### Endogenous Retroviral Elements in Human DNA<sup>1</sup>

C. Leib-Mösch, 2 R. Brack-Werner, T. Werner, M. Bachmann, O. Faff, V. Erfle, and R. Hehlmann

III. Medizinische Klinik Mannheim der Universität Heidelberg, D-6800 Mannheim [C. L-M., M. B., O. F., R. H.]; and GSF-Abteilung für Molekulare Zellpathologie [R. B-W., V. E.] and GSF-Institut für Säugetiergenetik [T. W.], D-8042 Neuherberg, Federal Republic of Germany

#### **Abstract**

Endogenous retroviruses and retroviral elements represent a substantial component of vertebrate genomes. They are inherited as stable Mendelian genes and may be activated spontaneously or by physical or chemical agents. In the human genome various retroviral elements have been detected by their relationship with mammalian endogenous and exogenous retroviruses. The structure of these elements resembles either full-length or truncated proviruses. The biological function of human retrovirus-related sequences is still unknown, but like other transposable elements, they may have contributed in shaping the eukaryotic genome. Furthermore, they exhibit a number of features giving them a potential for involvement in carcinogenesis. Expression of endogenous retroviral elements has been detected in various human tissues and cell lines and in some cases appears to be associated with human neoplasias.

#### Introduction

A substantial portion of the eukaryotic genome is thought to have been shaped by reintegration of products of reverse transcription (retroposition) (1). The information generated in this manner includes both sequence elements, which seem to have used cellular mechanisms for passive retroposition, as well as retroelements containing reverse transcriptase- and/or integrase-related sequences possibly initiating their own retrotransposition. Members of the former group are called retrosequences (2) and include SINES<sup>3</sup> and processed pseudogenes (3). Among the retroelements which in themselves may have the capacity to transpose are nonviral elements such as LINES (4) as well as retrovirus-like elements with structural analogies to infectious retroviruses. Indeed, transposition of LINES has been shown to take place in humans, causing disease by insertional mutation in at least one case (5, 6). Thus retroelements are implicated to be mobile genetic elements with a potential to act as causative agents of disease.

Retrovirus-like elements in the human genome are intriguing because they resemble infectious murine, primate, and human retroviruses frozen into a proviral state for what in many cases has been shown to be millions of years (7, 8). Generally, they seem to lack the extracellular phase characteristic of retroviruses, relying instead on the replicative machinery of the cell for their propagation. Hence, they are endogenous constituents of the human genome and consequently inherited as stable Mendelian genes. Since they make up at least 0.1 to 0.6% of

human DNA (9, 10), they contribute substantially to the architecture of the human genome.

The human retroviral elements analyzed so far show sequence similarities to C-type and to A-, B-, and D-type murine and primate retroviruses as well as human retroviruses. Since they cover the full range of structural variations detected for infectious retroviruses, they present a wealth of viral information in the human genome. In this paper, we will try to give an outline of the multiformity of human retrovirus-like elements in addition to discussing the biological implications of the presence of these retroelements in the human genome.

#### C-Type Retrovirus-related Sequences

Human C-type retrovirus-related elements (Table 1) were discovered by their homology to primate endogenous and exogenous retroviruses. In one of the first approaches, cloned African green monkey DNA containing sequences related to MuLV and BaEV was used to screen a human genomic library under low stringency hybridization conditions (11, 12). Two types of human endogenous retroviral elements were isolated in this manner. One type, represented in clone 4-1, consists of full-length proviral structures of 8.8 kilobases with retroviral LTRs. The other type, clone 51-1 and related sequences, contains 4.1 kilobases of gag-pol-related sequences bound by highly repetitive sequences composed of 72 to 76 base pair-imperfect repeats (13). In total, the full-length and the truncated sequences are each represented by approximately 35 to 50 copies per human haploid genome. The 4-1 and 51-1 elements have been found to be widely dispersed over the human genome by Southern analyses of DNA from somatic rodent × human hybrid cell lines (14). Clone NP-2, a full-length proviral sequence related to 4-1, was localized in two copies on the Y chromosome (15). Conservation of cellular flanking sequences suggests that the second copy of NP-2 results from gene duplication rather than from provirus insertion. Sequence analysis of the full-length clone 4-1 revealed several termination codons and frame shifts in the reading frames of all three retroviral genes, thereby precluding its expression in the form of infectious virus particles (16).

Another human C-type retroviral sequence (ERV1) was isolated by Bonner et al. (7) with the help of a fragment from a cloned chimpanzee retrovirus-like sequence homologous to the polymerase genes of BaEV and MoMuLV. Low stringency screening of a human genomic library with the same cloned chimpanzee fragment and a probe containing BaEV LTR led to the isolation of a full-length human endogenous provirus termed ERV3 (17). In addition, various BaEV-related clones were isolated by two other groups by probing a human genomic library with either BaEV LTR (18) or with a BaEV gag-pol fragment (19) under low stringency hybridization conditions. ERV1 is a truncated provirus of about 8 kilobases that lacks a 5' LTR. It occurs in only one copy per human genomic equivalent and was localized by hybridization of DNA from somatic cell hybrids (20) as well as in situ hybridization (21) to chromosome 18 q22/q23. ERV3 is also a single copy sequence and

<sup>&</sup>lt;sup>1</sup> Presented at the "XIVth Symposium of the International Association for Comparative Research on Leukemia and Related Diseases," October 8-12, 1989, Vail, CO.

<sup>&</sup>lt;sup>2</sup> To whom requests for reprints should be addressed, at GSF-Abteilung für Molekulare Zellpathologie, Ingolstädter Landstr. 1, D-8042 Neuherberg, Federal Republic of Germany.

<sup>&</sup>lt;sup>3</sup> The abbreviations used are: SINES, short interspersed nuclear sequences; LINES, long interspersed nuclear sequences; MuLV, murine leukemia virus; MoMuLV, Moloney MuLV; BaEV, baboon endogenous virus; LTR, long terminal repeat; ERV, endogenous retrovirus; SMRV, squirrel monkey retrovirus; SSV, simian sarcoma virus; SSAV, simian sarcoma-associated virus; RFLP, restriction fragment length polymorphism; CHIMP, endogenous chimpanzee retrovirus-like element; AKV, endogenous ecotropic retrovirus of the AKR mouse; MMTV, mouse mammary tumor virus; IAP, intracysternal A-type particle; HERV, human endogenous retrovirus; HTLV, human T-cell lymphotropic virus; BLV, bovine leukemia virus; EBV, Epstein Barr virus; snRNP, small nuclear ribonucleoprotein; cDNA, complementary DNA.

