Oncogene (2003) 22, 1528–1535

www.nature.com/onc

Quantitation of HERV-K *env* gene expression and splicing in human breast cancer

Feng Wang-Johanning[1], Andra R Frost[2], Bixi Jian[1], Lidia Epp[1], Danielle W Lu[3] and Gary L Johanning[4]

[1]*Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA;* [2]*Department of Pathology, University of Alabama at Birmingham, Birmingham, AL, USA;* [3]*The Lauren V. Ackerman Laboratory of Surgical Pathology, Department of Pathology and Immunology, Washington University Medical Center, St Louis, MO, USA;* [4]*Department of Nutrition Sciences, University of Alabama at Birmingham, Birmingham, AL, USA*

Human endogenous retroviruses (HERVs) comprise up to 8% of the human genome. In previous studies, we demonstrated that type 1 HERV-K envelope (*env*) transcripts are expressed in most human breast cancers, but not in normal breast tissues. In the current study, we report that type 2 HERV-K *env* transcripts are also present in human breast cancers. By real-time RT–PCR, the expression of HERV-K *env* transcripts was 5–10-fold higher in breast cancer cell lines treated with estradiol and progesterone than in cells without treatment, and expression was significantly higher in most breast cancer tissues than in normal breast tissues. Furthermore, both types of HERV-K *env* transcripts were capable of being spliced into subgenomic env transcripts and various splice donor and acceptor sites were detected in breast cancers. The selective expression and distribution of multiple HERV-K endogenous retroviral element splice variants in breast cancer, but not in normal controls, suggests that they are novel breast tumor markers.
Oncogene (2003) **22,** 1528–1535. doi:10.1038/sj.onc.1206241

Keywords: breast carcinoma; retrovirus; HERV envelope; and real-time RT-PCR

Human endogenous retroviruses (HERVs) comprise up to 8% of the human genome, as analysed by the Human Genome Sequencing Project (Lander *et al*., 2001). As a rule, retroviruses start to lose infectivity as a result of the accumulation of mutations, a process accentuated by selection against a replicating pathogen. These HERVs are mostly noninfectious, replication-defective, retroviral remnants transmitted as stable Mendelian genes. However, in humans, reports of detection of HERV mRNA, HERV proteins, and even HERV particles in human diseases, especially cancer, have increased in recent years. The pathogenic potential of nondefective endogenous retroviruses has so far only been demonstrated in mice, in which they may induce tumors and immunological disorders (Asch *et al*., 1993; Natoli *et al*., 1996). It has been suggested that there are 170 copies of HERV-K genomic DNA within the human genome (Tristem, 2000). Several other reports suggest the existence of at least 10 different sequence families in the human genome with varying degrees of homology to HERV-K10 and mouse mammary tumor virus (MMTV) *pol* genes, as well as to each other (Franklin *et al*., 1988; Medstrand and Blomberg, 1993; Barbulescu *et al*., 1999).

Of the many HERV families, only K appears to have the full complement of open reading frames (ORFs) typical of replication-competent mammalian retroviruses (LTR-gag-pro/pol-env-LTR) (Ono *et al*., 1986; Mayer *et al*., 1999). HERV-K retrovirus-like particles that encode for viral particles in the teratocarcinoma cell line GH (Lower *et al*., 1993; Lower *et al*., 1995) and the breast cancer cell line T47D (Ono *et al*., 1987) have been described. In teratocarcinoma cell lines, HERV-K *gag* genes are expressed, leading to the production of viral core proteins and virus particles (Lower *et al*., 1995).

Previously, we reported that type 1 HERV-K transcripts with coding potential for the envelope region are expressed frequently in human breast cancer. In this study, we investigated expression of type 2 HERV-K *env* mRNAs and the splicing pattern of both types of HERV-K in human breast cancer tissues. Messenger RNA was isolated from 31 breast cancers; cDNAs were then synthesized. A sense primer specific to the 292 bp region of HERV-K type 2 (5′-GTATGCTGCTTG-CAGCCTTGATGAT-3′, nucleotide (nt) 6674–6698; Accession number: AF074086, Figure 1a) (Mayer *et al*., 1999), and the same HERV-K antisense primer described previously (Wang-Johanning *et al*., 2001) were used to detect type 2 HERV-K *env* transcripts by RT–PCR in human breast cancer tissues. Some breast cancer tissues expressed only type 1 (1104 bp; Figure 1b, upper panel) HERV-K *env* transcripts, some expressed only type 2 (1194 bp; Figure 1b, middle panel), and some expressed both types of HERV-K *env* region transcripts with the same or varying intensities. β-Actin was used as an internal control for each amplified cDNA sample

*Correspondence: F Wang-Johanning, Department of Medicine, University of Alabama at Birmingham, 1824 6th Avenue South, Birmingham, AL 35294-3300, USA; E-mail: fengj@uab.edu
Received 18 September 2002; revised 8 November 2002; accepted 12 November 2002

