NEOPLASIA
www.neoplasia.com

Volume 13 Number 11 November 2011 pp. 1081–1092 1081

Activation of the Long Terminal Repeat of Human Endogenous Retrovirus K by Melanoma-Specific Transcription Factor MITF-M[1,2]

Iyoko Katoh*, Anna Mírová*,†,3, Shun-ichi Kurata‡, Yasushi Murakami§, Kenji Horikawa§, Natsuko Nakakuki§, Takunobu Sakai§, Kunihiko Hashimoto§, Ayako Maruyama§, Takaaki Yonaga§, Nahoko Fukunishi‡, Kohji Moriishi* and Hirohisa Hirai¶

*Department of Microbiology, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Yamanashi, Japan; †Charles University, Prague, Czech Republic; ‡Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; §HERV-K Research Group, Faculty of Medicine, University of Yamanashi, Yamanashi, Japan; ¶Primate Research Institute, Kyoto University, Aichi, Japan

Abstract

The human and Old World primate genomes possess conserved endogenous retrovirus sequences that have been implicated in evolution, reproduction, and carcinogenesis. Human endogenous retrovirus (HERV)-K with 5′LTR-*gag-pro-pol-env-rec/np9*-3′LTR sequences represents the newest retrovirus family that integrated into the human genome 1 to 5 million years ago. Although a high-level expression of HERV-K in melanomas, breast cancers, and teratocarcinomas has been demonstrated, the mechanism of the lineage-specific activation of the long terminal repeat (LTR) remains obscure. We studied chromosomal HERV-K expression in MeWo melanoma cells in comparison with the basal expression in human embryonic kidney 293 (HEK293) cells. Cloned LTR of HERV-K (HML-2.HOM) was also characterized by mutation and transactivation experiments. We detected multiple transcriptional initiator (Inr) sites in the LTR by rapid amplification of complementary DNA ends (5′ RACE). HEK293 and MeWo showed different Inr usage. The most potent Inr was associated with a TATA box and three binding motifs of microphthalmia-associated transcription factor (MITF). Both chromosomal HERV-K expression and the cloned LTR function were strongly activated in HEK293 by transfection with MITF-M, a melanocyte/melanoma–specific isoform of MITF. Coexpression of MITF and the HERV-K core antigen was detected in retinal pigmented epithelium by an immunofluorescence analysis. Although malignant melanoma lines MeWo, G361, and SK-MEL-28 showed enhanced HERV-K transcription compared with normal melanocytes, the level of MITF-M messenger RNA persisted from normal to transformed melanocytes. Thus, MITF-M may be a prerequisite for the pigmented cell lineage–specific function of HERV-K LTR, leading to the high-level expression in malignant melanomas.

Neoplasia (2011) 13, 1081–1092

Abbreviations: MITF, microphthalmia-associated transcription factor; 5′ RACE, rapid amplification of complementary DNA ends; RPE, retinal pigmented epithelium

Address all correspondence to: Iyoko Katoh, PhD, Department of Microbiology, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Chuo, Yamanashi 409-3898, Japan. E-mail: iyoko@yamanashi.ac.jp

[1]This work was supported in part by the Cooperation Research Program of Primate Research Institute, Kyoto University (H. H.) and by a Grant-in-Aid for Scientific Research (C) 21590434 (I. K.) from MEXT Japan.

[2]This article refers to supplementary materials, which are designated by Figures W1 to W3 and Tables W1 and W2 and are available online at www.neoplasia.com.

[3]International Federation of Medical Students' Associations, 2007.

Received 8 June 2011; Revised 21 September 2011; Accepted 23 September 2011

DOI 10.1593/neo.11794

Introduction

The human genome contains 2.7×10^6 retroelements that cover 42% of the total chromosomal DNA (3 billion nucleotides) [1]. Unlike other forms of retroelements, endogenous retroviruses (ERVs), which make up 8% of the present-day human genomic sequences [1], are derived from archaic infectious retroviruses that integrated into hominids including humans and great apes. Human ERVs (HERVs) represent ERVs prevailing in the human genome [2]. Studies indicated roles of the envelope protein of HERV-W, syncytin, in placental development [3,4] and in neuroinflammation [5].

HERV-K represents the newest family of HERVs that was integrated into the genome approximately 1 to 5 million years ago [6]. HERV-K proviruses with well-conserved 5′LTR-*gag-pro-pol-env-rec/np9*-3′LTR structure, in which LTR signifies the long terminal repeat, exist at 30 to 50 loci on the human chromosomes [1,7,8]. HERV-K types I and II encode regulatory proteins, Np9 and Rec, respectively, at their 3′ termini, which was expressed through double splicing [9,10]. Rec and Np9 exert an oncogenic function by interacting with cellular zinc-finger transcription factors including the promyelocytic leukemia zinc finger (PLZF) protein [11–13]. Intriguingly, high-level HERV-K expression was detected in breast cancers, teratocarcinomas, and melanomas [9,14–17]. Very importantly, intensification of HERV-K expression has been implicated in malignant transformation of melanomas [18–20].

HERV-K transcription is enhanced by stepwise incubation with estradiol and progesterone in breast cancer cells [21,22], although it remains unclear whether the LTR function is directly controlled by the nuclear receptors. Involvement of CpG methylation in the LTR-controlling mechanism was also suggested [23]. A model of the U3-R-U5 framework of the LTR has been proposed very recently [24], in which Sp1 and Sp6 were found essential for activation of the TATA-less promoter. However, the mechanism underlying the cancer-specific HERV-K expression remains unknown.

On the other hand, microphthalmia-associated transcription factor (MITF, also termed Mi) is essential for the development of retinal pigmented epithelium (RPE) and neural crest–derived melanocytes [25]. In addition to the genes essential for melanin synthesis and melanosome formation, a number of genes that support cell cycle progression and survival are induced by MITF [26,27]. MITF, having a basic helix-loop-helix, leucine-zipper (bHLH-LZ) structure [28,29], recognizes the E-box, CA(C/T)GTG, as well as the M-box (TCAYRTG or CAYRTGA) sequences in the promoter/enhancer region of the target genes, where Y and R represent pyrimidine and purine, respectively [25,30]. Importantly, the *MITF* gene (also referred to as *MI*, *WS2A*, *bHLHe32*) is a proposed lineage-specific oncogene, based on the gene amplification and its growth-promoting function in metastatic melanomas [27,31]. Having multiple first exons, *MITF* gives rise to various isoforms, such as MITF-A, -B, -H and M, having different N-terminal peptides [25,29]. Only the MITF-M isoform (MITF variant 4, NM_000248) undergoes melanocyte/melanoma–specific transcription in response to the α-melanocyte–stimulating hormone signaling [29], whereas other isoforms are controlled by distinct promoters and transcribed in many tissues and cancer cell lines [32].

We studied HERV-K transcription controlling mechanisms by determining the initiator (Inr) sites and enhancer sequences in the LTR. Our results indicate that an arrangement of (MITF motifs)–(TATA box)–(Inr) functions as the core enhancer/promoter inducible by MITF-M.

Materials and Methods

Cell Culture

Adenovirus-transformed human embryonic kidney 293 (HEK293) cells and melanoma lines, MeWo and G361, were supplied by the Japanese Collection of Research Bioresources (Osaka, Japan) and maintained with minimum essential medium supplemented with sodium pyruvate, nonessential amino acids, antibiotics, and 10% fetal bovine serum (FBS). SK-MEL-28 melanoma cells were from the American Type Culture Collection. Human epidermal melanocytes from lightly pigmented adult skin (HEMa-LP) were purchased from Invitrogen (Carlsbad, CA), and were cultured in Medium 254 (calcium/serum–free) with human melanocyte growth supplement (Invitrogen).

