>). MAP analysis was performed in 50-kb increments utilizing the default detection and clipping parameters (clip = 0.5 ρ_{max} , where ρ = MAP). When viral integration occurred antiparallel to unfinished genomic sequence, MAP analysis was performed on the reverse complement of that sequence to facilitate statistical analysis. The minimum analyzed interval was 5 kb. The upstream and downstream IPs, IP_{up} and IP_{down} , were the cellular nucleotides adjacent to the 5' and 3' ends, respectively, of the viral coding strand, where "coding strand" refers to the 5'-to-3' nucleotide sequence indexed in the public sequence databases. When an integration occurred in oriented cellular sequence and the viral and cellular coding strands were antiparallel, the IPs and distances D were assigned with respect to the cellular "coding strand." For each viral insertion, D_{up} and D_{down} were calculated as the distance in kilobase pairs from IP_{up} or IP_{down} to the closest HSR upstream or downstream, respectively, taking the endpoints of each analyzed interval as HSRs. When an IP occurred within an HSR, $D = 0$. Mean distances d_{up} and d_{down} were calculated for the subsets of viral integrations listed (see Table 2).

Validation of MAP algorithm. The ability of the MAP algorithm to accurately predict DNA sequences that attach to the nuclear matrix was verified by analysis of genomic regions containing MARs characterized in vivo or in vitro. When the default detection parameters were used, there was excellent correlation between HSRs and known MARs in the serpin gene cluster (41), the apolipoprotein B gene (32), the immunoglobulin κ enhancer (54), the immunoglobulin heavy-chain μ enhancer (12), and intron 13 of topoisomerase I (43). Recognition of the MAR in intron 2 of topoisomerase I (43) and the corticosteroid-binding globulin gene (41) intronic MAR improved when the clip was $< 0.4 \rho_{max}$, where ρ = MAP.

Statistical methods. We performed Monte Carlo simulations to test the hypothesis that viral integration occurred randomly with respect to HSRs. For each subset of viral integrations, we chose insertion points at random from the genomic intervals in which the in vivo integrations occurred and calculated the mean distance d_{up} or d_{down} to the nearest HSR for each of 10,000 independent simulations. P gives the fraction of simulated mean distances less than or equal to the observed mean; its value thus estimates the probability of the observed mean under the assumption that viral integration occurs randomly with respect to HSRs.

Database deposition. Cellular sequences flanking integrated HPV18 in cervical carcinoma are available under GenBank accession numbers AF339133, AF339134, AF339135, AF339136, AF339137, AF339138, and AF339139.

RESULTS

We isolated cellular sequences flanking integrated HPV18 genomes from five of nine cervical tumors studied, including squamous-cell carcinoma, adenocarcinoma, and small-cell cervical cancers (Fig. 1). Sequences flanking HPV16, HBV, SV40, and HTLV-1 integrations, isolated from both neoplastic and nonneoplastic viral pathologies, were obtained from the scientific literature and public sequence databases, for a total of 54 viral flanking sequences. After BLAST localization, we determined the MAP of the genomic region surrounding each viral integration site. We identified HSRs, DNA regions demonstrating a significant potential for attachment to the nuclear matrix, and determined the distance, D , in kilobase pairs from the viral integration point to the closest HSR (Fig. 2). D_{up} and D_{down} represent the distances from the upstream or downstream viral insertion point, IP_{up} or IP_{down} , to the closest HSR upstream or downstream, respectively.

Accession numbers of viral flanking sequences, highest-scoring BLAST homologies, and corresponding distances D to the closest HSR are given in Table 1. The mean distance between

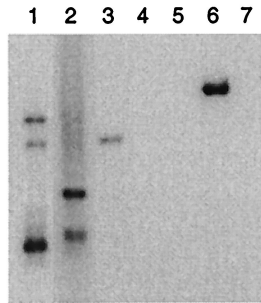


FIG. 1. Southern hybridization of cellular sequences flanking integrated HPV18 in cervical carcinoma. Alu-PCR products were hybridized with an HPV18-specific oligonucleotide probe. Lanes 1 and 2, HPV18(+) cervical squamous-cell carcinoma; lane 3, HPV18(+) cervical adenocarcinoma; lane 4, HPV16(+) cervical squamous-cell carcinoma; lane 5, normal human genomic DNA; lane 6, Alu-HPV18 hybrid plasmid (2.3 kb); and lane 7, negative control. Tumors 1 and 2 harbored multiple HPV18 integrations.

successive HSRs in the analyzed intervals surrounding the viral integration sites was 10.2 kb; no significant difference was observed between the mean distances in transformed cells (9.7 kb) and nonneoplastic cells (11.9 kb). The observed mean distances to the closest HSR, d_{up} and d_{down} , for viral integrations in transformed and nonneoplastic cells are given in Table 2.