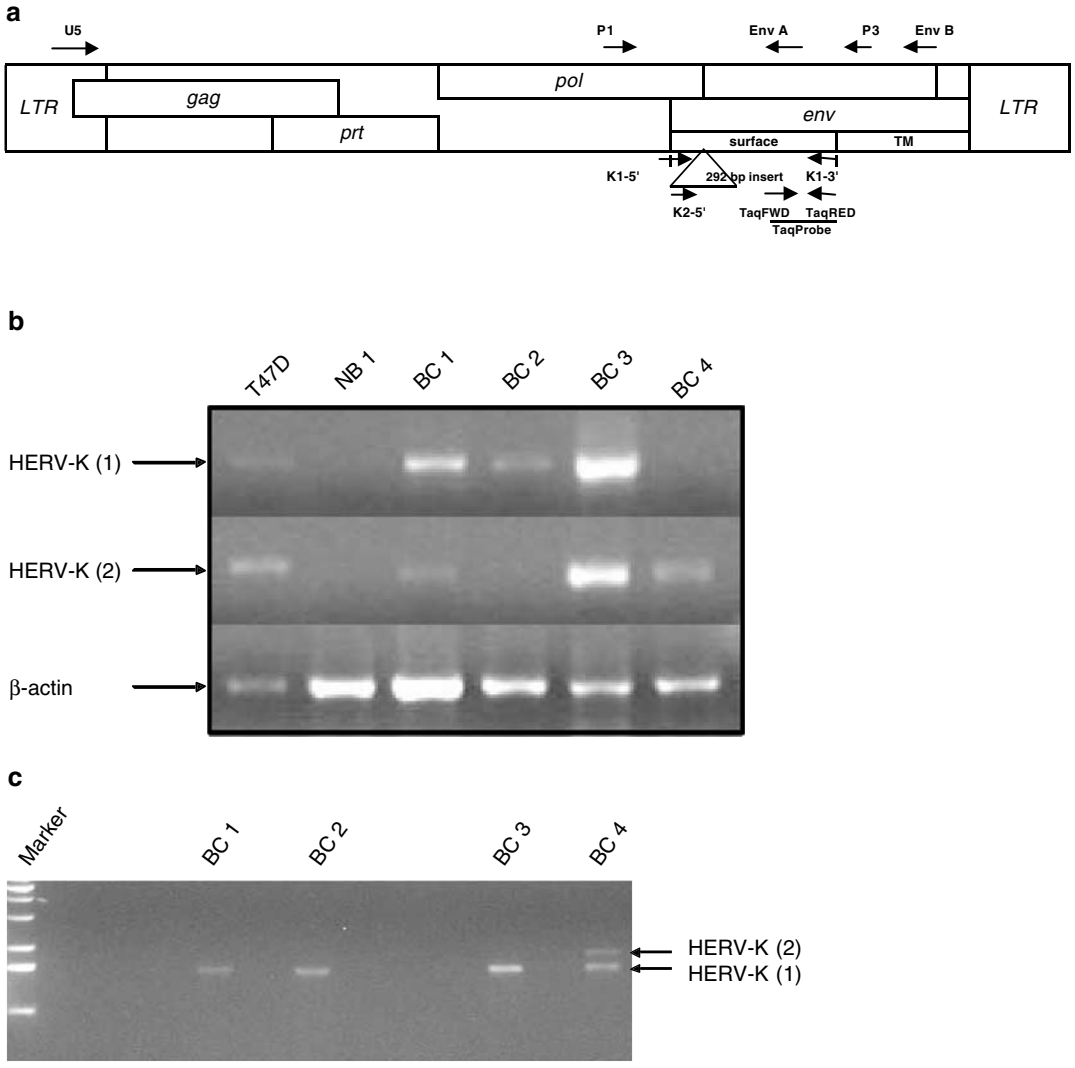

Figure 1 (**a**) Position of primers and probe in the HERV-K *env* gene. K1-5′ and K1-3′ primers were used to amplify type 1 HERV-K env surface protein. K2-5′ and K1-3′ primers were used to amplify type 2 HERV-env surface protein. U5 and EnvA, or U5 and EnvB primers were used to amplify spliced subgenomic transcripts. P1 and P3 primers were used to amplify full-length HERV-K transcripts. TaqFWD and TaqREV primers, and TaqProbe probe were used for quantitative RT–PCR. A triangle containing the 292 bp insert between pol and env regions represents the type 2 HERV-K. (**b**) Expression of type 1 and type 2 HERV-K *env* transcripts in breast tumor and normal tissues. HERV-K RT–PCR products obtained from T47D (a HERV-K positive breast cell line), NB (normal breast tissue), and BC 1–4 (breast cancer tissues number 1–4). The HERV-K (1) arrow shows the position of the 1104 bp fragment representing type 1 HERV-K surface region. The HERV-K (2) arrow shows the position of the 1194 bp fragment representing type 2 HERV-K surface region. The β-actin bands were used as internal controls. (**c**) Expression of HERV-K full-length transcripts in breast cancer tissues. The HERV-K (1) arrow shows the position of the 1616 bp fragment representing type 1 HERV-K full-length transcripts. The HERV-K (2) arrow shows the position of the 1908 bp fragment representing type 2 HERV-K full-length transcripts