Plasmids

The HERV-K LTR sequences (GenBank AF074086.2) were amplified by polymerase chain reaction (PCR) from a Bac clone, RP11-33P21 having approximately 154-kb sequences of chromosome 7. The 5′ and 3′ primers used are *Nhe*I-KLTR5-2 (5′-ccgctagctgtggggaaaagcaagagagatca-3′) and *Bgl*II-KLTR3-2 (5′-gcagatctcacctgtgggtgtttctcgtaaggt-3′). The amplified LTR sequences (nos. 1-945) were inserted into pGL3-Basic (Promega, Madison, WI) at the *Nhe*I/*Bam*HI sites. pGL3-Control (Promega), in which the SV40 early promoter is linked to the Luc gene, was used as a positive control. *In vitro* mutagenesis was carried out with the Gene Tailor Site–Directed Mutagenesis System and Platinum Taq High Fidelity (Invitrogen). The MITF-M clone, pHMI-9, was provided by Dr Shigeaki Shibahara through the RIKEN Bioresource Center (RDB 1786). The MITF-M–encoding sequences were excised by *Hin*dIII/*Xba*I digestion and recloned into pRc/CMV (Invitrogen) at the *Hin*dIII/*Xba*I sites.

Luciferase Assay

Cells (10^5 in each well of a 24-well plate) were seeded 24 hours before transfection. Plasmid transfection was with Effectene (Qiagen, Hilden, Germany). Forty-eight hours after the transfection, cells were lysed in Glo-lysis buffer (Promega) for determination of the luciferase activity with the Steady-Glo luciferase assay system (Promega) [33]. For induction with phorbol 12-myristate 13-acetate (PMA), transfected cells (at 16 hours) were starved in the medium containing 0.5% FBS for 8 hours and incubated with or without 20 ng/ml of the drug (Sigma-Aldrich, St Louis, MO) for an additional 24 hours.

Rhesus Monkey Tissues

Tissues of rhesus monkeys, a 13-year-old female and a 6-year-old male, were obtained immediately after killing by exsanguination through the bilateral carotid arteries under deep anesthesia using ketamine hydrochloride and sodium pentobarbital at the Primate Research Institute, Kyoto University, in accordance with the guidelines of the institute.

RNA Purification

Tissues and cultured cells were homogenized in RNAwiz (Ambion, Austin, TX) to purify RNA. Isopropanol-precipitated RNA was further purified with the RNeasy MinElute Cleanup system (Qiagen). DNA fragments possibly contained in the RNA preparations were digested with DNase I (Invitrogen, amplification grade) before reverse transcription (RT)–PCR. The High-Pure RNA Isolation Kit (Roche Diagnostics, Mannheim, Germany) was used for normal melanocytes.

Reverse Transcription–Polymerase Chain Reaction

Semiquantitative RT-PCR was performed with the SuperScript One-Step RT-PCR System with Platinum Taq DNA Polymerase (Invitrogen). KP1-2 (targeting nos. 911-933, 5′-ctctcgtcccaccttacgagaaa-3′), KP1 (nos. 916-933, 5′-tcccaccttacgagaaa-3′), KP3 (*gag*, 5′-gagccattaccggctctgctacat-3′), and KP7 (Rec, 5′-acaaaaccgccatcgtcatcatgg-3′) were newly designed. KP-g3 (5′ end of *gag*, 5′-tttgccccattatcacccta-3′) and KP-e1 (*env*, 5′-gtgacatcccgcttaccatgtgataagtg-3′) have been described [9]. For the detection of the MITF isoforms, MITF-A-F (5′-aaagtcaaccgctgaagagcag-3′), MITF-B-F2 (5′-gtatgcattttggtttttccca-3′), MITF-H-F (5′-atgttcatgccatgctcctt-3′), and MITF-M-F2(ATG) (5′-atgctggaaatgctagaatataatcact-3′) were used as forward primers. Reverse primers were MITF-b1b-R (5′-actcacgggcactctctgtt) and MITF-R2 (5′-gaatgtgtgttcatgcctgg-3′). Human 18S ribosomal RNA (rRNA) was amplified with Hu-18S-F (5′-GTAACCCGTTGAACCCCATT-3′) and Hu-18S-R (5′-ccatccaatcggtagtagcg-3′). Rhesus monkey 18S rRNA was tested with Mmula-18S-F (5′-aattccgataacgaacgaga-3′) and Mmula-18S-R (5′-atctaagggcatcacagacc-3′). These primers were synthesized by Invitrogen.

Determination of the Transcriptional Initiation Sites

The 5′ ends of HERV-K full-length RNA were determined by the 5′ rapid amplification of complementary DNA ends (5′ RACE) system (Invitrogen). We used a primer, KP7, targeting the *env*/*rec*/*np9* region as the gene-specific primer in the RT reaction. KP-g3 (5′-tttgccccattatcacccta-3′) targeting a 5′ terminal *gag* region and *Bgl*II-GSP3-RACE5 (5′-ttccgagctcgctcgagcgtaccttcaccctaga-3′) targeting the leader region were used at the PCR and nested PCR steps, respectively. Products were digested with *Spe*I and *Bgl*II to be cloned into pUC19 and sequenced.

Immunofluorescence Experiments

Human adult normal tissue panels (T8234605) and human fetus tissue arrays (BE01015) were was purchased from Biochain (Hayward, CA) and US Biomax (Rockville, MD), respectively. Immunofluorescence staining was performed as described [34]. After the antigen retrieval by incubation in the LAB solution (Polysciences, Warrington, PA) for 15 minutes at 60°C, cells were permeabilized and incubated with an anti-MITF (rabbit IgG, ab20663; Abcam, Cambridge, United Kingdom) antibody (at a dilution of 1:250) and an anti–HERV-K core (HERM-1831-5, mouse monoclonal; Austral Biologicals, San Ramon, CA) antibody (1:500) for 90 minutes. HERM-1831-5 represents a mouse monoclonal antibody (clone 4D4/F7) obtained by immunization of mice with HERV-K gag antigen (82 kDa). Tissue slides were next incubated with secondary antibodies conjugated with Alexa Fluor 488 or 594 (Invitrogen) for 60 minutes, followed by a 10-minute incubation with Hoechst 33342 (Sigma-Aldrich) for mild nuclear staining. For the control experiment, the anti–HERV-K antibody (1 μg) was neutralized by incubation with the HERV-K gag protein (50 μg) (HERV-K CA; ProteinTech Group, Chicago, IL) in phosphate-buffered saline (500 μl) for 18 hours at 4°C. Fluorescence images were captured by Nikon E600 with the Digital Sight DS-Fi1 system (Nikon, Tokyo, Japan).

Results

Differences in the HERV-K LTR Regulation between HEK293 and MeWo Cells

The LTR region (nucleotides nos. 1-945) of HERV-K (HML-2. HOM, human endogenous MMTV-like), also referred to as ERVK6, on chromosome 7 (GenBank AF074086.2) was amplified by PCR from BAC clone RP11-33P21 and inserted into pGL3-Basic. This locus has a tandem repeat of the HERV-K genome with three LTRs: LTR-1, -2, and -3 (Figure 1*A*) [35]. Our clones HKLTR-1 to -5 were ~99% identical to the reported sequences. We use the nucleotide numbering of AF074086.2 throughout this report, in which nos. 1 and 968 corresponded to the 5′ and 3′ ends of LTR-1 [35]. On the basis of the human genome sequence data (NT_007819.16 and NT_079592.2) and our sequencing analyses, however, we place an A residue at no. 796 so that nos. 793 to 780 octamer reads ′5′-AATAAA-TA-3′ instead of ′5′-AATGAATA-3′.