Viral integrations in transformed cells. Six cellular sequences upstream of high-risk HPV integrations, isolated from five independent cervical tumors, were analyzed. Although the

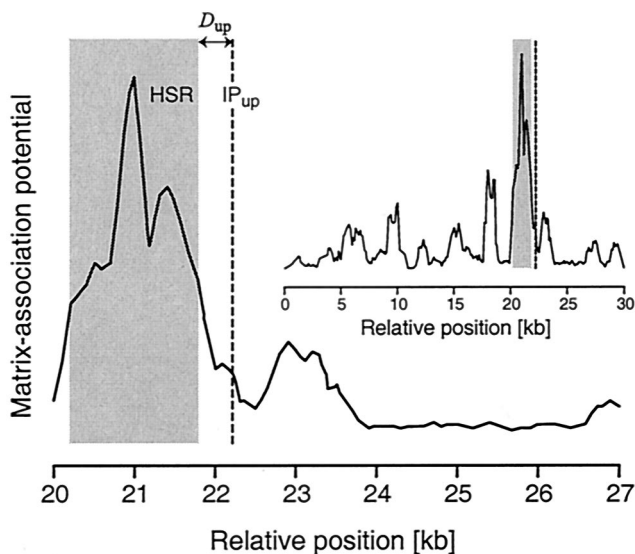


FIG. 2. MAP of the genomic region surrounding an HPV18 integration in a cervical carcinoma. The HSR, defined using a mathematical algorithm that weights sequence motifs associated with cellular MARs, demonstrates a significant potential for attachment to the nuclear matrix. D_{up} is the distance from the IP_{up} to the HSR. In this instance, the position of the downstream viral IP is unknown. The inset shows the full 30-kb analyzed interval, which contains a single HSR. In our study, MAP and HSR boundaries were defined using the MAR prediction algorithm's default detection and clipping parameters. The MAP of the genomic region illustrated here is shown with unclipped peaks for clarity.

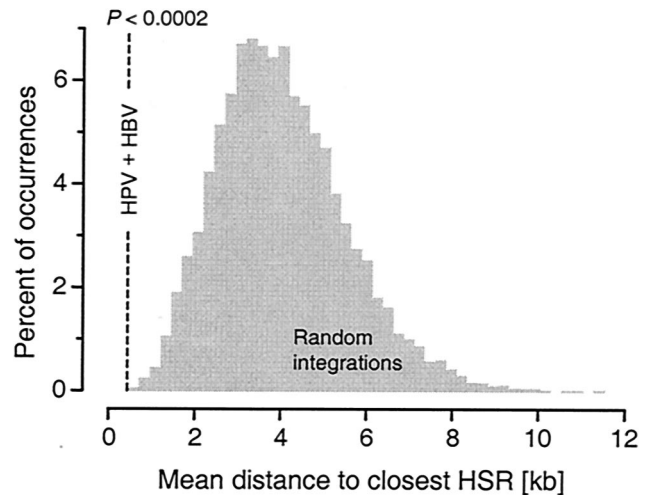


FIG. 3. Mean distance to an HSR, d_{up} , for high-risk HPV and HBV integrations in tumor cells and the distribution of means in Monte Carlo simulations. The distribution was computed under the assumption that viral integration occurs randomly with respect to HSRs. P gives the fraction of simulated mean distances less than or equal to the observed mean; here, $P < 0.0002$.

mean distance between successive HSRs in the genomic intervals in which the HPV integrations occurred was almost 10 kb, six of six in vivo integrations localized within approximately 1/2 kb of an upstream HSR. The mean distance to an HSR upstream, d_{up} , in HPV-associated tumors, together with that observed in the cervical carcinoma cell line SiHa, which harbors a single HPV16 integration (34), was 0.57 kb ($n = 7$, $P < 0.005$). HPV downstream flanks, however, showed no significant correlation with HSRs ($d_{down} = 7.3$ kb, $n = 10$). The difference between the HPV upstream and downstream means, d_{up} and d_{down} , was statistically significant ($P < 0.004$).

Similarly, HBV integrations in hepatocellular carcinoma cells occurred within HSRs ($d_{up} = 0.0$ kb, $n = 3$, $P < 0.002$ and $d_{down} = 0.0$ kb, $n = 3$, $P < 0.002$). Taken together, the mean distance d_{up} to the closest HSR for HPV and HBV integrations in tumor cells was 0.401 kb ($n = 10$, $P < 0.0002$). Random integrations simulated in those same genomic intervals occurred an average of 4 kb from an HSR (Fig. 3).

Although our sample size for SV40 integrations in transformed cell lines was small, the results were particularly intriguing. Here, SV40 downstream, but not upstream, flanking sequences were adjacent HSRs ($d_{down} = 0.66$ kb, $n = 3$, $P < 0.01$ and $d_{up} = 10.2$ kb, $n = 2$). Interestingly, the SV40 onco-genes, large and small T antigens, are transcribed from the minus strand of this bidirectionally transcribed virus. Our results suggest that the effect of a nearby MAR on the expression of an integrated viral genome is directional and is manifest when a MAR is 5' on the viral coding strand.