(Figure 1b, lower panel). Of the human breast cancer samples analysed by RT–PCR, 71% of breast cancer samples were positive for HERV-K type 1 mRNA expression and 52% were positive for type 2 expression ($n = 31$ HERV-K positive breast cancer samples). The RT–PCR products obtained from an invasive ductal carcinoma (BC3; Figure 1b) were completely sequenced and the results are shown in Table 1. Three clones representing type 1 HERV-K were 99–100% homologous to HERV-K 102 (the prototypical HERV-K type 1 sequence) with 0–3 bp difference (in a total 1104 bp amplified). Meanwhile, of five clones containing a 292 bp insert representing type 2 HERV-K, three of five were 98% homologous to HERV-K 109 (Accession number: AF164615) (Barbulescu *et al*., 1999) with at least 13 bp difference in a total of 1194 bp, and two of five were 97–98% homologous to HERV-K (I) (Accession number: AB047209.1) (Sugimoto *et al*., 2001).

The predominance of full-length type 1 HERV-K provirus transcript expression over type 2 expression was verified by Northern blot analysis in our earlier study (Wang-Johanning *et al*., 2001). In the current study, full-length HERV-K provirus transcription was detected in human breast cancer tissues

Table 1 cDNA sequence analysis of a breast cancer sample (invasive ductal carcinoma)

Transcripts	*Primer used*	*Clone no.*	*Homology (%)*	*HERV-K*	*Accession*	*Nt*
Env region (type 1)	K1-5′ and K1-3′	165K10C3	100	K102	AF164610.1	6491–7552
	K1-5′ and K1-3′	165AC4E	99	K102	AF164610.1	6491–7575
	K1-5′ and K1-3′	165KC2	99	K102	AF164610.1	6471–7575
Env region (type 2)	K2-5′ and K1-3′	165K22EC	98	K109	AF164615.1	6660–7501
	K2-5′ and K1-3’	165K22C11	97	K (I)	AB047209.1	7533–8680
	K2-5′ and K1-3′	165K22C18	98	K (I)	AB047209.1	7533–8680
	K2-5′ and K1-3′	165K22C6	98	K109	AF164615.1	6659–7194
Full-length (type 1)	P1 and P3	165P1C2	98	K102	AF164610.1	6072–7690
	P1 and P3	165P1C8	95	K (II)	AB047240.1	7016–8632
Spliced subgenomic Type 1 and type 2	U5 and Env A	165U2	99	cORF	X82271.1	12–533
	U5 and Env A	165U3	99	K102	AF164610.1	833–878/5997–6658
	U5 and Env A	165U4	99	cORF	X82271.1	12–533
	U5 and Env A	165U5	99	K102	AF164610.1	835–878/5997–6658
	U5 and Env B	165EBC22	99	K102	AF164610.1	835–1076/8117–8186
	U5 and Env B	165EBC23	99	K102	AF164610.1	835–1076/8117–8186

using the P1 (CAACAGGAATTCCTTATAATTCCC, integrase region, nt 6072–6095, according to HERV-K102 sequence; Figure 1a) and P3 antisense (CAATCTTGTAGAATTCTTTTGCC, *env* region, nt 7690–7668, according to HERV-K102) primers described previously (Lower *et al*., 1995). This primer pair amplifies only cDNA derived from full-length transcripts because the sequences to which the P1 primer binds in the *gag/pol* intron are spliced out into subgenomic *env* mRNAs. PCR using primers P1 and P3 to amplify human breast cancer mRNAs revealed a band at 1616 bp (type 1; see Figure 1c; lower band) in most HERV-K-positive breast cancers and a band at 1908 (type 2; upper band) in a minority HERV-K-positive cancers (Figure 1c). Two clones representing full-length type 1 HERV-K transcripts were obtained from an invasive ductal carcinoma (BC3 from Figure 1b) and sequenced. These clones showed 95–98% homology to HERV-K (II) and HERV-K 102 sequences (Table 1).