In the luciferase gene (Luc) reporter assays, HKLTR-1 was three-fold less active than the control SV40 early promoter (pGL3-control) in human embryonic kidney HEK293 cells without the large T antigen. The strong Luc expression in MeWo (Figure 1*B*, *right*) seemed to reflect the high-level chromosomal HERV-K expression. PMA (20 ng/ml) enhanced the HERV-K LTR and the SV40 promoter two-fold and three-fold, respectively, in serum-starved HEK293 cells (Figure 1*B*, *left*). Intriguingly, the LTR was not influenced by PMA in serum-starved MeWo cells, whereas the SV40 promoter was 1.4-fold enhanced (Figure 1*B*, *right*). The LTR function in MeWo cells may not be further potentiated by protein kinase C pathway stimulation. Other HERV-K LTR clones, HKLTR-2 to -5, were indistinguishable from HKLTR-1 in these and the following assays (data not shown).

Deletion analysis of the HERV-K LTR indicated that the promoter activity is regulated by different regions between HEK293 and MeWo cells (Figure 1, *C* and *D*). We were not able to identify a specific region that responded to PMA in HEK293. Most intriguingly, both Del-7 (nos. 684-945) and 7-LTR (nos. 1-704) were able to induce Luc expression in HEK293 cells, whereas Del-7, but not 7-LTR, could exert the promoter function in MeWo. We hypothesized the presence of at least two promoters, one in 7-LTR and the other in Del-7, which are activated in different modes between the two cell lines.

Multiple Initiation Sites in Chromosomal HERV-K LTRs

To determine the initiator (Inr) sequences in the HERV-K LTR on the chromosomal DNA, we performed a 5′ RACE experiment with RNA purified from MeWo and HEK293. To determine the 5′ terminus of the full-length viral RNA, we designed gene-specific primers: KP-7 (targeting *rec*/*np9*), KP-g3 (*gag*), and *Bgl*II-GSP3-RACE5 (5′ leader) for the RT reaction, PCR, and nested PCR, respectively (Figure 2*A*). Seventeen of 40 clones obtained from MeWo were derived from the HERV-K LTR-driven viral gene transcripts (Figure 2*B*, *subpanel (1)*). Four clones contained unrelated 50 to 80 nucleotide sequences; the other 19 are described in the next section.

Within the HERV-K LTR-derived 5′ RACE products, we found 11 clones starting at nucleotide no. 826, four clones starting at no. 796, and single clones starting at no. 910 or no. 918 (Figure 3*A*). All of the detected initiation sites were within the Del-7 region, supporting the Luc assay results in MeWo (Figure 1*D*). However, nos. 910 (T/U residue) and 918 (C residue) were inappropriate as a conventional initiation nucleotide and were speculated to be 5′ ends of processed RNA. Nucleotide BLAST (Basic Local Alignment Search Tool, National Center for Biotechnology Information) searches indicated that each of the no. 826-initiated transcripts, with an exception of clone-3, was templated by a certain HERV-K locus on chromosome 1, 6, 7, 19, or 22 on the basis of the nucleotide substitutions at a few

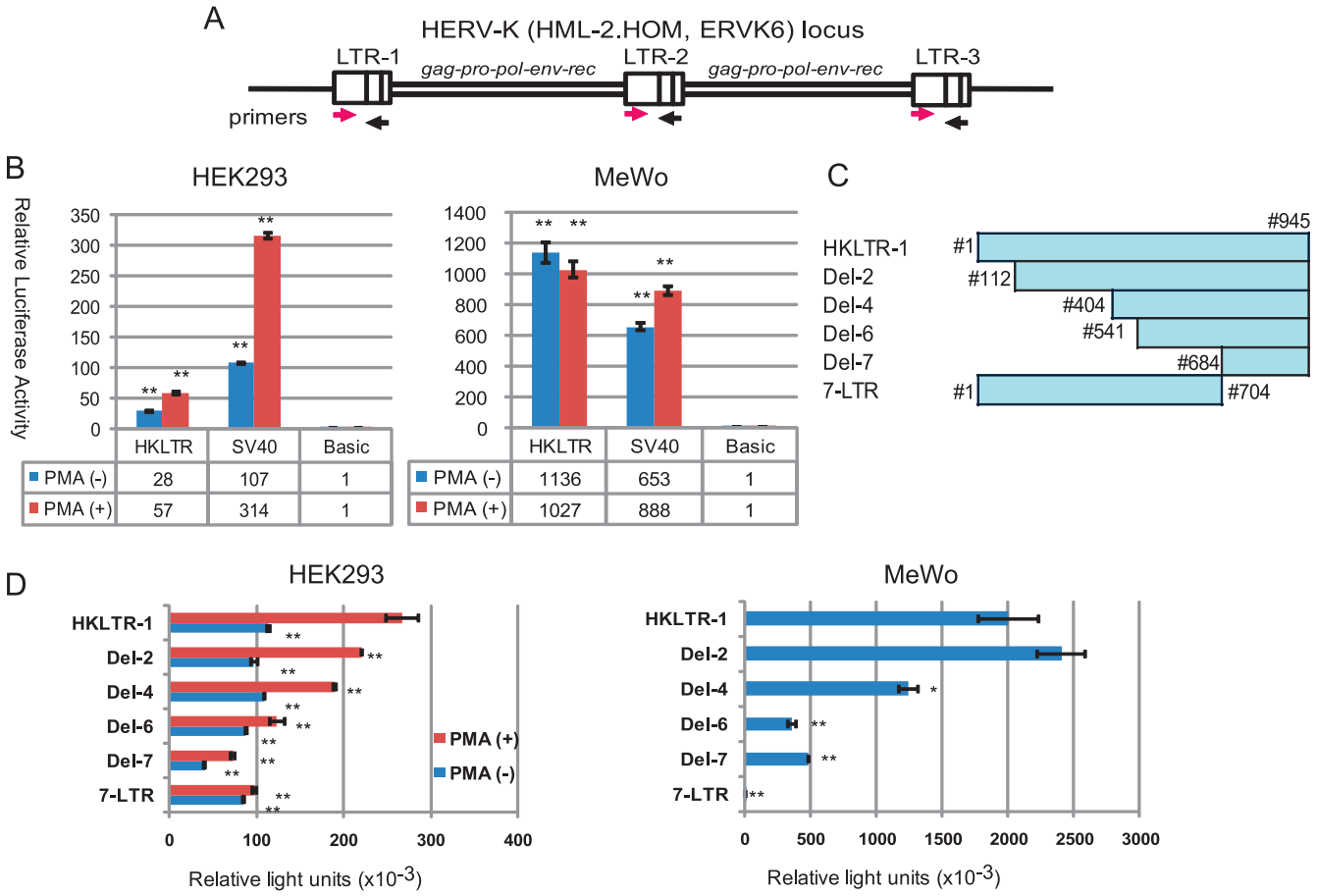

Figure 1. Detection of the promoter function in cloned HERV-K LTR. (A) The HERV-K locus (HML-2.HOM, ERVK6) on chromosome 7. A tandem repeat of the LTR-*gag-pro-pol-env-rec/np9*-LTR structure containing LTR-1, -2, and -3 are illustrated as described [35]. The forward and reverse primers used for the LTR sequence amplification are indicated by magenta and black arrows, respectively, at the positions of their target sequences. (B) Luciferase (luc) gene reporter assay with HKLTR-1-Luc in HEK293 and MeWo cells. The pGL3-control plasmid containing the SV40 early promoter (SV40) and pGL3-Basic without a promoter (Basic) were examined as the positive and negative controls of the experiments. Luciferase activity was determined after stimulation with (+) or without (−) PMA (20 ng/ml) in serum reduced (0.5%) medium. Measurements (mean ± SD) are presented in relation to the enzyme activity (=1) obtained by pGL3-Basic. Experiments were carried out in triplicate two times (**$P < .01$ *vs* Basic). (C) Deletion mutants examined. Nucleotide numbers (#) contained in each plasmid are indicated. (D) Luc assay with the cloned HERV-K LTR and its deletion mutants. Luciferase activity in transfected HEK293 cells was determined after stimulation with (+) or without (−) PMA (20 ng/ml) in serum-reduced (0.5%) medium. Luciferase activities in MeWo cells incubated with 10% FBS were analyzed. Results (mean ± SD) are shown in relative light units (RLUs). Statistical significance: **$P < .01$ and *$P < .05$ *versus* HKLTR (without PMA stimulation).