Table 1 C-type-related human endogenous retroviral elements

Prototype	Length (kilobases)	Genomic structure	Chromosomal localization	Refs.
4-1	8.8	Full-length provirus		11-14, 16, 47-49
NP-2	8.8	Full-length provirus	Y	15
51-1	4.4	gag-pol		11-14
ERVI	8	gag-pol-env-LTR	18q22-23	7, 20, 21
ERV3	9.9	Full-length provirus	7	17, 50, 51, 84
S71	5.5	gag-SNRS-pol-LTR	18q21	24-28
HuRRS-P	8.1	Provirus?	•	22, 23

was mapped to chromosome 7 (17). Partial sequence analysis of ERV3 and ERV1 revealed that both clones contain termination codons within the *pol* and *gag* genes. The *env* gene of ERV3, however, contains a long open reading frame that is capable of encoding a polypeptide of approximately 650 amino acids.

A further family of probably C-type-related retroviral elements comprising about 20 to 40 copies in human DNA was detected by virtue of the observation that several retroviruses use tRNA<sup>Pro</sup> molecules as primers for reverse transcription. Screening a human genomic library with either a mouse tRNA<sup>Pro</sup> (22) or a synthetic oligonucleotide complementary to a retroviral tRNA<sup>Pro</sup> primer-binding site (23) yielded various retroviral sequences of 5 to 9 kilobases. Nucleotide sequencing of DNA sections upstream of the putative primer binding site showed all characteristic features of C-type retroviral LTRs.

We have found human DNA to contain a number of endogenous sequences related to SSV and SSAV. Under relaxed hybridization conditions, multiple distinct bands are obtained not only when probing human genomic DNA with the total SSAV genome but also with subgenomic fragments derived from the gag and pol genes of SSAV. Upon screening a human genomic library with a probe containing the complete SSAV genome as well as with probes derived from various regions of the SSAV genome under low stringency conditions, we isolated one strongly hybridizing clone termed S71 (24). screening of the human library with S71 under high stringency conditions yielded exclusively clones that contained the S71 retroviral element but no other SSAV-related sequences, indicating that human SSAV-related sequences are less similar to each other than the members of the 4-1/51-1 family. Low stringency hybridization of human DNA with specific S71 probes derived from the gag and pol region, however, revealed that the human genome contains at least 15 to 20 copies of retroviral elements related to S71/SSAV.

S71 was localized to chromosome 18, Band q21, by Southern blot analysis of DNA from rodent × human hybrid cell lines as well as in situ hybridization (25). Thus, the long arm of chromosome 18 carries two proviral sequences, S71 and ERV1. S71 is located in a region of high biological significance, since Band q21 of chromosome 18 also contains the yes-1 oncogene and the bcl-2 breakpoint cluster region of t(14;18) translocations observed in follicular and diffuse B-cell lymphomas. An S71 pol-LTR-containing probe detects RFLPs for two enzymes (EcoRI and BglII) in various human DNAs (26). Tracing the inheritance of S71 within two families by virtue of the EcoRI RFLP demonstrated that S71 is passed on as a stable Mendelian gene, thereby confirming the endogenous nature of this retroviral sequence in the human genome (27).

S71 is an incomplete provirus of 5.5 kilobases with C-type retrovirus-related gag and pol regions and a 3' LTR-like sequence (Fig. 1). The nucleotide sequence of S71 was determined and compared with the corresponding regions of SSV/SSAV (25, 28). The S71 gag sequence consists of matrix protein p15-,

p12-, capsid protein p30- and nucleocapsid protein p10-related sections and aligns in a largely collinear manner with all four protein-coding regions of the SSV gag gene. Although synthesis of a complete gag precursor protein is precluded by frame shifts and termination codons, there is an open reading frame comprising the complete p15 and part of the p12 region. The deduced amino acid sequence of the S71 gag region shows an overall identity of about 40% with the SSV gag gene.

The pol-related region in S71 corresponds to the 3' half of retroviral pol genes beginning in the tether region normally located 3' of the reverse transcriptase-encoding section and extending through the endonuclease region. The S71 pol-related sequences also align with the corresponding region of the SSAV pol gene in a collinear manner. A comparison of the deduced amino acid sequence with those of the human retroviral element 4-1 and mammalian C-type retroviruses is shown in Table 2. The reading frame of S71 pol sequences is, like that of 4-1, interrupted by several termination codons. Furthermore, the S71 element is lacking over half the pol gene, emcompassing the protease-coding region and the complete reverse transcriptase domain. In a biological sense, expression of sequences enabling random reverse flow of genetic information from RNA to DNA would pose a great threat for the evolutionary stability of the human genome. A prerequisite for the maintenance of such sequences in the human genome is therefore a very rigid control mechanism precluding their random expression. The numerous termination codons, frame shifts, and deletions observed in the pol sequences of C-type-related human endogenous retroviral elements may be a significant factor contributing to this stringent control.

Sequence comparison of S71 with SSV/SSAV indicates a higher degree of conservation on the amino acid level than on

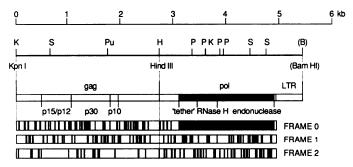


Fig. 1. Genomic organization of the C-type-related human endogenous retroviral element S71. kb, kilobases. Abbreviations: B, BamHI; H, HindIII; K, KpnI; P, PstI; Pu, PvuII; S, SacI.

Table 2 Amino acid identities of the tether, RNase H, and endonuclease region of S71 pol with other human and mammalian retroviral pol genes

	Tether (%)	RNase H (%)	Endonuclease (%)
AKV	49	53	58
4-1	45	44	40
SSV			52
SSAV	52	49	52

the nucleotide level. Furthermore, the distribution of nucleotide substitutions among the three positions of the amino acid codons is clearly biased toward silent third position changes (28). This bias suggests functional constraints effective at the protein level, indicating transient functionality of these proteins through evolution. The gag and pol region of S71 is separated by a 1130-base pair nonretroviral sequence. The predicted amino acid sequence suggests a possible open reading frame of 121 amino acids immediately 3' of the p10<sup>gag</sup> region.