Viral integrations that disrupted growth control genes. Some HPV and HBV integrations disrupted genes involved in the control of cellular growth, including *TP73*, *ERBB2*, *MYCN*, *TERT*, *RARB*, and *CCNA2* (10, 16, 34, 39, 45, 52). The integrations in *ERBB2*, *MYCN*, and *TERT* occurred in intronic, untranslated exonic, and regulatory cellular sequences, respectively. Increased expression of the virally disrupted gene was

TABLE 1. Genomic localizations of integrated tumor virus genomes in tumor cells, tumor cell lines, and nonneoplastic cells and the distance D in base pairs from viral IP to closest HSR^a

Cell type or lines	Virus	Source	Accession no. or reference	Localization	D_{up}	D_{down}	
Tumor cells	HPV18	CxCa	AF339133	AC008870	690		
	HPV18	CxCa	AF339134	AC008072	168		
	HPV18	CxCa	AF339135	AC040933	0		
	HPV18	CxCa	AF339136	AC013415	417		
	HPV16	CxCa	M33610	AC007982	0		
	HPV16	CxCa	L43000	AC068359	888		
	HPV18	CxCa	AF339137	AC027364		554	
	HPV18	CxCa	AF339138	AC087590		2,612	
	HPV16	CxCa	M33611	AC007982		3,738	
	HPV16	CxCa	M33613	AF152363		6,215	
	HPV16	CxCa	M33612	AC007282		12,314	
	HPV16	CxCa	M33616	AL133377		15,459	
	HPV16	CxCa	L42999	AL033527		12,794	
	HPV16	CxCa	M33614	U66722		12,445	
	HPV16	CxCa	51	AC026075		3,940	
	HPV16	CxCa	M59869	AL136528 ^b	32,221		
	HPV16	CxCa	M59869	AL136528 ^b		4,566	
	HPV16	CxCa	13	AC004518	10,637		
	HPV16	CxCa	13	AC004518		24,766	
	HPV16	CxCa	51	AC005325	10,746		
	HPV18	CxCa	AF339139	AB025285		4,007	
		HBV	HCC	AJ000499	AC022525	0	
		HBV	HCC	M27097 ^c	AL049696	0	
		HBV	HCC	M16635	AC026982	0	
		HBV	HCC	AJ000499	AC022525		0
		HBV	HCC	M27097 ^c	AL049696		0
		HBV	HCC	M16635	AC026982		0
		HBV	HCC	K02716	AC011323	347	
		HBV	HCC	K02717	AC011323		2,940
		HBV	HCC	S76119	AF128893		1,642
		HBV	HCC	52	AC079341	38	
		HBV	HCC	52	AC079341		356
Tumor cell line	HPV16	SiHa	AF001600	AL158191	1,849		
	HPV16	SiHa	AF001599	AL356416		2,870	
	HPV45	IC-4	AJ242956	AC010145	16,071		
	HPV45	IC-4	AJ242956	AC010145		2,949	
	HBV	HUH2-2	M11225	AC005203	10,931		
	HBV	HUH2-2	M11225	AC005203		258	
	SV40	VA13	D00846	AC012178	20,454		
	SV40	VA13	D00847	AC012178		1,291	
	SV40	AG34	U08313	AC009075	0		
	SV40	AG34	U08313	AC009075		440	
	SV40/Ad5	H13.1	X71401	AL049569		248	
	HTLV-1	MT-4	S80210	AC036142		266	
	HTLV-1	MT-4	S80215	AL136981	0		
	HTLV-1	MT-4	S80213	AC021942		0	
	HTLV-1	MT-4	S80212	AC007222		26,450	
Nonneoplastic cells	HTLV-1	HAM/TSP	AY003898	AL163209	4,862		
	HTLV-1	HAM/TSP	AY003897	AC016195		5,193	
	HTLV-1	HAM/TSP	AY003891	AC011753		2,079	
	HTLV-1	HAM/TSP	AY003890	AL023806		0	
	HTLV-1	HAM/TSP	AY003892	AC016941		56,400	
	HTLV-1	HAM/TSP	AY003888	AL512422		8,995	
	HTLV-1	HAM/TSP	AY003886	AL163204	6,965		
	HTLV-1	HAM/TSP	AY003894	AC012559		13,317	
	HTLV-1	HAM/TSP	AY003893	AC010552		2,182	
	HTLV-1	HAM/TSP	AY003899	AL132712		13,933	

^a CxCa, cervical carcinoma; HCC, hepatocellular carcinoma. When an IP occurred within an HSR, $D = 0$.^b $E > 10^{-25}$, but both flanks localized to the same clone.^c Clip = 0.4 (see Materials and Methods).