positions (Figure W1*B*). We failed to identify the templates for the no. 796-initiated transcript (four clones) having an A-to-T alteration at no. 929 in the human genomic sequences (Figure W1*A*). Cell line–specific chromosomal alteration and locus-specific transcriptional activation were speculated for the minor initiation at no. 796. Thus, no. 826 seemed to serve as the most potent Inr (Inr826), which was functional at several loci on the human chromosomes. Interestingly, we found, 25 bp upstream of Inr826, octanucleotide 5′-AATAAATA-3′ (nos. 793-800) matching the TATA box consensus sequence TATA(A/T)A(A/T) [36,37] (Figure 3*B*).

Within the 40 clones of the 5′ RACE products from HEK293 RNA, we identified 7 starting at no. 460, 3 starting at no. 813, and 2 starting at no. 815 (Figures 3*A* and W2). These were also consistent with the Luc assay result that both 7-LTR and Del-7 acted as an independent promoter in HEK293 cells. Nucleotide no. 460 seemed to be the major initiation site in this cell line and was tentatively termed *Inr460*. We noted that all of the 5′ RACE clones starting at Inr460 shared 33 nucleotide substitutions when compared with HERV-K (HML-2.HOM) and were derived from a locus on chromosome 3.

Detection of HERV-K–like Transcripts without a 5′ LTR

Nineteen of the 40 clones obtained from MeWo had 5′-terminal unknown sequences preceding the 5′ leader sequences of HERV-K (Figures 2*B*, *subpanel (2)*, and W1*C*). We could identify the template sequences at 7q22, which showed a 3.6-kb deletion spanning the *pol* and *env* genes (ref: NT_007933.15 and NT_2079596). Because of the long deletion, these transcripts may have been more efficiently reverse transcribed and amplified in the 5′ RACE experiment than the full-length HERV-K messenger RNA (mRNA).

Requirement of the Inr Sequences for the LTR Function

To test whether the major Inr sites (Inr460 and Inr826) and the TATA box-like sequences (793TATA) are necessary for the LTR

function, we constructed mutants termed *HKLTR-460m*, *-826m*, and *793TATAm* (Figure 3*B*). In MeWo cells, the 460m mutation even increased the luciferase activity 1.4-fold (Figure 3*C*), suggesting the possibility that no. 460 is involved in the mechanism to render 7-LTR inactive in MeWo (Figure 1*D*). The 793TATAm and 826m mutations decreased the Luc expression by 97% and 88%, respectively (Figure 3*C*), indicating that both 793TATA and Inr826 play an essential role in the cloned LTR.

Unlike the result in MeWo, luciferase activity produced from HKLTR-460m was 15% decreased in HEK293 (Figure 3*C*). Although Inr826-initiated transcription was not detected by 5′ RACE, Luc expression by 793TATAm and 826m were decreased 60% and 30%, respectively. Thus, 793TATA and Inr826 were also vital in HEK293. Inr460 seemed to contribute to the LTR function in HEK293, but less significantly. The predominant appearance of the Inr460-initiated transcript (Figure 3*A*) may have caused by biased Inr usage at the unique locus (Figure W2*A*) in the nonproducer cell line.

Activation of the Cloned HERV-K LTR by MITF-M

Because HERV-K is efficiently expressed in melanomas and teratocarcinomas, transcription regulatory factor(s) governing the pigmented tissue development could be considered as a potential HERV-K inducer(s). MITF, the MITF-M isoform in particular, is a melanocytic master transcription factor and an oncogene of melanoma [38–40]. Nucleotide sequences corresponding to E-box and related sequences including CACATG, CATGTG, and CATATG serve as the core target sites of MITF to induce gene expression [25]. Intriguingly, three hexanucleotides resembling the MITF binding core motifs were detected upstream of 793TATA within the Del-7 region (Figure 4*A*): MITF-1 (CACATC starting at no. 757, with one nucleotide mismatch