A 535-base pair region with all structural features of a retroviral LTR, including potential signal sequences essential for transcriptional control, constitutes the 3' terminus of the S71 element (25). Comparison of the S71 sequence with the aligned nucleotide sequence of 11 retroviral LTRs, 7 of which were derived from human endogenous retroviral elements and 4 from infectious proviruses, demonstrated a common sequence motif, all or part of which is reflected in 6 of the 7 human endogenous LTRs analyzed. In the S71 LTR-like sequence this motif contains a 9-base pair region with eight matches to the enhancer core consensus sequence present in a number of viral enhancers (29). Sequences with potential enhancer function, such as this common motif or direct repeats, two of which are also contained in the S71 LTR-like sequence, may enable human endogenous LTRs to influence the expression of adjacent cellular genes in cis.

The genomic organization of S71 shows interesting structural analogies with the replication-defective, transforming SSV provirus. Both elements contain a complete gag gene, but are missing a large part of the polymerase gene, including the entire reverse transcriptase domain and the protease sequences. Both elements are also lacking most or all of the envelope gene. Finally, in both genomes part of a retroviral gene has been replaced by nonretroviral sequences. In SSV part of the envelope gene is replaced by the sis oncogene which is essential for the SSV-transforming activity. In S71, part of the pol gene is replaced by a presumably nonretroviral sequence (Fig. 1). These structural similarities imply that analogous mechanisms may have been involved in the generation of both elements.

We used conserved regions within the p30<sup>peag</sup> and the endonuclease region of the pol gene for phylogenetic analysis of various proviruses and retroviral elements (28). Fig. 2 shows a composite phylogenetic tree based on separate alignments of gag- and pol-derived sequences. In the case of ERV1 and the endogenous CHIMP sequence, data were only available for the endonuclease part of the pol gene. Since defective endogenous sequences do not face continuous functional selection, a com-

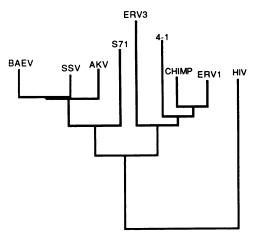


Fig. 2. Phylogenetic relationship of primate and murine retroviral sequences.

mon rate of amino acid changes with active viral sequences cannot be assumed. Therefore the branch length in Fig. 2 may not correlate with the real evolutionary distances, but should not interfere with the overall topology of the tree. Our data suggest that 4-1, ERV3, ERV1, and CHIMP are closely related, whereas S71 clusters with the infectious mammalian C-type retroviruses AKV, BaEV, and SSV/SSAV in a separate subgroup.

#### **B-Type Retrovirus-related Sequences**

Besides C-type-related proviruses, the human genome contains a large number of retroviral elements showing homology to the B-type MMTV, as well as to the Syrian hamster IAP and to the D-type SMRV (Table 3). B- and D-type retroviruses and A-type particles are assumed to originate from a common progenitor on the basis of homologous *pol* sequences that differ from those of mammalian C-type retroviruses (30, 31). Members of the group of B-type-related human retroviral elements were isolated by low stringency hybridization with DNA probes encompassing either various regions of the MMTV genome (32–35) or by using a probe from the polymerase gene of the Syrian hamster IAP (36).

Franklin et al. (37) analyzed 100 recombinant clones which they had isolated from the DNA of a human breast cancer cell line by screening with an MMTV gag-pol fragment. By crosshybridization experiments with subcloned fragments from some of these isolates they identified nine distinct subgroups of MMTV-related sequences among these clones. Although all subgroups hybridize with the MMTV gag-pol probe, they are not closely related to each other. The largest subgroup, comprising 64% of the isolated clones, was found to contain sequences most homologous to MMTV DNA. This group consists of retroviral genomes of 6 to 10 kilobases, including the well-characterized clones HLM-2 (38), NMWV4 (34), HM16 (33), and HERV-K10 (39), and is estimated to comprise about 50 copies per human haploid genome. Members of this group were mapped to human chromosomes 1 (HLM-2) and 5 (HLM-25) and chromosomes 7, 8, 11, 14, and 17 (40).

In the case of HERV-K10, the complete nucleotide sequence was determined (36, 39). HERV-K10 is a full-length provirus 9.2 kilobases in length with LTRs of 968 base pairs at both ends. Thus of all known retroviral LTRs, HERV-K LTRs rank second in length after the LTRs of MMTV. The pol region of HERV-K10 is closely related to that of A- and D-type retroviruses and especially to the B-type pol gene. Interestingly, HERV-K10 is until now the only human retroviral genome that contains an open reading frame large enough to allow synthesis of full-length polymerase proteins including reverse transcriptase. The env gene of HERV-K10 shows unusually high similarity with the MMTV env region, even on the level of secondary structure of the predicted amino acid sequences, including potential glycosylation sites (41).

MMTV-related endogenous sequences were also detected in the genomes of great apes and Old World monkeys, but not in New World monkeys and prosimians (8). This suggests that progenitor viruses for MMTV-related human retroviral elements must have entered the genomes of Old World anthropoids after their divergence from New World monkeys and prior to branching off from the ancestors of the large hominoids.

#### **HTLV-related Sequences**

A third group of retroviral elements is distinguished by sequence relationship with human T-cell lymphotropic viruses

Table 3 B-type-related human endogenous retroviral elements

Prototype	Length (kilobases)	Genomic structure	Chromosomal localization	Refs.
HLM-2	9	Full-length provirus	1	8, 32, 38, 40
HLM-25	9	Full-length provirus	5	40
HM16	6-8	LTR-(gag?)-pol-LTR		33
NMWV4	6–7	LTR-gag-pol-LTR		34, 35, 37, 76, 85
HERV-K10	9.2	Full-length provirus		36, 39, 41

Table 4 HTLV-related human endogenous retroviral elements

Prototype	Length (kilobases)	Genomic structure	Refs.	
RTVL-H2	5, 5-6	LTR-gag-pol-LTR	42, 43	
HRES-1/1	ND <sup>a</sup>	LTR-gag	46	

a ND, not determined.