The expression of HERV-K was increased if T47D cells were treated with β-estradiol (10 nM) on day 1 followed by progesterone (100 nM) on day 2 as described by our group and others (Ono *et al*., 1987) (Figure 2a). Therefore, the change in expression of HERV-K in breast cancer cells in response to hormones was quantified by real-time RT–PCR. The HERV-K primers and probes used were: forward (TaqFWD; 5′-ATTGGCAACACCGTATTCTGCT-3′; nt 7413–7434 according to HERV-K102 sequence or 7706–7727 according to HML-2.HOM sequence), reverse (TaqREV; 5′-CAGTCAAAATATGGACGGATGGT-3′; nt 7517–7495 for K102 or 7810–7789 for HML-2.HOM sequence) and probe (TaqProbe; 5′-FAM-ACACAGGGATCCACACGCCCTCTCTT-TAMRA-3′; nt 7469–7444 for HERV-K102 or 7762–7737 for HML-2.HOM sequence). The HERV-K primers and probe for TaqMan real-time RT–PCR are able to amplify a 155 bp length PCR product from both types of HERV-K *env* regions from HERV-K10 (HUMERVKA), HERV-K102 (AF164610), HERV-K103 (AF164611), HERV-K104 (AF164612), HERV-K107 (AF164613), HERV-K108 (AF164614), HERV-K109 (AF164615), HERV-K113 (AY037928.1), HERV-K115 (AY037929.1), and HML-2.HOM (AF074086.2). Using glyceraldehyde-3-phosphate dehydrogenase (GAPDH; Perkin-Elmer, Foster City, CA, USA) as an endogenous control, quantitation of the amount of target (copy number/pg GAPDH RNA) in unknown samples is accomplished using the standard curve method, with each unknown HERV-K RNA sample being normalized on the basis of GAPDH content. One-step real-time RT–PCR was performed using an ABI PRISM 7700 sequence detector as follows. Known amounts of HERV-K cRNA molecules (1×10^6, 1×10^5, 1×10^4, and 1×10^3 copies) were used to generate an absolute standard curve. The copy number of HERV-K mRNA was then determined by linear extrapolation of the C_T values using the equation of the line obtained from the absolute HERV standard curve. These values were then divided by the relative amounts of GAPDH in an endogenous control (normal breast RNA; Biochain Institute Inc.). The results (Figure 2b) clearly indicate that the HERV-K *env* mRNA copy number per ng RNA increased from 4–10-fold in response to hormone treatment in most of the breast cancer cell lines studied, including BT-20, ZR-75-1, MCF-7, T47D, and SKBr3. However, hormone treatment did not increase copy number in MDA-MB-231 cells, likely because this cell line does not express estrogen and progesterone receptors. The HERV-K *env* mRNA copy number was lower in normal human mammary epithelial cells (HMEC; Figure 2b) than in any of the breast cancer cell lines, with or without hormone treatment.

The expression of HERV-K in human breast biopsies was subsequently evaluated by real-time RT–PCR analysis as described previously (Wang-Johanning *et al*., 2002). The results of one of these analyses are shown in Figure 2c. The expression of HERV-K *env* mRNA was elevated in approximately 70–80% of breast cancers compared with normal breast tissues. The ratio of HERV-K mRNA copy number in breast cancers to the average copy number in 10 normal breast control tissues was calculated. The HERV-K *env* expression increased from six to 20-fold in ductal carcinoma *in situ*