TABLE 2. Mean distances from viral IP to closest HSR (kb)^a

Subset	Virus	d_{up}	d_{down}
a	HPV	0.57 ($n = 7$, $P < 0.005$)	7.3 ($n = 10$)
	HBV	0.00 ^b ($n = 3$, $P < 0.002$)	0.00 ($n = 3$, $P < 0.002$)
	SV40	10.2 ($n = 2$)	0.66 ($n = 3$, $P < 0.01$)
	HTLV-1 in ATL		0.09 ^c ($n = 3$)
b	HPV + HBV in growth control genes	11.6 ($n = 7$)	5.2 ($n = 8$)
c	HTLV-1 in HAM/TSP	5.9 ($n = 2$)	12.8 ($n = 8$)

^a d_{up} and d_{down} give the mean distance from the viral IP to the closest HSR for the following subsets of viral integrations: a, integrations in tumors and transformed cell lines; b, integrations in tumors and cell lines that disrupted genes involved in the control of cellular growth; and c, integrations in nonneoplastic cells. P values are given when observed means are significantly different from values obtained under the assumption of random integration.

^b When an IP occurred within an HSR, $D = 0$.

^c Upstream and downstream.

demonstrated for *MYCN*, *RARB*, and *CCNA2* (16, 45, 53). The insertional activation of host genes, by either direct mutation or transcriptional activation, is reminiscent of cellular transformation induced by the nonacutely transforming retroviruses; here, viral integration promotes tumorigenesis irrespective of viral gene expression. Interestingly, the positions of HPV and HBV integrations in cellular growth control genes showed no association with HSRs ($d_{\text{up}} = 11.6$ kb, $n = 7$ and $d_{\text{down}} = 5.2$ kb, $n = 8$).

HTLV-1 proviral integrations in nonneoplastic and transformed cells. Finally, HTLV-1 integrations in nonneoplastic T cells showed no correlation with HSRs ($d_{\text{up}} = 5.9$ kb, $n = 2$ and $d_{\text{down}} = 12.8$ kb, $n = 8$). Several studies of HTLV-1 integration in adult T-cell leukemia have been published, but few flanking sequences have been deposited in the public databases. Our analysis of HTLV-1 insertions in malignant cells was therefore limited to the ATL cell line MT-4. MT-4 harbors multiple HTLV-1 integrations and expresses high levels of HTLV-1 Tax protein (26). Although HTLV-1 proviral insertions in nonneoplastic cells showed no association with HSRs, three of the four integrations isolated from this transformed cell line (8) were immediately adjacent to HSRs ($d = 0.09$ kb).

Thus, tumor virus integrations in transformed cells and cell lines were adjacent to cellular sequences identified as MARs (Fig. 4). Viral integrations in nonneoplastic cells, and in tumor cells in which the integration disrupted genes involved in cellular growth control, showed no association with these regions.

DISCUSSION

Here we demonstrate a significant correlation between the positions of diverse integrated tumor virus genomes and sequences identified as MARs. Where these viral integrations in tumor cells do not disrupt genes involved in the control of cellular growth, they lie near MARs, thereby providing strong evidence for an association with tumorigenesis. MAR-mediated effects on the expression of integrated viral oncogenes might confer a proliferative advantage to a cell and its progeny, promoting tumor development.

Due to their limited genetic content, the small tumor viruses depend on the host cell machinery for replication. The viral oncoproteins HPV16 and HPV18 E6/E7, SV40 T antigen, and HTLV-1 Tax target regulators of the cell cycle, including p53, Rb, and CBP/p300, to promote cellular proliferation and a

stable environment for viral replication (7, 25). Because they functionally inactivate cellular tumor suppressors, the viral oncoproteins play an important role in the development of human cancers. However, only a small percentage of cells infected with oncogenic viruses progress to malignancy. In cervical carcinogenesis, malignant progression correlates with papillomaviral integration into host chromatin. Hepatocarcinogenesis, though incompletely understood, likely involves a complex interplay between chronic liver inflammation and the effects of HBV gene expression on hepatocellular growth control and resistance to apoptosis. Again, the integration of HBV DNA into host chromatin represents an important step in the development of hepatocellular carcinoma.

In HPV-related cancers, the integration of viral DNA into the human genome is believed to lead to transcriptional deregulation of the viral transforming genes E6 and E7. Integration frequently disrupts the viral E2 open reading frame (3, 46); loss of the E2 gene product relieves its repression of the HPV E6/E7 promoter, increasing expression of the viral oncoproteins (47). In the absence of integration, mutational inactivation of E2 has also been demonstrated to increase the transforming potential of HPV16 in vitro (42). However, no parallel mechanism explaining the role of integration in the development of cancers associated with other oncogenic viruses has been elucidated, suggesting that a more unifying mechanism explaining the role of integration in cellular transformation exists.