HTLV-I and -II (Table 4). A multicopy family of endogenous retroviral elements termed RTVL-H was detected fortuitously by analysis of the  $\beta$ -globin gene cluster (42). The RTVL-H sequences are present in approximately 1000 copies per human haploid genome. Nucleotide sequence analysis of one member of this family (RTVL-H2) revealed a *pol*-related region homologous to parts of C-type retrovirus *pol* genes (43). A segment of the *gag* region of RTVL-H2 shows 55 to 60% amino acid identity with a 50-amino acid region of the *gag* nucleic acid binding proteins encoded by HTLV-I and -II and BLV.

A recently developed strategy for the isolation of retrovirus-related sequences applies synthetic oligonucleotides, homologous to highly conserved regions of retroviral pol genes, to the polymerase chain reaction technique (44, 45). A wide spectrum of retrovirus-related sequences, including B- and C-type-related elements and KpnI (LINE 1) family repeats, was amplified using mixed oligonucleotides reflecting consensus sequences of numerous reverse transcriptase genes or unique oligonucleotides whose sequences were derived from the polymerase genes of HTLV-I and -II. Some of the HTLV-related sequences isolated by this approach showed a hybrid-like character with sequence homologies to HTLV-I as well as to Mason-Pfitzer monkey virus (45).

An HTLV-I-related retroviral element termed HRES-1/1 was isolated by low stringency screening of a human genomic library with an HTLV-I LTR-gag probe (46). Nucleotide sequence analysis of approximately 2 kilobases revealed a single 5' LTR-like sequence and a gag-related sequence containing two potential open reading frames encoding a  $M_r$  25,000 and 15,000 protein. The  $M_r$  25,000 protein of HRES-1/1 shows 32% and 39% amino acid identity with HTLV-I and HTLV-II p19, respectively.

#### **Expression of Human Endogenous Retroviral Elements**

Although all human endogenous retroviral elements examined so far seem to be replication defective, some of them have been shown to be transcriptionally active in human tissues and cell lines. Several discrete mRNA species hybridizing to LTR and env DNA probes derived from the 4-1 element were detected in human placenta, spleen, normal colon mucosa, and primary colon cancers as well as in colon cancer cell lines (SW1116, HCT, Caco2), in a breast carcinoma cell line (T47D), and in a T-cell acute lymphocytic leukemia cell line (8402) (47–49). In colon tumors an increase of 1.7- and 3.0-kilobase LTR-env-related transcripts was observed compared with normal colon tissue, whereas a 3.6-kilobase transcript abundant in normal colon mucosa was decreased in tumor cells (47). Partial cDNA clones of 4-1 env-related mRNA transcripts were iso-

lated from human placenta. Sequence analysis of two placental cDNA clones, however, revealed in-frame termination codons so that neither of them could encode full-length *env* proteins (49).

The env region of another C-type-related full-length retroviral element, ERV3, contains an open reading frame of approximately 650 amino acids. This potential polypeptide exhibits characteristic features of a retroviral glycoprotein, including several potential glycosylation sites and sequences typical for transmembrane proteins (50). An ERV3 env-specific cDNA of 2.85 kilobases was isolated from a human fetal cDNA library and found to be identical to parts of ERV3 by DNA sequence analysis. Three polyadenylated RNAs of 9, 7.3, and 3.5 kilobases were identified in human placental chorion and characterized by Northern blotting and S1 nuclease mapping (51). The RNAs were found to be spliced mRNAs lacking the gag and most of the pol gene. The two larger mRNAs extended through the polyadenylation site in the 3' LTR and contained adjacent cellular sequences.

Screening cDNA libraries from human placenta, a human osteosarcoma cell line, and a phorbolmyristate acetate-stimulated B-cell line with various S71 probes under low stringency hybridization conditions yielded several strongly hybridizing clones. Preliminary sequencing data indicate that at least some of the isolated S71-related clones contain sequences corresponding to the nonretroviral sequence detected in S71.<sup>4</sup>

The MMTV-related human proviral sequences HERV-K was found to be expressed as an 8.8-kilobase full-length mRNA transcipt in cell lines from breast carcinoma (T47D), gastric carcinoma (Kato-III), malignant melanoma (HMT-2), and epidermoid carcinoma (HEp-2, Hela). Stimulation of HERV-K expression was observed in steroid-treated T47D cells (41). Hybridization of RNA from five breast cancer cell lines, HeLA and A431 cells, and human placenta with probes from various MMTV-related endogenous elements yielded mRNA transcripts ranging from 1.2 to 12 kilobases in size (37).

Expression of the HTLV-related HRES-1/1 sequence was shown in MA-T cells, melanoma cells, HL-60 promyelocytic leukemia cells, MOLT-4 T-cell leukemia cells, normal placenta, and EBV-transformed normal human peripheral blood B-lymphocytes (46). The 6-kilobase transcripts hybridized with an HRES-1/1 probe containing the LTR-gag-related sequences.

In spite of abundant transcription of endogenous retroviral sequences in various cells, the corresponding proteins have not yet been identified. However, detection of retrovirus-related antigens in human tissue and sera provides evidence for expression of human endogenous retroviral sequences at the protein level. We have previously reported that antibodies against the structural protein p30geag of SSV/SSAV and BaEV recognize proteins in human leukemic sera (52). Furthermore, serum proteins related to the SSAV gp70 protein seem to be of diagnostic value for the prognosis of patients with acute leukemias or chronic myelogenous leukemia in blast crisis (53). Antigens cross-reacting with antibodies against various retro-

<sup>&</sup>lt;sup>4</sup> M. Bachmann et al., manuscript in preparation.

viral proteins have also been detected in human leukemic cells, placenta, and choriocarcinoma (54-57). Antibodies that were raised against a synthetic undecapeptide derived from the gagrelated region of ERV1 identified a  $M_r$  75,000 protein in renal adenocarcinoma, placenta, and trophoblastic tumors (58, 59).