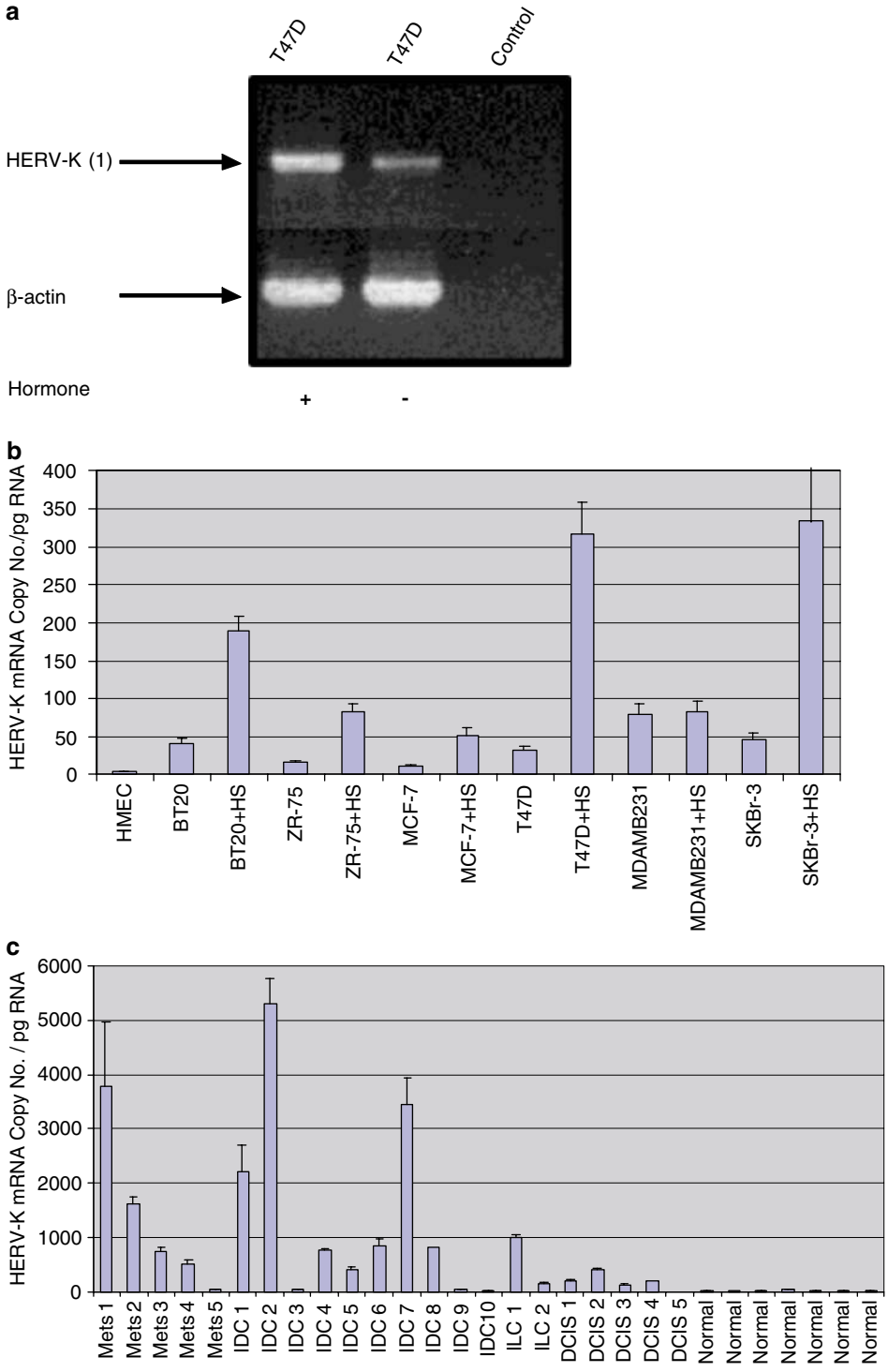

Figure 2 Expression of HERV-K *env* transcripts in breast cell lines. Evaluation of expression of HERV-K *env* gene in breast cell lines by RT–PCR or real-time RT–PCR. (**a**) RT–PCR analysis of expression of HERV-K *env* transcripts in T47D breast cancer cells with or without estradiol and progesterone treatment. (**b**) RT–PCR analysis of expression of HERV-K *env* transcripts in breast cancer cell lines with or without estradiol and progesterone treatment. H.S.: hormone-stimulated breast cell lines; HMEC: normal human mammary epithelial cells. (**c**) Quantitation of HERV-K *env* transcript expression in human breast tissues by real-time RT–PCR: Quantitation of HERV-K *env* transcripts in human breast tissue RNA obtained from subjects with ductal carcinoma *in situ* (DCIS), invasive ductal carcinoma (IDC), and normal controls (NC) without breast cancer

and 20- to 200-fold in invasive ductal carcinoma samples with or without metastasis to the lymph nodes, relative to expression in normal breast tissue.

To evaluate more closely expression in epithelial cells of uninvolved and cancer tissues, we employed an LCM technique to isolate separately breast carcinoma cells

and uninvolved normal epithelial cells; the purity of the populations of cancer and normal cells obtained from the same patient was greater than 90%. Frozen sections fixed in 70% ethanol and stained with hematoxylin and eosin (H&E) stain were used, and cells were microdissected from the sections using an LCM system with an infrared diode laser (PixCell II System, Arcturus Engineering, Mountain View, CA, USA). RNA from the microdissected cells was isolated and treated with DNase, followed by quantitation of the level of expression of HERV-K in the different cell populations by real-time RT–PCR. Real-time RT–PCR quantitation of HERV-K mRNA levels in LCM cells (approximately 300–2000 cells) derived from cancer or uninvolved adjacent epithelium were measured, and the copy number was 5–8-fold greater in tumor epithelial cells than in uninvolved normal epithelial cells in most of the patients studied (data not presented). Approximately 25% of breast cancer tissues showed a very low expression of HERV-K in both tumor cells and adjacent uninvolved ductalobular epithelium obtained by LCM. Factors that contribute to increased HERV-K expression in some breast cancers but not in others are currently unknown.

Expression of both types of HERV-K full-length or *env* transcripts has been observed by other groups. However, only type 2 HERV-K splicing to a subgenomic *env* mRNA has been documented in teratocarcinoma cell lines (Lower *et al*., 1995), or only type 1 HERV-K splicing to a subgenomic *env* mRNA has been documented in the T47D cell line (Etkind *et al*., 1997). In this study, we employed the same primer pairs described previously to identify type 1 and type 2 splicing patterns (Etkind *et al*., 1997). Both types of HERV-K *env* transcripts were able to undergo splicing into subgenomic *env* mRNAs (Figure 3). Two bands, 790 bp (representing type 2 HERV-K) and 497 bp (representing type 1 HERV-K), were amplified from HERV-K positive breast cells or tissues by RT–PCR using U5 and EnvA primers. The difference in the size of the two bands is 292 bp. Four clones (165U2, 165U4, 165U3, and 165U5), derived from breast cancer patient #165 (BC3; see Figure 1b) with invasive ductal carcinoma (see Tables 1 and 2), were amplified via nested RT–PCR using U5 and e*nv* A primer pairs (790 bp product). After sequencing, clones #165U2 and 165U4 were found to be 99% homologous to type 2 HERV-K sequences, such as HERVCORF (X82271.1) and HERV-K(HML-2.HOM) (AF074086.2), which contain a 292 bp insert between the *pol* and *env* regions. Clones #165U3 and 165U5 were 99 and 98% homologous, respectively, to type 1 HERV-K sequences, such as HERV-K102 (AF164610), which contain the 292 bp deletion.