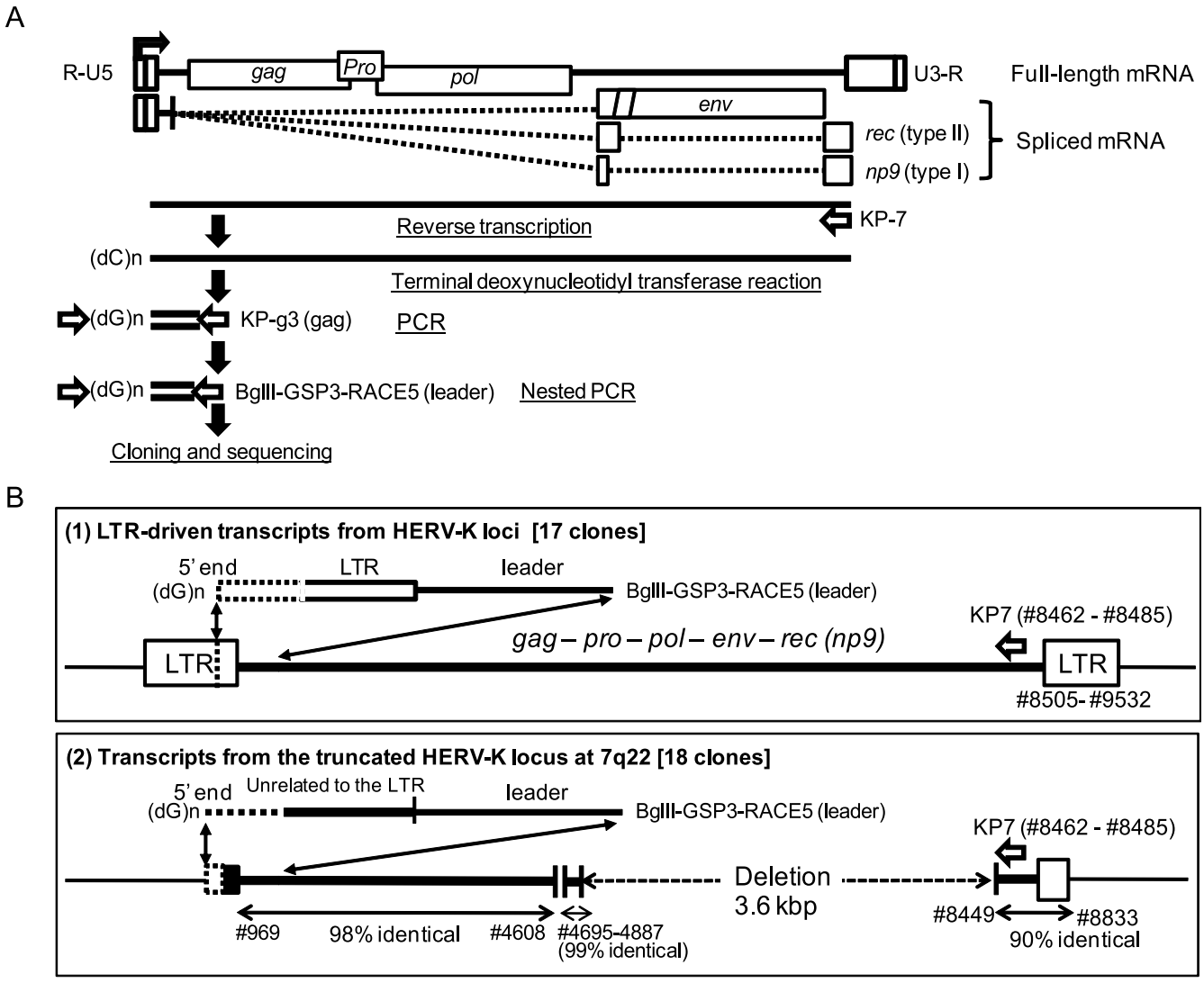

Figure 2. Summary of the 5′ RACE experiments. (A) Strategy for determination of the 5′ end of the viral RNA. HERV-K full-length and spliced RNAs are schematically shown with the *gag*, PR, *pol*, *env*, and *rec*/*np9* genes (top). Experimental processes are illustrated with gene specific primers. (B) Products detected in the experiments. (1) Detection of the transcriptional start points in HERV-K LTR. Twenty-one of 50 clones contained a 3′ terminal region of the 5′ LTR followed by the leader sequences. (2) Expression of a HERV-K-like element rearranged by the loss of 5′ LTR and a 3.6-kb deletion spanning *pol* and *env*. The transcripts without a LTR sequence were aligned with the segmented HERV-K locus at 7q22. Twenty-four of fifty 5′ RACE clones contained this structure.

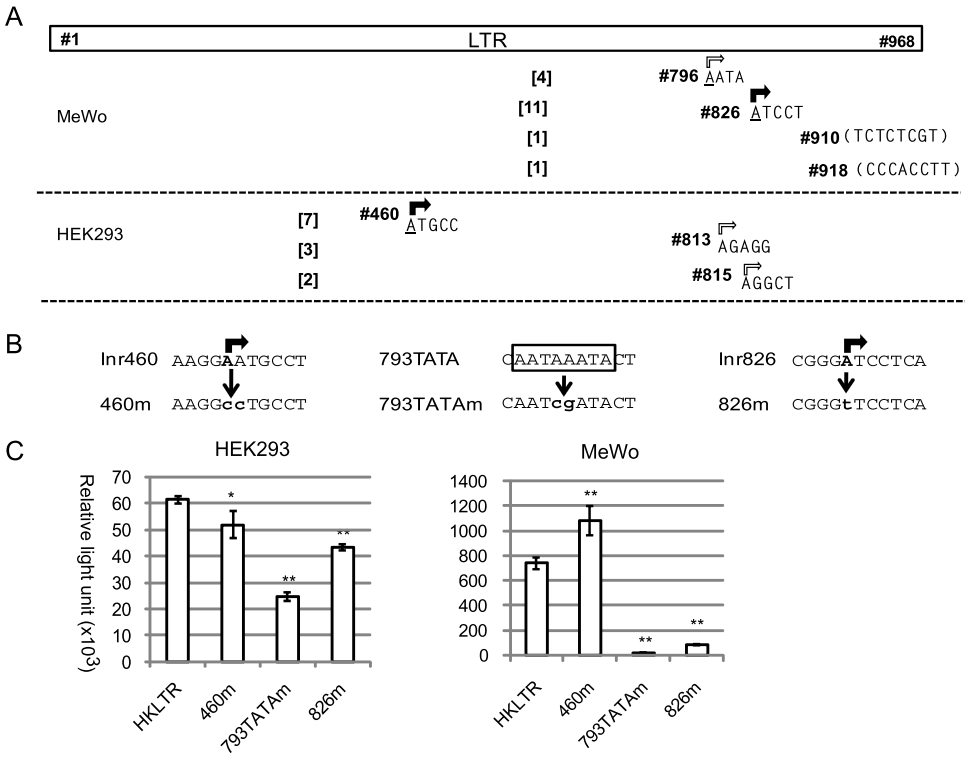

Figure 3. Multiple Inrs detected in the HERV-K LTR. (A) Results of the 5′ RACE analyses. Positions of the 5′ ends (rectangular arrows) of the HERV-K mRNA from MeWo and HEK293 cells are schematically shown. Number of detected clones (from among the 40 clones) is in brackets. (B) Mutation of Inr460, 793TATA, and Inr826. The original nucleotide sequences and mutated sequences in 460m, 793TATAm, and 826m are indicated. Major and minor initiation sites are indicated by filled and open rectangular arrows, respectively. The 793TATA sequences, 5′-TATA(T/A)A(A/T)(A/G)-3′, corresponding to the TATA box consensus are boxed. (C) Luc assay with HKLTR-1 and the Inr mutants in HEK293 and MeWo cells. Luciferase activity (mean ± SD) is shown in RLUs. Experiments were performed in triplicate two times. Statistical significance: $^{**}P < .01$ and $^{*}P < .05$ *versus* HKLTR-1.

underlined), MITF-2 (CTTGTG at no. 744), and MITF-3 (CACATG at no. 688). An additional motif, MITF-4 (CATATG at no. 832), was found six bases downstream of Inr826 (Figure 5*A*). We inserted one or two nucleotide mutations into each MITF motif to generate mutants termed *MITF-1m* (CggATG), *MITF-2m* (CggGTG), *MITF-3m* (CggATC), and *MITF-4m* (CggATG). Results of the Luc assays indicated that MITF-1, -2, and -3, but not MITF-4, contribute significantly to the promoter function of the LTR in MeWo cells (Figure 5*C*).

The basal-level Luc expression in HEK293 was not greatly altered by any of the MITF motif mutations (Figure 5*D*). When HKLTR-luc was cotransfected with a CMV promoter–driven MITF–M expression vector, a substantial increase in Luc expression was observed (~140-fold increase in the luciferase activity). This strong transactivation by MITF-M was cancelled in the MITF-1m, MITF-2m, and MITF-3m mutants, but not in MITF-4m. In Western blot analysis, the MITF-M protein (419 amino acids long) appeared as a minor band of ~45 kDa in MITF-M–transfected HEK293 (293/MITF-M) cells as well as in MeWo (Figure 5*B*). In contrast, HEK293 cells transfected with the control vector (293/vec) lacked this protein. The Del-7 region (nos. 684-945) as well as the entire LTR clone exhibited a dose-dependent response to MITF-M in HEK293, whereas 7-LTR was insensitive to MITF-M (Figure W3). Thus, forced expression of MITF-M in HEK293 caused strong HERV-K LTR activation depending on the MITF-1, -2, and -3 motifs.

Induction of Chromosomal HERV-K by MITF-M

We analyzed *MITF* gene expression in HEK293 and MeWo by RT-PCR with isoform-specific primers (Figure 6, *A* and *C*). *MITF-A* and *MITF-B* were transcribed equally in these cell lines. *MITF-H* was expressed more efficiently in HEK293 than in MeWo. MITF-M was clearly expressed in MeWo but not in HEK293. Total MITF expression did not significantly differ between the cell lines as analyzed with the pan-MITF-2 primer pairs. HEK293 cells transfected with the MITF-M expression vector showed a substantial increase in the MITF-M mRNA without an alteration in other isoforms. These results were consistent with the Western blot analysis results, in which two dominant MITF bands appeared equally in the two cell lines, whereas MITF-M was detectable in MeWo, but not in HEK293 (Figure 5*D*) as observed in various cancer cell lines [32].