Particles with retrovirus-like morphology have been visualized by electron microscopy in various human tissues and cell lines, many of which are of neoplastic origin (60-63). The release of particles from clonal derivatives of the T47D human breast carcinoma cell line was found to be steroid dependent (61). To examine whether there is a connection between production of retrovirus-like particles and expression of the MMTV-related endogenous sequence HERV-K in steroidtreated T47D cells (41), we have constructed synthetic peptides from the putative outer membrane region of the env gene of HERV-K10. At least three of the peptides were found to react with polyclonal and monoclonal antibodies raised against MMTV and/or with antibodies against the virus-like particles produced by the T47D cell line. In addition, polyclonal antipeptide sera show an immunological cross-reaction with T47D particles. These data strongly suggest an immunological relationship between HERV-K10 env gene products and the viruslike particles produced by the human breast cancer cell line T47D.5 Although neither HERV-K10 nor any other known human endogenous retroviral element seems to be able to produce infectious retroviral particles, as a consequence of inframe termination codons, the formation of pseudotypes might be possible by complementation of structural components derived from different retrovirus-like sequences.

#### Discussion

Endogenous retroviruses and retroviral elements have been detected in the DNA of many vertebrate species including primates. As a rule, they persist as silent retroviral copies in their host cell genome and are transmitted through the germ line. Human endogenous retroviral sequences resemble in structure either full-length or truncated proviruses. Although their genomes have generally been found to be defective, they represent a reservoir of viral genes which may be activated spontaneously, by recombination events or by radiation and chemical agents. Their biological function is still unknown, but, like other transposable elements, they may have played a role in shaping the eukaryotic genome by intracellular transposition events. Furthermore, they might be involved in physiological processes such as protection against superinfection by related exogenous retroviruses.

The pathogenic potential of endogenous retroviruses and retroviral elements has been shown for murine leukemia viruses (64-68), mouse mammary tumor viruses (69-71), and intracysternal A-type particles (72-74). Once activated, they can act as mutagens and interfere with normal cellular functions. Expression of cellular genes can be influenced in *cis* by transcription control elements present in retroviral LTRs (75). Proviral integration causing activation of an adjacent cellular protooncogene by promoter insertion has been reported for IAPs in mouse plasmacytoma, for example (72). Analysis of genomic DNA from two human breast cancer cell lines revealed an additional band that hybridized to the human endogenous MMTV-related element NMW1E (76). This fragment is not found in DNA from normal human blood cells and may represent the recent integration of MMTV-related DNA. A tumor-specific c-myc

rearrangement caused by insertion of a human LINE 1 element into the second intron of the c-myc locus was observed in a human breast carcinoma (6). Insertion mutagenesis can also lead to inactivation of cellular genes. Transposition of human LINE 1 elements was found to have inactivated the Factor VIII gene in two cases of Haemophilia A (5).

The high copy number and dispersion of closely related endogenous retroviral elements to various human chromosomes raise the possibility of an involvement of these sequences in chromosomal aberrations. Homologous sequences on different chromosomes could be sites for recombination events leading to genetic rearrangements as has been demonstrated for *Alu* repeats (77). Some evidence for this hypothesis is provided by the observation that one member of the multicopy family of RTVL-H elements is located close to the breakpoint of three naturally occurring deletions in the  $\beta$ -globin gene cluster (43). Furthermore, the retroviral sequence S71 was found to map to the same band as the major breakpoint region of t(14;18) chromosomal translocations in B-cell lymphomas. The same chromosomal region is also involved in deletions frequently occurring in colon carcinomas (78).

A further level of potential pathogenic activity may be the expression of retroviral proteins. Even replication-defective proviruses can give rise to products such as the p15E envelope proteins which have been shown to possess immunosuppressive and antiinflammatory activity (79). A possible role for endogenous retroviral sequences has also been suggested in the initiation of autoimmune diseases. Mammalian C-type retroviral gag proteins have been shown to share an antigenic determinant with the U1 snRNP-associated  $M_r$  70,000 protein (80). The activation of endogenous retroviral sequences followed by the expression of retroviral gag proteins may therefore lead to the production of autoantibodies against U1 snRNPs and the initiation of autoimmunity. The presence of retrovirus-related antigens and antibodies cross-reacting with various retroviral proteins in sera from patients with autoimmune diseases suggests a possible involvement of gene products expressed by human endogenous retroviral sequences (81-83). Although there is growing evidence that human endogenous retroviral elements cannot simply be considered as silent genes that have lost their biological relevance during evolution, their actual contribution in biological processes and their possible involvement in human neoplasia demand further investigation.

#### References

- Weiner, A. M., Deininger, P. L., and Efstratiadis, A. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem., 55: 631-661, 1986.
- Temin, H. M. Reverse transcriptases: retrons in bacteria. Nature (Lond.), 339: 254-255, 1989.
- Deininger, P. L. SINEs: short interspersed repeated DNA elements in higher eucaryotes. *In:* D. E. Berg and M. M. Howe (eds.), Mobile DNA, pp. 619– 636. Washington, DC: American Society of Microbiology, 1989.
- Hutchison, C. A., III, Hardies, S. C., Loeb, D. D., Shehee, W. R., and Edgell, M. H. Lines and related retroposons: long interspersed repeated sequences in the eucaryotic genome. *In:* D. E. Berg and M. M. Howe (eds.), Mobile DNA, pp. 593-617. Washington, DC: American Society of Microbiology, 1989.
- Kazazian, H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G., and Antonarakis, S. E. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature (Lond.), 332: 164-166, 1988.
- Morse, B., Rotherg, P. G., South, V. J., Spandorfer, J. M., and Astrin, S. M. Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. Nature (Lond.), 333: 87-90, 1988.
- 7. Bonner, T. I., O'Connell, C., and Cohen, M. Cloned endogenous retroviral sequences from human DNA. Proc. Natl. Acad. Sci. USA, 79: 4709-4713,
- 8. Mariani-Constantini, R., Horn, T. M., and Callahan, R. Ancestry of a human

<sup>&</sup>lt;sup>5</sup> O. Faff et al., manuscript in preparation.