If integrated tumor virus genomes are subject to position effects of the type noted for endogenous (23) and therapeutic (36) transgenes, expression of integrated viral sequences might be maintained or increased in some cellular loci. Cells harboring such integrants would manifest higher levels of the viral oncoproteins, exhibit increased proliferative capacity, and undergo clonal expansion. Interestingly, variation in HPV16 E7 protein levels has been noted among subclonal populations of cervical epithelial cells harboring exclusively integrated HPV16 DNA (29). Despite disruption of the E2 open reading frame in all populations, significant variation in cellular E7 protein levels was demonstrated and in some cases equaled that seen in Caski, a cervical cancer cell line harboring an estimated 60 to 600 copies of HPV16 (34). The observed variation in E7 protein levels in independent populations of epithelial cells harboring integrated HPV16 might reflect posi-

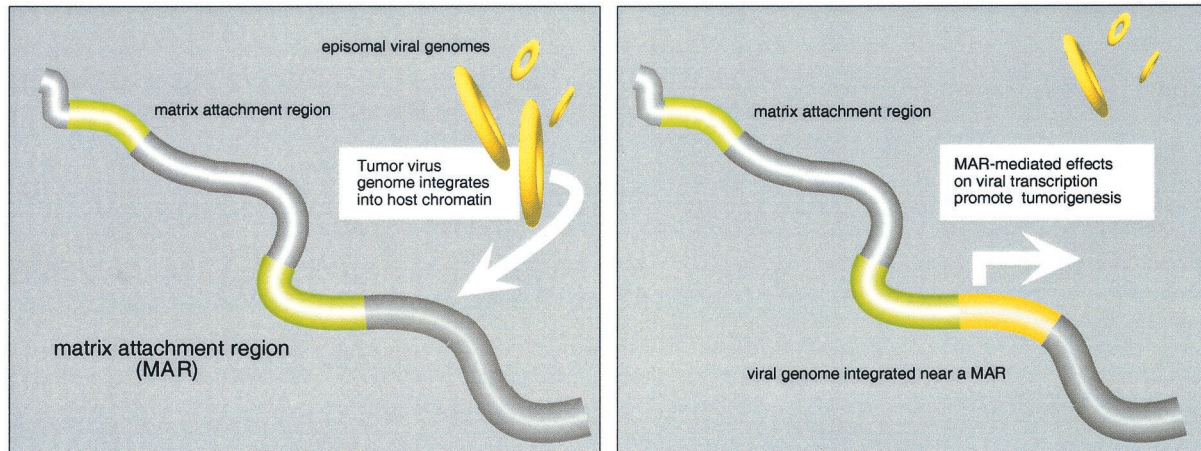


FIG. 4. Integrated tumor virus genomes in transformed cells and cell lines are adjacent to MARs. Viral integrations in nonneoplastic cells, and in tumor cells in which the integration disrupted genes involved in cellular growth control, show no association with these regions. MAR-mediated effects on the expression of integrated viral oncogenes might confer a proliferative advantage to a cell and its progeny, promoting tumor development.

tion-mediated effects on the expression of the integrated viral oncogenes.

Some viral integrations disrupt genes involved in the control of cellular growth, indirectly conferring a proliferative advantage to the host cell. We show that, in the absence of such alterations, integrated DNA and RNA tumor viruses in tumors and tumor cell lines lie near cellular sequences identified as MARs. The proximity of viral integrants to cellular MARs could simply reflect enhanced recombigenicity in these genomic loci. However, the random distribution of viral integrants in nonneoplastic cells and in tumor cells with disruptions in growth control genes argues against an absolute mechanistic requirement for integrations within or near MARs.

A previous report located a single woodchuck hepatitis virus integration in a woodchuck hepatocellular carcinoma near a MAR (18). However, the ubiquity of MARs in mammalian genomes implies that analysis of a single viral integration site cannot establish an association between integrated viral genomes and MARs, let alone with tumorigenesis. Associations of viral integration sites with local AT richness (11) and polypurine or polypyrimidine (9) or topoisomerase I motifs (31) have also been reported. Although no correlation of topoisomerase I motifs with MARs has been demonstrated, AT richness and polypurine and polypyrimidine tracts are fundamental sequence characteristics of MARs; our study thus confirms and extends previous observations. We propose that the proximity of integrated viral genomes to cellular MARs demonstrated here represents a unifying mechanism underlying the role of viral integration in the development of diverse human cancers.

In this study, MARs were identified using a computational algorithm that statistically weights sequence motifs characteristic of MARs identified by traditional biochemical methods. These methods rely upon nuclear isolation, followed by separation of nuclear matrix and loop fractions and hybridization with specific probes (in vivo assay), or alternatively by reassociation of putative MARs to the purified nuclear matrix (in vitro assay). Because these assays are subject to experimental variability (17), many studies employ both in vivo and in vitro

biochemical approaches to more definitively characterize putative MARs in DNA sequences. In our laboratory, we observed a strong correlation between the positions of MARs identified in vivo and/or in vitro and those predicted in silico. The concordance between the boundaries of experimentally determined MARs and HSRs typically averaged less than 100 or 200 bp over a 50-kb analyzed interval. Further, the MAR-Finder algorithm has been validated in several independent studies (24, 40). By providing data about the relative frequency of specific sequence motifs within each HSR, the use of a computational approach for the characterization of MARs in virally transformed cells may help to identify subsets or clusters of sequence motifs that typify those classes of MARs playing roles in the modulation of cellular gene expression, providing the beginnings of a resolution of these diverse regulatory elements at the sequence level.