Also, the splice donor (SD) and splice acceptor (SA) sites, respectively, for type 2 HERV-K were at nt 1076 and 6433, based on HML-2.HOM (AF074086.2) sequence numbering (Table 2), and for type 1 HERV-K were at nt 876 (following the poly (A) $(A)_{27}GG(A)_{12}$ tail), and at nt 5997, based on the HERV-K102 sequence (Table 2). This is the first report that shows that both types of HERV-K are transcriptionally active in breast cancer.

Importantly, the four clones derived from this breast cancer patient encode ORFs with no stop codons, suggesting that they have the potential for expression at the protein level. Other SD and SA sites derived from breast cancer cell lines and tissues are shown in Table 2. Clones from other breast cancer cells and tissues (with the exception of MDA-MB-231 and MCF-7 cell lines) did not have defective ORFs, pointing to the potential for env protein expression in breast cancer (Table 2). The putative translated proteins from clones representing type 2 HERV-K are highly homologous (>95%) to the HERV-K env protein (AF164614), with a putative nuclear localization signal (NLS; aa 13–20; RRRRHRNR) and a leucine-rich nuclear export signal (NES; aa 53–64; LKKLTQLATKYL) (Boese *et al*.,

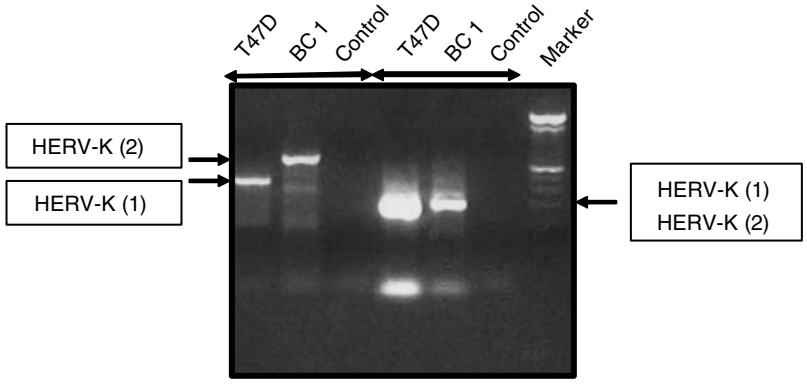

Figure 3 Detection of splicing into subgenomic transcripts in breast cancer cells and tissues by RT–PCR. RNA obtained from hormone-treated T47D cells (lanes 1 and 4) and breast cancer tissue (an invasive ductal carcinoma patient) (lanes 2 and 5) was amplified by RT–PCR using primers (Figure 1) specific to the HERV-K U5 region and either EnvA (lanes 1–3) or EnvB (lanes 4–6) regions. Lanes 3 and 6 are negative controls. The 790 bp arrow shows the 790 bp fragment obtained from the breast cancer tissue specimen. The 497 bp arrow shows the 497 bp fragment obtained from the hormone-treated T47D human breast cancer cell line. The 311 bp arrow shows the 311 bp fragment obtained from T47D cells and the breast cancer tissue (lanes 4 and 5, respectively)

Table 2 Analysis of SD and SA in breast cancer cell lines and tissues

Tissue	*165U2*[1]	*165U3*[2]	*165U4*[1]	*165U5*[2]	*177U26*[1]	*177U29*[1]	*178U11*[2]	*178U15*[2]	*165UB22*	*165UB23*
SD	1076	876	1076	876	961	961	928	928	1076	1076
SA	6433	5997	6433	5997	6948	6948	6399	6399	8117	8117
ORF[3]	√	√	√	√	√	√	√	√	√	√
Cell lines[4]	*T47DU3*[2]	*T47DU4*[2]	*MCF7c28*[5]	*231U3*[5]	*ZR75c33*[6]	*Tera2U2*[2]	*Tera2U3*[2]	*Tera1U13*[1]	*Tera1U15*[1]	*Tera1UB4*[1]
SD	883	883	2078	2078	1076	1076	1076	1076	1035	1076
SA	6222	6222	7599	7599	8410	6433	6433	6433	6756	6433
2nd SD[7]						6492	6492			
2nd SA[8]						6946	6794			
ORF[3]	√	√	N/A[9]	N/A	N/A	X	X	√	X	√

[1]Type 2 HERV-K, nt numbered according to HML-2.HOM sequence (AF074086.2); [2]type 1 HERV-K, nt numbered according to HERV-K102 sequence (AF164610); [3]ORF: open reading frame. √ means ORF without stop codon, and X means ORF with one or more than one stop codon; [4]231: MDA-MB-231; ZR75: ZR75-1; [5]type 1 HERV-K, nt numbered according to HERV-K (II) sequence (AB047240); [6]Type 2 HERV-K, nt numbered according to HERV-K113 sequence (AY037928.1); [7]Second splice donor; [8]Second splice acceptor; [9]N/A: could not be determined

2000b). The putative translated proteins from clones representing type 1 HERV-K are highly homologous (95%) to the pol/env ORF (AAA88033.1). There were at least two stop codons between the pol and env regions, indicating that the env protein does not undergo fusion with the pol protein in these clones.