- endogenous retrovirus family. J. Virol., 63: 4982-4985, 1989.
- Brack-Werner, R., Leib-Mösch, C., Werner, T., Erfle, V., and Hehlmann, R. Human endogenous retrovirus-like sequences. *In:* R. Neth (ed.), Modern Trends in Human Leukemia VIII, pp. 464-477. Berlin/Heidelberg/New York: Springer-Verlag, 1989.
- Callahan, R. Two families of human endogenous retroviral genomes. Eukaryotic transposable elements as mutagenic agents. Banbury Rep., 30: 91– 100, 1988.
- Martin, M. A., Bryan, T., Rasheed, S., and Khan, A. S. Identification and cloning of endogenous retroviral sequences present in human DNA. Proc. Natl. Acad. Sci. USA, 78: 4892-4896, 1981.
- Repaske, R., O'Neill, R. R., Steele, P. E., and Martin, M. A. Characterization and partial nucleotide sequence of endogenous type C retrovirus segments in human chromosomal DNA. Proc. Natl. Acad. Sci. USA, 80: 678-682, 1983.
- Steele, P. E., Rabson, A. B., Bryan, T., and Martin, M. A. Distinctive termini characterize two families of human endogenous retroviral sequences. Science (Wash. DC), 225: 943-947, 1984.
- Steele, P. E., Martin, M. A., Rabson, A. B., Bryan, T., and O'Brien, S. J. Amplification and chromosomal dispersion of human endogenous retroviral sequences. J. Virol., 59: 545-550, 1986.
- Silver, J., Rabson, A., Bryan, T., Willey, R., and Martin, M. A. Human retroviral sequences on the Y chromosome. Mol. Cell Biol., 7: 1559-1562, 1987
- Repaske, R., Steele, P. E. O'Neill, R. R., Rabson, A. B., and Martin, M. A. Nucleotide sequence of a full-length human endogenous retroviral segment. J. Virol., 54: 764-772, 1985.
- O'Connell, C., O'Brien, S., Nash, W. G., and Cohen, M. ERV3, a full-length human endogenous provirus: chromosomal localization and evolutionary relationships. Virology, 138: 225-235, 1984.
- Noda, M., Kurihara, M., and Takano, T. Retrovirus-related sequences in human DNA: detection and cloning of sequences which hybridize with the long terminal repeat of baboon endogenous virus. Nucleic Acids Res., 10: 2865-2878, 1982.
- McClain, K., and Wilkowski, C. Activation of endogenous retroviral sequences in human leukemia. Biochem. Biophys. Res. Commun., 133: 945-950, 1985.
- O'Brien, S. J., Bonner, T. I., Cohen, M., O'Connell, C., and Nash, W. G. Mapping of an endogenous retroviral sequence to human chromosome 18. Nature (Lond.), 303: 74-77, 1983.
- Renan, M. J., and Reeves, B. R. Chromosomal localization of human endogenous retroviral element ERV1 to 18q22→q23 by in situ hybridization. Cytogenet. Cell Genet., 44: 167-170, 1987.
- Harada, F., Tsukada, N., and Kato, N. Isolation of three kinds of human endogenous retrovirus-like sequences using tRNA<sup>Pro</sup> as a probe. Nucleic Acids Res., 15: 9153-9162, 1987.
- Kröger, B., and Horak, I. Isolation of novel human retrovirus-like sequences by hybridization to synthetic oligonucleotides complementary to the tRNA<sup>pro</sup> primer-binding site. J. Virol., 61: 2071-2075, 1987.
- Leib-Mösch, C., Brack, R., Werner, T., Erfle, V., and Hehlmann, R. Isolation of an SSAV-related endogenous sequence from human DNA. Virology, 155: 666-677, 1986.
- Brack-Werner, R., Barton, D. E., Werner, T. Foellmer, B. E., Leib-Mösch, C., Francke, U., Erfle, V., and Hehlmann, R. Human SSAV-related endogenous retroviral element: LTR-like sequence and chromosomal localization to 18q21. Genomics, 4: 68-75, 1989.
- Leib-Mösch, C., Barton, D. E., Brack-Werner, R., Foellmer, B., Werner, T., Rohrmeier, D., Francke, U., Erfle, V., and Hehlmann, R. Genetic characterization of a human endogenous retroviral element located on chromosome 18q21. In: R. Neth (ed.), Modern Trends in Human Leukemia VIII, pp. 461-463. Berlin/Heidelberg/New York: Springer Verlag, 1989.
- Leib-Mösch, C., Barton, D., Geigl, E.-M., Brack-Werner, R., Erfle, V., Hehlmann, R., and Francke, U. Two RFLPs associated with the human endogenous retroviral element S71 on chromosome 18q21. Nucleic Acids Res., 17: 2367, 1989.
- 28. Werner, T., Brack-Werner, R., Leib-Mösch, C., Backhaus, H., Erfle, V., and Hehlmann, R. S71 is a phylogenetic distinct human endogenous retroviral element with structural and sequence homology to simian sarcoma virus (SSV). Virology, 155: 225-238, 1990.
- 29. Weiher, H., König, M., and Gruss, P. Multiple point mutations affecting the simian virus 40 enhancer. Science (Wash. DC), 219: 626-631, 1983.
- Chiu, I.-M., Callahan, R., Tronick, S. R., Schlom, J., and Aaronson, A. Major pol gene progenitors in the evolution of oncoviruses. Science (Wash. DC), 223: 364-370, 1984.
- Ono, M., Toh, H., Miyata, T., and Awaya, T. Nucleotide sequence of the Syrian hamster intracisternal A-particle gene: close evolutionary relationship of type A particle gene to types B and D oncovirus genes. J. Virol., 55: 387– 394, 1985.
- Callahan, R., Drohan, W., Tronick, S., and Schlom, J. Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. Proc. Natl. Acad. Sci. USA, 79: 5503-5507, 1982.
- Deen, K. C., and Sweet, R. W. Murine mammary tumor virus pol-related sequences in human DNA: chracterization and sequence comparison with the complete murine mammary tumor virus pol gene. J. Virol., 57: 422-432, 1986.
- 34. May, F. E. B., and Westley, B. R. Structure of a human retroviral sequence related to mouse mammary tumor virus. J. Virol., 60: 743-749, 1986.
- 35. Westley, B., and May, F. E. B. The human genome contains multiple