We next examined these cells for production of full-length HERV-K mRNA encoding *gag-pol*, spliced *env* mRNA, and doubly spliced *rec/np9* mRNA (Figure 6, *B* and *D*). Forward primers (KP1, KP1-2) were designed to hybridize sequences downstream of Inr826. Consistent with the previous report [9], high-level expression of the spliced *env* mRNA and the doubly spliced *rec* mRNA was evident in MeWo, whereas neither the *env* nor the *rec* RT-PCR product was detectable in HEK293. The *gag*-encoding viral RNA was found in HEK293 at a lower level, whereas it was increased three-fold in MeWo. Interestingly, MITF-M–transfected HEK293 cells significantly enhanced the *env* and *rec* mRNA synthesis. The *gag*-encoding RNA was also increased

two-fold by MITF-M. Thus, forced expression of MITF-M caused chromosomal HERV-K activation in HEK293.

Elements Conserved between the Human and Rhesus-type HERV-K LTRs

We detected 113 LTR copies with 95% or higher sequence identity to the HERV-K LTR (HML-2.HOM) in the human genomic sequences (Table W1). MITF-1, MITF-2, MITF-3, 793TATA, and Inr826 were conserved in 108 LTRs, 102 of which had an equivalent MITF motif (5′CACGTG3′) in place of MITF-3 (5′CACATG3′). Inr460 was also conserved in 86 LTRs but inactivated by a nucleotide substitution (A-to-C) in 27 LTRs.

In the rhesus monkey genome, we detected 33 LTR-like elements with 80% or higher nucleotide identity to the HERV-K LTR (Table W2). The GGA(Inr460)ATGC sequence was disrupted in 12 LTRs but existed in other LTRs in a modified form (AGAAGGC), which may possibly retain the Inr function. Twenty-seven LTRs had 793TATA, whereas four showed a single nucleotide substitution in the octanucleotide. None of the rhesus LTRs had an Inr826-like sequence. However, a heptanucleotide matching the consensus of Inr, $PyPyA^{+1}N(T/A)PyPy$ [41,42], was inserted at a position corresponding to Inr826 in 13 LTRs. Three MITF motifs were contained in three LTRs, two MITF motifs in six LTRs, and only one MITF motif in 23 LTRs. The rhesus version of HERV-K could possibly be controlled by mechanisms similar to those found in human HERV-K.

Detection of MITF and HERV-K Expression in RPE

Rhesus monkey tissues were examined for MITF and rhesus-type HERV-K expression by RT-PCR (Figure 7). The human MITF primer pairs shown previously (Figure 6*A*) were also used in this experiment because the MITF isoform sequences were identical between the two species. The primer pair (KP-1/KP-g3) targeting the sequences shared by the human and rhesus-type HERV-K was used to detect the *gag*-encoding transcript. Rhesus 18S ribosomal RNA was amplified for normalization.

The *MITF-A* mRNA level was similarly high in the kidney, brain, seminal vesicle, testis, and retina but was undetectable in the mammary gland and ovary under the assay conditions. Enhanced *MITF-M* expression was observed not only in the retina but also in the seminal vesicle. The spleen and brain seemed to produce MITF-M mRNA less predominantly. Total *MITF* expression was most intense in the kidney, retina, and seminal vesicle but very weak in the mammary gland and ovary. The *MITF* expression profiles in the rhesus tissues were consistent with the expression pattern extracted from the human EST database and conventional analyses [32], although the seminal vesicle has not been examined previously.

Rhesus-type HERV-K expression was most active in the pigmented tissues, retina, and seminal vesicle [43], supporting the observation that the rhesus-type LTRs also contain MITF motifs (Table W2). Less predominant but correlated expression of HERV-K and MITF-M were detected in the spleen and brain. Other tissues including the testis and kidney showed a level of HERV-K mRNA, which was not correlated to MITF-M mRNA.

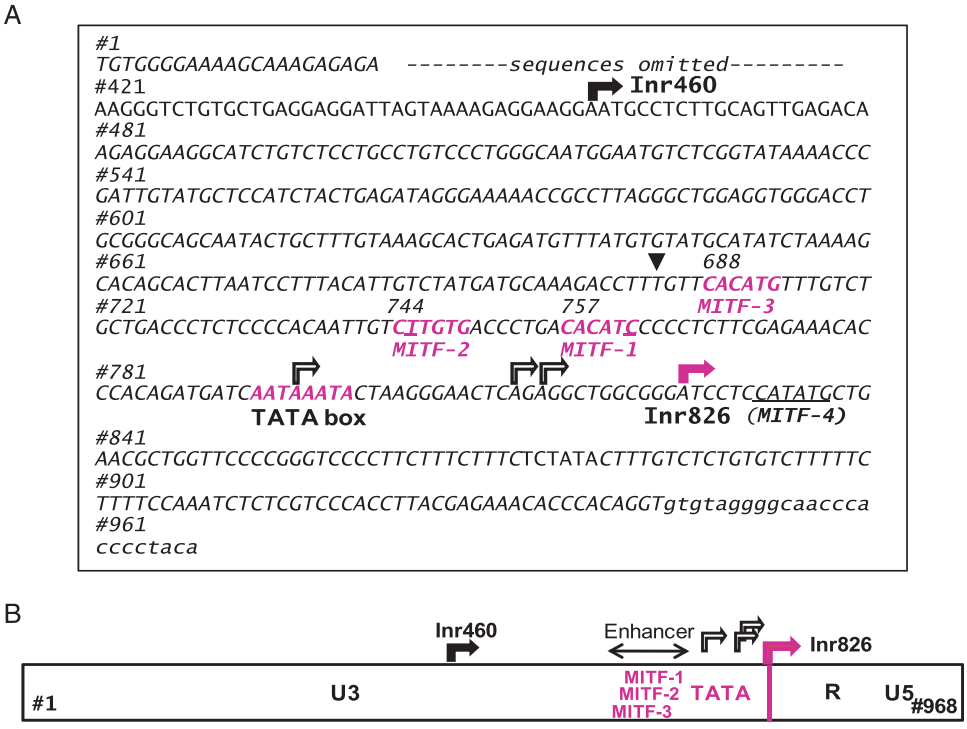

Figure 4. Transcriptional regulatory elements in HERV-K LTR. (A) Nucleotide nos. 1 to 968 of HERV-K LTR are shown according to the numbering in GenBank AF074086.2 (HML-2.HOM). Rectangular arrays indicate the major initiation sites (Inr460 and Inr826) determined by 5′ RACE. MITF-M–responsive sequences (MITF-1, -2, and -3) and 793TATA are highlighted (magenta). MITF-4 is underlined. A list of MITF binding core sequences used in the known MITF-induced genes is provided in reference [25]. The first nucleotide (no. 684) of the Del-7 region is marked by an arrowhead. (B) Proposed framework of HERV-K LTR. Inr826 defines the U3-R junction. Relative positions of Inr826, 793TATA, and the enhancer region formed by MITF-1, -2, and -3 are indicated. Inr460 is also indicated.