The potential for protein expression may explain, in part, why certain viral genes are preferentially activated in breast cancer. In contrast to our observations that the *env* gene of HERV-K is expressed in breast cancer, Yin *et al*. (1997) found that the *pol* gene of HERV-K was no more expressed in malignant than nonmalignant breast tissue. The HERV-K10 gene possesses a *pol* ORF but disrupted *gag* and *env* (Ono *et al*., 1986). The variants of HERV-K in breast cancer patients that we describe in our manuscript do not have stop codons in their *env* ORFs, and thus have the potential to be translated to proteins. If these mRNAs are translated, this may account for increased expression of *env* but not *pol* in breast cancer.

Both types of HERV-K env transcripts were detected in T47D cells treated with estradiol and progesterone; however, only type 1 HERV-K splicing in T47D cells was detected and the SD and SA sites were at nt 883 and 6222 according to the HERV-K 102 sequence (AF164610) (Table 2). The type 1 HERV-K splicing pattern was also detected in hormone-treated MCF-7 and MDA-MB-231 breast cancer cell lines, and the SD and SA sites were at nt 2076 and 7599, based on the HERV-K (II) sequences (AB047240). In contrast to clones from most breast cancer cells and tissues, clones derived from teratocarcinoma cells, especially those that were highly homologous to type 1 HERV-K sequences, encoded ORFs with one or more stop codons.

A 311 bp band was detected in T47D cells and in breast cancer tissue using the U5 and *env* B primer pair. No sequence differences among clones of the 311 bp bands were detected in the breast cancer cell lines or tissues analysed. The PCR product of 311 bp has an SD site at nt 1076 and an SA site at nt 8117, based on the HERV-K102 or nt 8410 based on the HML-2.HOM sequences. The RT–PCR sequences of breast cancer tissues revealed that the stop codon described for the putative *env* region of HERV-K10 (M1234; Ono *et al*., 1986) or HERV-K107 (Barbulescu *et al*., 1999) is absent.

The expression of both types of HERV-K, including full-length HERV-K transcripts, *env* transcripts, and spliced subgenomic *env* transcripts, was detected in the majority of human breast cancer tissues analysed. HERV-K expressed in breast cancer has the full complement of ORFs typical of infective retroviruses. The *env* transcripts of type 1 HERV-K (homology to HERV-K102 and HERV-K (II)) were frequently detected in RT–PCR clones from multiple tissues, but the relative abundance of their transcripts varied among tissues. The *env* transcripts of type 2 HERV-K revealed more varied transcriptional profiles, with homology to a greater variety of HERV-K sequences. Increased expression of HERV-K *env* transcripts was associated with hormone treatment, suggesting that hormone-responsive elements may exist in the regulatory regions of the HERV-K genome.

Our LCM/RT–PCR results raise the question of whether HERV-K expression in breast cancer cells might be specifically associated with tumor development. Such expression might form the basis for using HERV-K as a diagnostic biomarker, and further analysis in cancer would be warranted. Furthermore, both types of HERV-K *env* transcripts are able to undergo splicing into subgenomic *env* mRNAs in breast cancer tissues. This suggests that within breast tumor tissue, there are populations of cells that have complex splicing patterns that lead to expression of both types of HERV-K *env* transcripts. Thus, cell lines or individual cells within a population of tumor cells may express one type of HERV-K *env* transcript, but cell populations within the tumor are capable of expressing both types of HERV-K *env*. HERV-K *env* expression in human breast cancer tissues strongly resembles the pattern observed for exogenous retroviruses: full-length transcripts are accompanied by subgenomic *env* mRNAs.

In addition, results of our sequence analysis reveal the presence of several novel alternatively spliced HERV *env* subgenomic transcripts and their smaller spliced products in breast cancer cells or tissues. Multiple SD and SA sites were detected in various breast cancer cell lines

and tissues. Interestingly, these spliced products encode ORFs without stop codons, indicating the potential for translation of these splice variants to proteins. Two spliced HERV-K *env* subgenomic transcripts, central open reading frame (cORF) and np9, have been associated with tumorigenesis. The translated product of cORF is biologically comparable to the human immunodeficiency virus type-1 (HIV-1) Rev protein, critical for nuclear export of unspliced mRNA (Lower *et al*., 1995, Magin *et al*., 1999, 2000; Yang *et al*., 1999; Boese *et al*., 2000a). The 14-kDa cORF protein is expressed in teratocarcinoma cell lines (Lower *et al*., 1995), and HERV-K cORF was reported to induce tumor formation in nude mice (Boese *et al*., 2000a). In the latter study, mice injected with full-length env-transduced Rat-1 cells developed tumors that produced the env splice variant ORF, rather than env itself. This result provides support for the idea that cORF rather than full-length env protein is responsible for cell transformation. Recently, a novel gene, np9, resulting from use of a unique HERV-K type 1-specific splice donor site, was confirmed to be expressed in various tumor tissues and transformed cell lines but not in normal, nontransformed cells, suggesting that the protein may function in tumorigenesis (Armbruester *et al*., 2002). It remains to be seen whether the variants that we have identified here will possess tumorigenic properties similar to the np9 variant just reported.