Our data are consistent with the hypothesis that tumor virus genomes are differentially expressed when integrated in close proximity to MARs. Specific mechanisms for the transcriptional activation of viral oncogenes in host chromatin remain to be elucidated. Functionally diverse, MARs define chromatin domains, augment the activity of flanking enhancers, and protect transgenes from position effects. These sequence elements lack a consensus motif; MAR-binding proteins typically recognize structural features of these AT-rich DNA elements, which display enhanced flexibility and a propensity for base-unpairing and DNA bending and unwinding (5).

HMGA (previously known as HMG-I/Y) is a MAR-binding, nonhistone chromatin protein that displays a marked affinity for the bent DNA and AT tracts characteristic of MARs (5). Remarkably, the HPV18 E6/E7 promoter was recently reported to be under the control of an enhanceosome containing HMGA and the JunB/Fra-2 heterodimer (6). Enhanceosomes are cooperatively assembled nucleosome complexes that incorporate DNA-bending architectural proteins such as HMGA with sequence-specific transcription factors (35). Binding of the architectural protein to a specific DNA structural motif facilitates the recruitment of transcription factors and the for-

mation of a stable, three-dimensional enhanceosome complex; the assembled enhanceosome induces a dramatic increase in the rate at which preinitiation complexes assemble at a given promoter (55).

Thus, the integration of an HPV18 genome adjacent to a MAR might facilitate the assembly of the HMGA-containing enhanceosome on the viral promoter, driving the transcription of the integrated HPV oncogenes. Experiments in our laboratory are currently under way to investigate the influence of MARs on the activity of the HPV18 promoter and specifically on the assembly of HMGA-containing enhanceosomes. Other integrated DNA and RNA tumor viruses may be activated through similar MAR-mediated effects on transcriptional initiation.

ACKNOWLEDGMENTS

We acknowledge the contributions of our colleagues in preparation of the manuscript. We thank Connie P. Matthews and Jenn Ihle Koo for tissue acquisition; George Bonnet, Marija Helt, Andy McShea, Mike Barrett, and Brooks Shera for helpful discussions; and Denise Galloway, Steve Schwartz, and Keith Fournier for insightful comments on the manuscript. pHPV18 was a gift of H. zur Hausen.

This work was supported by grants NCI CA42792 (J.K.M.) and NIDCD DC03687 (C.A.S.).

REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Anderson, S., K. Shera, J. Ihle, L. Billman, B. Goff, B. Greer, H. Tamimi, J. McDougall, and A. Klingelutz. 1997. Telomerase activation in cervical cancer. *Am. J. Pathol.* **151**:25–31.
- Baker, C. C., W. C. Phelps, V. Lindgren, M. J. Braun, M. A. Gonda, and P. M. Howley. 1987. Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. *J. Virol.* **61**:962–971.
- Batzer, M. A., M. Alegria-Hartman, and P. L. Deininger. 1994. A consensus Alu repeat probe for physical mapping. *Genet. Anal. Tech. Appl.* **11**:34–38.
- Bode, J., C. Benham, A. Knopp, and C. Mielke. 2000. Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements). *Crit. Rev. Eukaryot. Gene Expr.* **10**:73–90.
- Bouallaga, I., S. Massicard, M. Yaniv, and F. Thierry. 2000. An enhanceosome containing the Jun B/Fra-2 heterodimer and the HMGI(Y) architectural protein controls HPV 18 transcription. *EMBO Rep.* **1**:422–427.
- Butel, J. S. 2000. Viral carcinogenesis: revelation of molecular mechanisms and etiology of human disease. *Carcinogenesis* **21**:405–426.
- Cavrois, M., S. Wain-Hobson, and E. Wattel. 1995. Stochastic events in the amplification of HTLV-I integration sites by linker-mediated PCR. *Res. Virol.* **146**:179–184.
- Choo, K. B., C. M. Chen, C. P. Han, W. T. Cheng, and L. C. Au. 1996. Molecular analysis of cellular loci disrupted by papillomavirus 16 integration in cervical cancer: frequent viral integration in topologically destabilized and transcriptionally active chromosomal regions. *J. Med. Virol.* **49**:15–22.
- Choo, K. B., H. H. Lee, L. N. Liew, K. Y. Chong, and H. F. Chou. 1990. Analysis of the unoccupied site of an integrated human papillomavirus 16 sequence in a cervical carcinoma. *Virology* **178**:621–625.
- Chou, K. S., A. Okayama, I. J. Su, T. H. Lee, and M. Essex. 1996. Preferred nucleotide sequence at the integration target site of human T-cell leukemia virus type I from patients with adult T-cell leukemia. *Int. J. Cancer* **65**:20–24.
- Cockerill, P. N., and W. T. Garrard. 1986. Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. *Cell* **44**:273–282.
- Cone, R. W., A. C. Minson, M. R. Smith, and J. K. McDougall. 1992. Conservation of HPV-16 E6/E7 ORF sequences in a cervical carcinoma. *J. Med. Virol.* **37**:99–107.
- Cook, P. R. 1999. The organization of replication and transcription. *Science* **284**:1790–1795.
- Cullen, A. P., R. Reid, M. Campion, and A. T. Lorincz. 1991. Analysis of the physical state of different human papillomavirus DNAs in intraepithelial and invasive cervical neoplasm. *J. Virol.* **65**:606–612.
- Dejean, A., L. Bougueleret, K. H. Grzeschik, and P. Tiollais. 1986. Hepatitis B virus DNA integration in a sequence homologous to v-erb-A and steroid receptor genes in a hepatocellular carcinoma. *Nature* **322**:70–72.
- Donev, R. M. 2000. The type of DNA attachment sites recovered from nuclear matrix depends on isolation procedure used. *Mol. Cell. Biochem.* **214**:103–110.
- D'Ugo, E., R. Bruni, C. Argentini, R. Giuseppetti, and M. Rapicetta. 1998. Identification of scaffold/matrix attachment region in recurrent site of wood-chuck hepatitis virus integration. *DNA Cell Biol.* **17**:519–527.
- Feitelson, M. A. 1999. Hepatitis B virus in hepatocarcinogenesis. *J. Cell. Physiol.* **181**:188–202.
- Fernandez, L. A., M. Winkler, W. Forrester, T. Jenuwein, and R. Grosschedl. 1998. Nuclear matrix attachment regions confer long-range function upon the immunoglobulin mu enhancer. *Cold Spring Harbor Symp. Quant. Biol.* **63**:515–524.
- Fernández, L. A., M. Winkler, and R. Grosschedl. 2001. Matrix attachment region-dependent function of the immunoglobulin μ enhancer involves histone acetylation at a distance without changes in enhancer occupancy. *Mol. Cell. Biol.* **21**:196–208.
- Forrester, W. C., L. A. Fernandez, and R. Grosschedl. 1999. Nuclear matrix attachment regions antagonize methylation-dependent repression of long-range enhancer-promoter interactions. *Genes Dev.* **13**:3003–3014.
- Gdula, D. A., T. I. Gerasimova, and V. G. Corces. 1996. Genetic and molecular analysis of the gypsy chromatin insulator of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **93**:9378–9383.
- Glazko, G. V., I. B. Rogozin, and M. V. Glazkov. 2001. Comparative study and prediction of DNA fragments associated with various elements of the nuclear matrix. *Biochim. Biophys. Acta* **1517**:351–364.
- Goodman, R. H., and S. Smolik. 2000. CBP/p300 in cell growth, transformation, and development. *Genes Dev.* **14**:1553–1577.
- Jeang, K.-T., D. Derse, M. Matocha, and O. Sharma. 1997. Expression status of Tax protein in human T-cell leukemia virus type 1-transformed MT4 cells: recall of MT4 cells distributed by the NIH AIDS Research and Reference Reagent Program. *J. Virol.* **71**:6277–6278.
- Jendraschak, E., and W. E. Kaminski. 1998. Isolation of human promoter regions by Alu repeat consensus-based polymerase chain reaction. *Genomics* **50**:53–60.
- Jenuwein, T., W. C. Forrester, L. A. Fernandez-Herrero, G. Laible, M. Dull, and R. Grosschedl. 1997. Extension of chromatin accessibility by nuclear matrix attachment regions. *Nature* **385**:269–272.
- Jeon, S., B. L. Allen-Hoffmann, and P. F. Lambert. 1995. Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *J. Virol.* **69**:2989–2997.
- Kalos, M., and R. E. Fournier. 1995. Position-independent transgene expression mediated by boundary elements from the apolipoprotein B chromatin domain. *Mol. Cell. Biol.* **15**:198–207.
- Konopka, A. K. 1988. Compilation of DNA strand exchange sites for non-homologous recombination in somatic cells. *Nucleic Acids Res.* **16**:1739–1758.
- Levy-Wilson, B., and C. Fortier. 1989. The limits of the DNase I-sensitive domain of the human apolipoprotein B gene coincide with the locations of chromosomal anchorage loops and define the 5' and 3' boundaries of the gene. *J. Biol. Chem.* **264**:21196–21204.
- Matsukura, T., S. Koi, and M. Sugase. 1989. Both episomal and integrated forms of human papillomavirus type 16 are involved in invasive cervical cancers. *Virology* **172**:63–72.
- Meissner, J. D. 1999. Nucleotide sequences and further characterization of human papillomavirus DNA present in the CaSki, SiHa and HeLa cervical carcinoma cell lines. *J. Gen. Virol.* **80**:1725–1733.
- Merika, M., and D. Thanos. 2001. Enhanceosomes. *Curr. Opin. Genet. Dev.* **11**:205–208.
- Neff, T., F. Shotkoski, and G. Stamatoyannopoulos. 1997. Stem cell gene therapy, position effects and chromatin insulators. *Stem Cells* **15**:265–271.
- Park, J. S., E. S. Hwang, S. N. Park, H. K. Ahn, S. J. Um, C. J. Kim, S. J. Kim, and S. E. Namkoong. 1997. Physical status and expression of HPV genes in cervical cancers. *Gynecol. Oncol.* **65**:121–129.
- Paulson, J. R., and U. K. Laemmli. 1977. The structure of histone-depleted metaphase chromosomes. *Cell* **12**:817–828.
- Quade, K., J. Saldanha, H. Thomas, and J. Monjardino. 1992. Integration of hepatitis B virus DNA through a mutational hot spot within the cohesive region in a case of hepatocellular carcinoma. *J. Gen. Virol.* **73**:179–182.
- Rogozin, I. B., G. V. Glazko, and M. V. Glazkov. 2000. Computer prediction of sites associated with various elements of the nuclear matrix. *Brief Bioinform.* **1**:33–44.
- Rollini, P., S. J. Namciu, M. D. Marsden, and R. E. Fournier. 1999. Identification and characterization of nuclear matrix-attachment regions in the human serpin gene cluster at 14q32.1. *Nucleic Acids Res.* **27**:3779–3791.
- Romanczuk, H., and P. M. Howley. 1992. Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. *Proc. Natl. Acad. Sci. USA* **89**:3159–3163.
- Romig, H., J. Ruff, F. O. Fackelmayr, M. S. Patil, and A. Richter. 1994. Characterisation of two intronic nuclear-matrix-attachment regions in the human DNA topoisomerase I gene. *Eur. J. Biochem.* **221**:411–419.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Sastre-Garau, X., M. Favre, J. Couturier, and G. Orth. 2000. Distinct patterns of alteration of myc genes associated with integration of human pap-