- sequences of varying homology to mouse mammary tumor virus DNA. Gene, 28:221-227,1984.
- Ono, M. Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. J. Virol., 58: 937-944, 1986.
- Franklin, G. C., Chretien, S. C., Hanson, I. M., Rochefort, H., May, F. E. B., and Westley, B. R. Expression of human sequences related to those of mouse mammary tumor virus. J. Virol., 62: 1203-1210, 1988.
- Callahan, R., Chiu, I. M., Wong, J. F., Tronick, S. R., Roe, B. A., and Aaronson, S. A. A new class of endogenous human retroviral genomes. Science (Wash. DC), 228: 1208-1211, 1985.
- 39. Ono, M., Yasunaga, T., Miyata, T., and Ushikubo, H. Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. *J. Virol.*, 60: 589-598, 1986.
- Horn, T. M., Huebner, K., Croce, C., and Callahan, R. Chromosomal locations of members of a family of novel endogenous human retroviral genomes. J. Virol., 58: 955-959, 1986.
- Ono, M., Kawakami, M., and Ushikubo, H. Stimulation of expression of the human endogenous retrovirus genome by female steroid hormones in human breast cancer cell line T47D. J. Virol., 61: 2059-2062, 1987.
- Mager, D. L., and Henthorn, P. S. Identification of a retrovirus-like repetitive element in human DNA. Proc. Natl. Acad. Sci. USA, 81: 7510-7514, 1984.
- Mager, D. L., and Freeman, J. D. Human endogenous retrovirus-like genome with type C pol sequences and gag sequences related to human T-cell lymphotropic viruses. J. Virol., 61: 4060-4066, 1987.
- Bangham, C. R. M., Daenke, S., Phillips, R. E., Cruickshank, J. K., and Bell, J. I. Enzymatic amplification of exogenous and endogenous retroviral sequences from DNA of patients with tropical spastic paraparesis. EMBO J., 7: 4179-4184, 1988.
- Shih, A., Misra, R., and Rush, M. G. Detection of multiple, novel reverse transcriptase coding sequences in human nucleic acids: relation to primate retroviruses. J. Virol., 63: 64-75, 1989.
- Perl, A., Rosenblatt, J. D., Chen, I. S. Y., DiVincenzo, J. P., Bever, R., Poiesz, B. J., and Abraham, G. N. Detection and cloning of new HTLVrelated endogenous sequences in man. Nucleic Acids Res., 17: 6841-6854, 1989.
- Gattoni-Celli, S., Kirsch, K., Kalled, S., and Isselbacher, K. J. Expression of type C-related endogenous retroviral sequences in human colon tumors and colon cancer cell lines. Proc. Natl. Acad. Sci. USA, 83: 6127-6131, 1986.
- Rabson, A. B., Steele, P. E., Garon, C. F., and Martin, M. A. mRNA transcripts related to full-length endogenous retroviral DNA in human cells. Nature (Lond.), 306: 604-607, 1983.
- Rabson, A. B., Hamagishi, Y., Steele, P., Tykocinske, M., and Martin, M. A. Characterization of human endogenous retroviral envelope RNA transcripts. J. Virol., 56: 176-182, 1985.
- Cohen, M., Powers, M., O'Connell, C., and Kato, N. The nucleotide sequence
  of the env gene from the human provirus ERV3 and isolation and characterization of an ERV3-specific cDNA. Virology, 147: 449-458, 1985.
- Kato, N., Pfeifer-Ohlsson, S., Kato, M., Larson, E., Rydnert, J., Ohlsson, R., and Cohen, M. Tissue-specific expression of human provirus ERV3 mRNA in human placenta: two of the three ERV3 mRNAs contain human cellular sequences. J. Virol., 61: 2182-2191, 1987.
- Hehlmann, R., Schetters, H., Erfle, V., and Leib-Mösch, C. Detection and biochemical characterization of antigens in human leukemic sera that crossreact with primate C-type viral proteins (M<sub>r</sub> 30,000). Cancer Res., 43: 392– 399, 1983.
- Hehlmann, R., Erfle, V., Schetters, H., Luz, A., Rohmer, H., Schreiber, M. A., Pralle, H., Essers, U., and Weber, W. Antigens and circulating immune complexes related to the primate retroviral glycoprotein SiSVgp70. Cancer (Phila.), 54: 2927-2935, 1984.
- 54. Derks, J. P. A., Hofmans, L., Bruning, H. W., and v. Rood, J. J. Synthesis of a viral protein with molecular weight of 30,000 (p30) by leukemic cells and antibodies cross-reacting with simian sarcoma virus p30 in serum of a chronic myeloid leukemia patient. Cancer Res., 42: 681-686, 1982.
- Jerabek, L. B., Mellors, R. C., Elkon, K. B., and Mellors, J. W. Detection and immunochemical characterization of a primate type C retrovirus-related p30 protein in normal human placentas. Proc. Natl. Acad. Sci. USA, 81: 6501-6505, 1984.
- Maeda, S., Yonezawa, K., Yoshizaki, H., Mori, M., Kobayashi, T., Akahonai, Y., Yachi, A., and Mellors, R. C. Leukemia serum reactive with retrovirusrelated antigen in normal human placenta. Int. J. Cancer, 38: 309-316, 1986.
- Suni, J., Närvänen, A., Wahlström, T., Lehtovirta, P., and Vaheri, A. Monoclonal antibody to human T-cell leukemia virus p19 defines polypeptide antigen in human choriocarcinoma cells and syncytiotrophoblasts of firsttrimester placentas. Int. J. Cancer, 33: 193-198, 1984.
- Suni, J., Närvänen, A., Wahlström, T., Aho, M., Pakkanen, R., Vaheri, A., Copeland, T., Cohen, M., and Oroszlan, S. Human placental syncytiotrophoblastic M, 75,000 polypeptide defined by antibodies to a synthetic peptide based on a cloned human endogenous retroviral DNA sequence. Proc. Natl. Acad. Sci. USA, 81: 6197-6201, 1984.
- Wahlström, T., Närvänen, A., Suni, J., Pakkanen, R., Lehtonen, T., Saksela, E., Vaheri, A., Copeland, T., Cohen, M., and Oroszlan, S. M. 75,000 protein, a tumor marker in renal adenocarcinoma, reacting with antibodies to a synthetic peptide based on a cloned human endogenous retroviral nucleotide sequence. Int. J. Cancer, 36: 379-382, 1985.
- Boller, K., Frank, H., Löwer, J., Löwer, R., and Kurth, R. Structural organization of unique retrovirus-like particles budding from human teratocarcinoma cell lines. J. Gen. Virol., 64: 2549-2559, 1983.