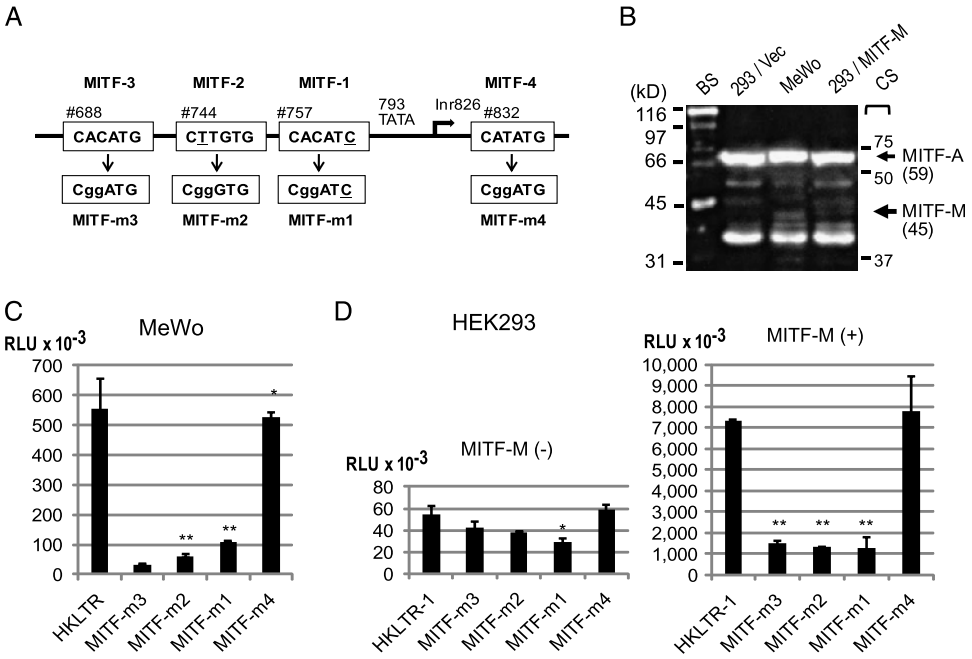

Figure 5. MITF-binding motifs essential for the HERV-K LTR activation by MITF-M. (A) (Top) An array of the MITF-binding motifs (MITF-1, -2, -3, and -4) is graphically shown in relation to the initiation sites (Inr826). (Bottom) Mutated nucleotides in MITF-m1, -m2, -m3, and -m4 are indicated in lower case. (B) Detection of MITF-M by Western blot analysis. Lysates of HEK293 (293), MeWo, and HEK293 transfected with MITF-M (293/MITF-M) were probed with an anti-MITF antibody. Molecular masses (kDa) and positions of the biotinylated protein standards (BS) and prestained color standards (CS) are indicated. MITF-A (520 amino acid residues, 59 kDa; GenBank BAA32288.1; National Center for Biotechnology Information Reference Sequence: NP_937802.1) and 44-kDa MITF-M (419 residues; NM_000248.3) are indicated on the basis of their molecular masses and the RT-PCR analyses shown in Figure 6. (C) Luc assay with the MITF motif mutants in MeWo. Experiments were performed in triplicate two times. Luciferase activity (mean ± SD) is shown in RLUs. Statistical significance: $^{**}P < .01$ and $^{*}P < .05$ *versus* HKLTR-1. (D) Transactivation of the HERV-K LTR by MITF-M in HEK293 cells. HKLTR-1 and the mutants were transfected in combination with a CMV promoter–driven MITF-M expression vector (+) or with the empty vector (−) at a weight ratio of 1:1.

We performed an immunofluorescence analysis with a human fetus tissue array (Figure 8). Because MITF-M–specific antibody was not available, an antibody (ab20663) recognizing the C-terminal domain shared by all of the MITF isoforms was used in combination with an anti–HERV-K core antibody (HERM-1831). In the retina/choroid sections, RPE displayed a strong label by MITF (*red*, Figure 8*A*, *column a*) as reported [44]. The anti–HERV-K antibody also stained RPE strongly (*green*). The choroid region was poorly reactive with these antibodies. In the control experiment, the anti–HERV-K antibody neutralized with the HERV-K gag peptide failed to stain the RPE (Figure 8*A*, *column b*). Other tissues including the liver (Figure 8*B*) were moderately or poorly reactive with the anti-MITF and anti–HERV-K antibodies. Immunofluorescence images of the seminal vesicle sections were not informative because of an intense autofluorescence of the lipofuscin granules in the epithelium [43] (data not shown).

MITF-M and HERV-K Expression in Normal and Transformed Melanocytes

Three malignant melanoma lines, G361, SK-MEL-28, and MeWo, were compared with normal human epidermal melanocytes for the levels of *MITF* and HERV-K expression by semiquantitative RT-PCR (Figure 9). The levels of total *MITF* isoforms and *MITF-M* were constant in the melanoma lines and normal melanocytes. However, G361, SK-MEL-28, and MeWo showed different levels of the *gag* (full-length) (1:4:2.7), *env* (1:3:1.4), and *rec* (1:2:1.5) mRNAs. In normal melanocytes cultured in the serum/calcium–free medium, the amounts of HERV-K mRNAs were only 3% to 10% of those in MeWo. When exposed to 1% FBS for 9 hours, the melanocyte culture increased the *gag* and *rec* mRNAs three-fold or higher without a further increase in *MITF-M* mRNA. Thus, the level of HERV-K transcription was not directly correlated with that of *MITF-M* in melanoma cells and melanocytes, consistent with the requirement of post-translational modifications including Ser73 phosphorylation by Erk2 for functional activation of MITF-M [25,28,45,46]. MITF-M was thought to determine the melanocyte/melanoma lineage specificity of HERV-K expression.

Discussion

We studied transcriptional activation of the HERV-K LTR in melanoma cells. The three MITF binding motifs, the TATA box (793TATA), and the major Inr site (Inr826) seemed to serve as a potent enhancer/promoter unit. In nonmelanoma HEK293 cells, both the cloned LTR function and chromosomal HERV-K expression were strongly activated by forced expression MITF-M, an RPE/melanocyte–specific isoform of MITF. Requirement of MITF-M for the LTR activation seemed to explain the lineage-specific expression of HERV-K.

Endogenous retrovirus expression may undergo varied pathways of transcriptional regulation depending on cell lineages and cellular contexts as reported previously [24,47]. Presence of the multiple Inr sites may reflect the complexity of cellular transcription regulatory mechanisms and the flexibility of retroviral LTRs, allowing ubiquitous