- illomavirus type 16 or type 45 DNA in two genital tumours. *J. Gen. Virol.* **81**:1983–1993.
46. Schwarz, E., U. K. Freese, L. Gissmann, W. Mayer, B. Roggenbuck, A. Stremlau, and H. zur Hausen. 1985. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature* **314**:111–114.
 47. Shah, K., and P. M. Howley. 1996. Papillomaviruses, p. 2077–2109. *In* B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), *Fields virology*, 3rd ed. Lip-pincott-Raven Publishers, Philadelphia, Pa.
 48. Singh, G. B., J. A. Kramer, and S. A. Krawetz. 1997. Mathematical model to predict regions of chromatin attachment to the nuclear matrix. *Nucleic Acids Res.* **25**:1419–1425.
 49. Stief, A., D. M. Winter, W. H. Stratling, and A. E. Sippel. 1989. A nuclear DNA attachment element mediates elevated and position-independent gene activity. *Nature* **341**:343–345.
 50. Stoler, M. H., C. R. Rhodes, A. Whitbeck, S. M. Wolinsky, L. T. Chow, and T. R. Broker. 1992. Human papillomavirus type 16 and 18 gene expression in cervical neoplasias. *Hum. Pathol.* **23**:117–128.
 51. Thorland, E. C., S. L. Myers, D. H. Persing, G. Sarkar, R. M. McGovern, B. S. Gostout, and D. I. Smith. 2000. Human papillomavirus type 16 inte-grations in cervical tumors frequently occur in common fragile sites. *Cancer Res.* **60**:5916–5921.
 52. Wang, J., X. Chenivresse, B. Henglein, and C. Brechot. 1990. Hepatitis B virus integration in a cyclin A gene in a hepatocellular carcinoma. *Nature* **343**:555–557.
 53. Wang, J., F. Zindy, X. Chenivresse, E. Lamas, B. Henglein, and C. Brechot. 1992. Modification of cyclin A expression by hepatitis B virus DNA integra-tion in a hepatocellular carcinoma. *Oncogene* **7**:1653–1656.
 54. Whitehurst, C., H. R. Henney, E. E. Max, H. W. Schroeder, Jr., F. Stuber, K. A. Siminovitch, and W. T. Garrard. 1992. Nucleotide sequence of the intron of the germline human kappa immunoglobulin gene connecting the J and C regions reveals a matrix association region (MAR) next to the en-hancer. *Nucleic Acids Res.* **20**:4929–4930.
 55. Yie, J., K. Senger, and D. Thanos. 1999. Mechanism by which the IFN-beta enhanceosome activates transcription. *Proc. Natl. Acad. Sci. USA* **96**:13108–13113.
 56. zur Hausen, H. 1991. Human papillomaviruses in the pathogenesis of ano-genital cancer. *Virology* **184**:9–13.