It is important to determine whether the HERV-K splice products are capable of producing proteins, and whether or not these proteins can serve as tumor-associated antigens in breast cancer. Recently, an HERV-K *env* sequence (HERV-K MEL) encoding an antigen was recognized by cytolytic T lymphocytes obtained from melanoma patients, but not from normal controls (Schiavetti *et al*., 2002). Although the 2.5 kb HERV-K MEL sequence contains many mutations that disrupt the ORFs coding for viral proteins, the antigenic peptide is encoded by a very short ORF present in the region of a spliced HERV-K MEL transcript. These results suggest that HERV env proteins or their peptides are a source of tumor-associated antigens. Some of our variant breast cancer clones were found to be capable of protein expression.

In conclusion, our studies have established that increased HERV-K *env* and spliced *env* expression is associated with breast cancer. Evaluation of expression of the HERV-K *env* gene and its splice variants may thus provide a new breast cancer screening tool, and serve as a novel target for detection, diagnosis, and treatment of breast cancer. The potential impact of the unique splice variants that we have described on breast tumorigenesis will require further investigation, but previous studies of other splice variants of the HERV-K *env* gene suggest that the variants we have described may play a role in the etiology of breast cancer. The results of our studies also indicate that HERV-K proteins are expressed in breast cancer, and therefore are capable of acting as tumor-associated antigens.

Acknowledgements
This work was supported in part by grant CRTG-99-247-01-CCE from the American Cancer Society, Grant 99-003147 from the Susan G Komen Foundation, and Grant DAMD17-00-1-0123 from the Department of the Army.

References

Armbruester V, Sauter M, Krautkraemer E, Meese E, Kleiman A, Best B, Roemer K and Mueller-Lantzsch N. (2002). *Clin. Cancer Res.*, **8,** 1800–1807.

Asch BB, Asch HL, Stoler DL and Anderson GR. (1993). *Int. J. Cancer*, **54,** 813–819.

Barbulescu M, Turner G, Seaman MI, Deinard AS, Kidd KK and Lenz J. (1999). *Curr. Biol.*, **9,** 861–868.

Boese A, Sauter M, Galli U, Best B, Herbst H, Mayer J, Kremmer E, Roemer K and Mueller-Lantzsch N. (2000a). *Oncogene*, **19,** 4328–4336.

Boese A, Sauter M and Mueller-Lantzsch N. (2000b). *FEBS Lett.*, **468,** 65–67.

Etkind PR, Lumb K, Du J and Racevskis J. (1997). *Virology*, **234,** 304–308.

Franklin GC, Chretien S, Hanson IM, Rochefort H, May FE and Westley BR. (1988). *J. Virol.*, **62,** 1203–1210.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR,

de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, Szustakowki J, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S and Chen YJ. (2001). *Nature*, **409,** 860–921.

Lower R, Boller K, Hasenmaier B, Korbmacher C, Muller-Lantzsch N, Lower J and Kurth R. (1993). *Proc. Natl. Acad. Sci. USA*, **90,** 4480–4484.

Lower R, Tonjes RR, Korbmacher C, Kurth R and Lower J. (1995). *J. Virol.*, **69,** 141–149.

Magin C, Hesse J, Lower J and Lower R. (2000). *Virology*, **274,** 11–16.

Magin C, Lower R and Lower J. (1999). *J. Virol.*, **73,** 9496–9507.

Mayer J, Sauter M, Racz A, Scherer D, Mueller-Lantzsch N and Meese E. (1999). *Nat. Genet.*, **21,** 257–258.

Medstrand P and Blomberg J. (1993). *J. Virol.*, **67,** 6778–6787.

Natoli F, Crowley MR, Asch HL, Stoler DL and Asch BB. (1996). *Cancer Lett.*, **99,** 121–127.

Ono M, Kawakami M and Ushikubo H. (1987). *J. Virol.*, **61,** 2059–2062.

Ono M, Yasunaga T, Miyata T and Ushikubo H. (1986). *J. Virol.*, **60,** 589–598.

Schiavetti F, Thonnard J, Colau D, Boon T and Coulie PG. (2002). *Cancer Res.*, **62,** 5510–5516.

Sugimoto J, Matsuura N, Kinjo Y, Takasu N, Oda T and Jinno Y. (2001). *Genomics*, **72,** 137–144.

Tristem M. (2000). *J. Virol.*, **74,** 3715–3730.

Wang-Johanning F, Frost AR, Johanning GL, Khazaeli MB, LoBuglio AF, Shaw DR and Strong TV. (2001). *Clin. Cancer Res.*, **7,** 1553–1560.

Wang-Johanning F, Lu DW, Wang Y, Johnson MR and Johanning GL. (2002). *Cancer*, **94,** 2199–2210.

Yang J, Bogerd HP, Peng S, Wiegand H, Truant R and Cullen BR. (1999). *Proc. Natl. Acad. Sci. USA*, **96,** 13404–13408.

Yin H, Medstrand P, Andersson ML, Borg A, Olsson H and Blomberg J. (1997). *AIDS Res. Hum. Retroviruses* **13**(6), 507–516.