- Keydar, I., Ohno, T., Nayak, R., Sweet, R., Simoni, F., Weiss, F., Karby, S., Mesa-Tejada, R., and Spiegelman, S. Properties of retrovirus-like particles produced by a human breast carcinoma cell line: immunological relationship with mouse mammary tumor virus proteins. Proc. Natl. Acad. Sci. USA, 81: 4188-4192, 1984.
- Löwer, R., Löwer, J., Frank, H., Harzmann, R., and Kurth, R. Human teratocarcinomas cultured in vitro produce unique retrovirus-like viruses. J. Gen. Virol., 65: 887-898, 1984.
- Mondal, H., and Hofschneider, P. H. Isolation and characterization of retrovirus-like elements from normal human fetuses. Int. J. Cancer, 30: 281– 287, 1982.
- Chattopadhyay, S. K., Cloyd, M. W., Linemeyer, D. L., Lander, M. R., Rands, E., and Lowy, D. R. Cellular origin and role of mink cell focusforming viruses in murine thymic lymphomas. Nature (Lond.), 295: 25-31, 1982.
- Green, N., Hiroshi, H., Elder, J. H., Schwartz, R. S., Khiroya, R. H., Thomas, C. Y., Tsichlis, P. N., and Coffin, J. M. Expression of leukemogenic recombinant viruses associated with a recessive gene in HRS/J mice. J. Exp. Med., 152: 249-264, 1980.
- Hartley, J. W., Wolford, N. K., Old, L. J., and Rowe, W. P. A new class of murine leukemia virus associated with development of spontaneous lymphomas. Proc. Natl. Acad. Sci. USA, 74: 789-792, 1977.
- Leib-Mösch, C., Schmidt, J., Etzerodt, M., Pedersen, F. S., Hehlmann, R., and Erfle, V. Oncogenic retrovirus from spontaneous murine osteomas. Virology, 150: 96-105, 1986.
- Schmidt, J., Erfle, V., Pedersen, F. S., Rohmer, H., Schetters, H., Marquart, K.-H., and Luz, A. Oncogenic retrovirus from spontaneous murine osteomas.
   I. Isolation and biochemical characterization. J. Gen. Virol., 65: 2237-2248, 1984.
- Bentvelzen, P. Host-virus interactions in murine mammary carcinogenesis. Biochim. Biophys. Acta, 355: 236-259, 1974.
- Cohen, J. C., and Varmus, H. E. Proviruses of mouse mammary tumor virus in normal and neoplastic tissues from GR and C3Hf mouse strains. J. Virol., 35: 298-305, 1980.
- Etkind, P. R., and Sarkar, N. H. Integration of new endogenous mouse mammary tumor virus proviral DNA at common sites in the DNA of mammary tumors of C3Hf mice and hypomethylation of the endogenous mouse mammary tumor virus proviral DNA in C3Hf mammary tumors and spleens. J. Virol., 45: 114-123, 1983.
- Canaani, E., Dreazen, O., Klar, A., Rechavi, G., Ram, D., Cohen, J. B., and Givol, D. Activation of the c-mos oncogene in a mouse plasmacytoma by insertion of an endogenous intracisternal A-particle genome. Proc. Natl.

- Acad. Sci. USA. 80: 7118-7122, 1983.
- Hawley, R. G., Shulman, M. J., and Hozumi, N. Transposition of two different intracysternal A particle elements into an immunoglobulin kappachain gene. Mol. Cell. Biol. 4: 2565-2572, 1984.
- Horowitz, M., Luria, S., Rechavi, G., and Givol, D. Mechanism of activation
  of the mouse c-mos oncogene by the LTR of an intracysternal A-particle
  gene. EMBO J., 3: 2937-2941, 1984.
- Hayward, W. S., Neel, B. G., and Astrin, S. M. Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis. Nature (Lond.), 290: 475-480, 1981.
- May, F. E. B., and Westley, B. R. Characterization of sequences related to the mouse mammary tumor virus that are specific to MCF-7 breast cancer cells. Cancer Res., 49: 3879-3883, 1989.
- Rouyer, F., Simmler, M.-C., Page, D. C., and Weissenbach, J. A sex chromosome rearrangement in a human XX male caused by *Alu-Alu* recombination. Cell. 51: 417-425, 1987.
- Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M. M., and Bos, J. L. Genetic alterations during colorectal-tumor development. N. Engl. J. Med., 319: 525-532. 1988.
- Snyderman, R., and Cianciolo, G. J. Immunosuppressive activity of the retroviral envelope protein P15E and its possible relationship to neoplasia. Immunol. Today, 5: 240-244, 1984.
- Query, C. C., and Keene, J. D. A human autoimmune protein associated with U1 RNA contains a region of homology that is cross-reactive with retroviral p30<sup>200</sup> antigen. Cell, 51: 211-220, 1987.
- Mellors, R. C., and Mellors, J. W. Antigen related to mammalian type-C RNA viral p30 proteins is located in renal glomeruli in human systemic lupus erythematosus. Proc. Natl. Acad. Sci. USA, 73: 233-237, 1976.
- Mellors, R. C., and Mellors, J. W. Type C RNA virus-specific antibody in human systemic lupus erythematosus demonstrated by enzymoimmunoassay. Proc. Natl. Acad. Sci. USA, 75: 2463-2467, 1978.
- Rucheton, M., Graafland, H., Fanton, H., Ursule, L., Ferrier, P., and Larsen, C. J. Presence of circulating antibodies against gag-gene MuLV proteins in patients with autoimmune connective tissue disorders. Virology, 144: 468– 480, 1985.
- O'Connell, C. D., and Cohen, M. The long terminal repeat sequences of a novel human endogenous retrovirus. Science (Wash. DC), 226: 1204-1206, 1984.
- May, F. E. B., Westley, B. R., Rochefort, H., Buetti, E., and Diggelmann, H. Mouse mammary tumour virus related sequences in human DNA. Nucleic Acids Res., 11: 4127-4139, 1983.



# **Cancer Research**

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

## **Endogenous Retroviral Elements in Human DNA**

C. Leib-Mösch, R. Brack-Werner, T. Werner, et al.

Cancer Res 1990;50:5636s-5642s.

**Updated version** Access the most recent version of this article at:

http://cancerres.aacrjournals.org/content/50/17\_Supplement/5636s

**E-mail alerts** Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications

Department at pubs@aacr.org.

**Permissions** To request permission to re-use all or part of this article, contact the AACR Publications

Department at permissions@aacr.org.