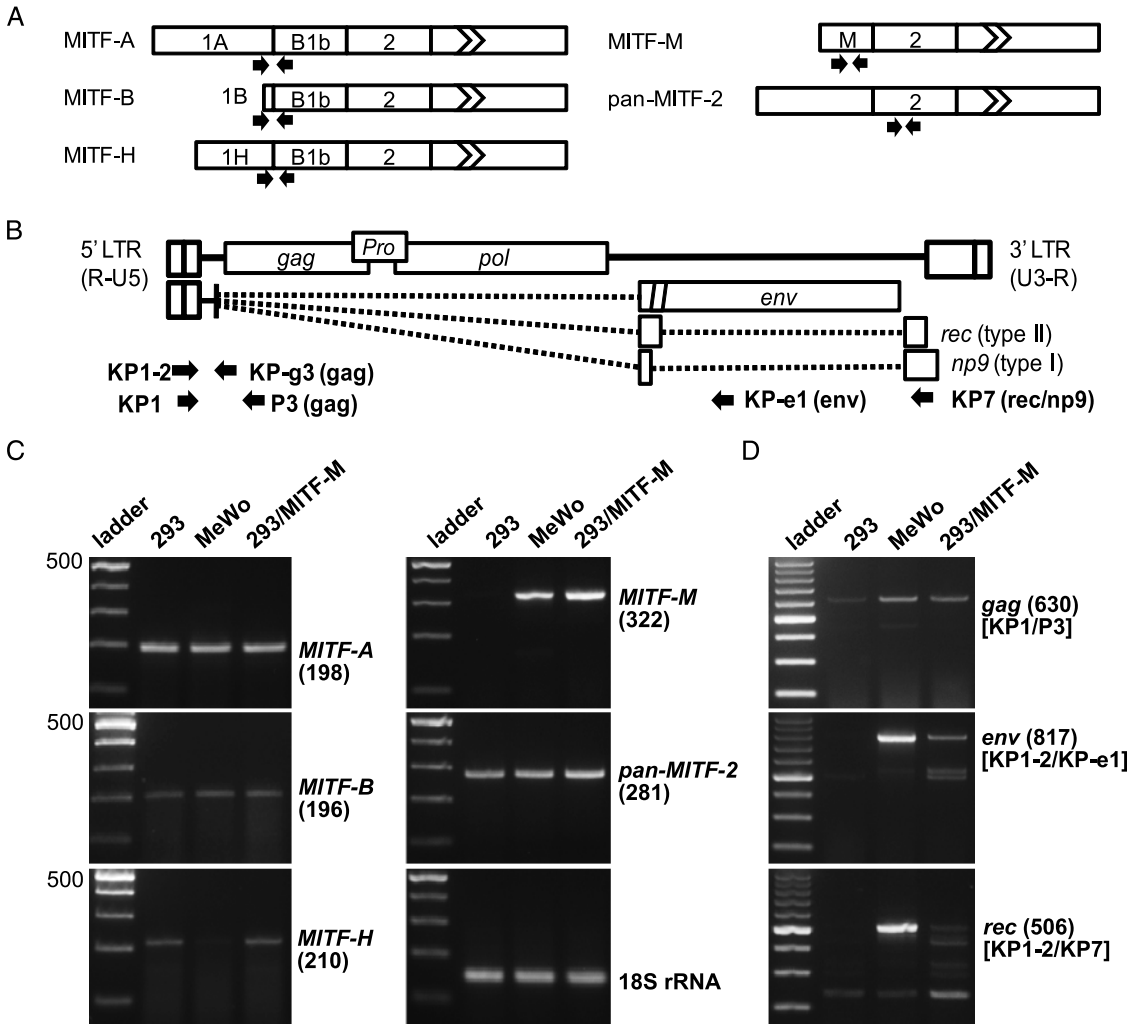

Figure 6. Induction of chromosomal HERV-K expression by forced expression of MITF-M. (A) MITF variant mRNAs with 5′ terminal isoform-specific exons are shown in ribbon-like structures. Exons 3 to 9 were compressed in the finned box. Positions of the forward and reverse primers are indicated by arrows. The pan-MITF-2 primer pairs were designed to detect exon 2 shared by the known MITF variants. (B) Structure of the *gag*-encoding full-length HERV-K RNA and spliced mRNAs encoding *env* and *rec*/*np9*. Splicing events are indicated by dotted lines. Arrows show positions of the primers used in RT-PCR. (C) Results of the RT-PCR experiments for MITF mRNAs. HEK293 (293), MeWo, and MITF-M transfected HEK293 (293/MITF-M) were analyzed. Position of the 500-bp marker and sizes (in bp) of the amplified products are indicated in parentheses. The control reaction for 18S ribosomal RNA is also shown. (D) RT-PCR analyzing the *gag*, *env*, and *rec* mRNAs. Primers applied are shown in brackets.

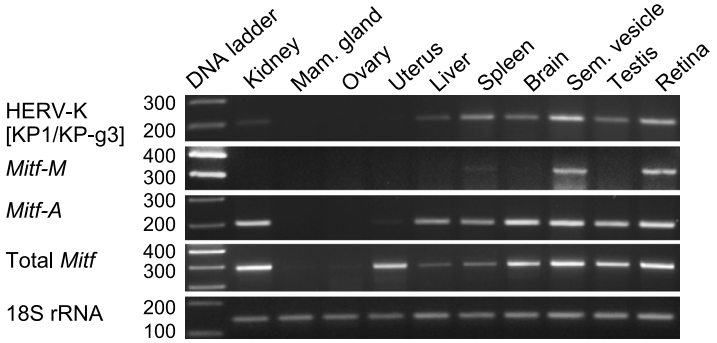

Figure 7. Expression of the older version HERV-K and MITF in rhesus monkey tissues. Rhesus monkey (*Macaca mulatta*) tissues including the mammary gland (Mam. gland) and seminal vesicle (Sem. vesicle) were examined by RT-PCR for expression of the rhesus HERV-K–like elements and the MITF variants. *Macaca* (Mac) 18S rRNA was also analyzed for standardization of the RNA amounts.

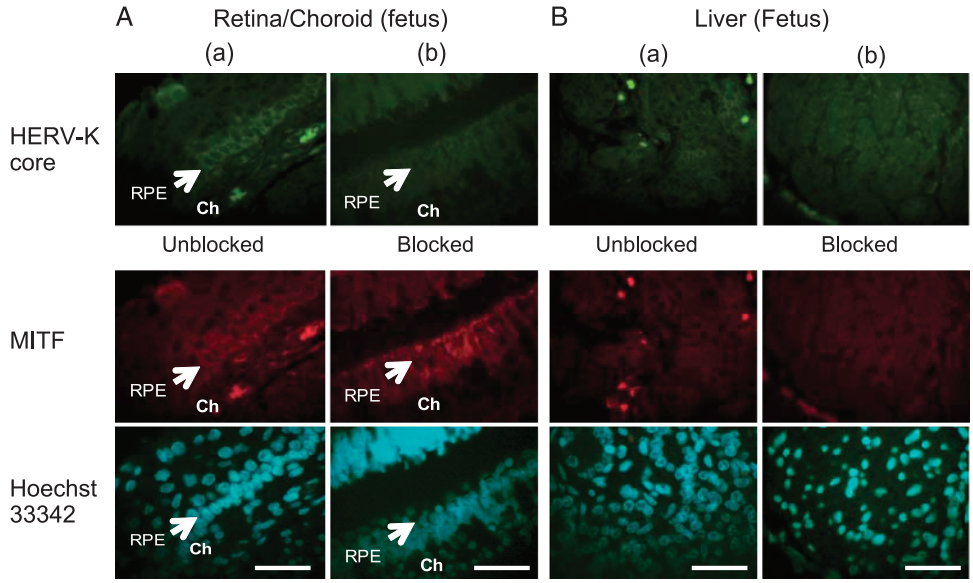

Figure 8. Detection of MITF and HERV-K in human RPE. A normal fetus tissue array (BE01015) was labeled for the MITF (red) and HERV-K core (green) proteins. Unblocked (column a) and blocked (column b) anti–HERV-K core (HERM-1831) IgG was used in combination with anti-MITF IgG (ab20663). Uncropped images of the retina/choroid and liver sections are shown. RPE (arrow) and the choroid (Ch) are indicated. Mild nuclear staining with Hoechst 33342 is also shown. Scale bar, 50 μm.

basal-level expression [1]. By focusing on the high-level expression of HERV-K in melanoma, we identified MITF-M as an inducer of the HERV-K LTR. The arrangement of MITF binding motifs (MITF-1, -2, and -3), 793TATA, and Inr826 (Figure 4*B*) may compose a typical enhancer/promoter structure for RNA polymerase II as found in all of the retroviral LTRs studied so far [48]. Because the 5′ end of the R region is defined by the Inr, the U3 region the HERV-K LTR is thought to span from no. 1 to no. 825. This region is long enough to recruit other regulatory proteins, which may cooperate with the initiation factors associated with the core sequences in melanomas.

In normal melanocytes under the calcium/serum–free conditions, MITF-M was highly expressed without inducing HERV-K transcription. Interestingly, serum stimulation enhanced HERV-K transcription in these cells. The transactivating function of MITF-M is known to be enhanced on Ser73 phosphorylation by MAP kinase in response to c-kit [28]. Varied signaling molecules that control MITF-M have been reported [25,45,46]. Melanocytes may gain the constitutive MITF-M–activating functions during the malignant transformation. HEK293 cells would be able to activate it but only if it is synthesized in the cells.

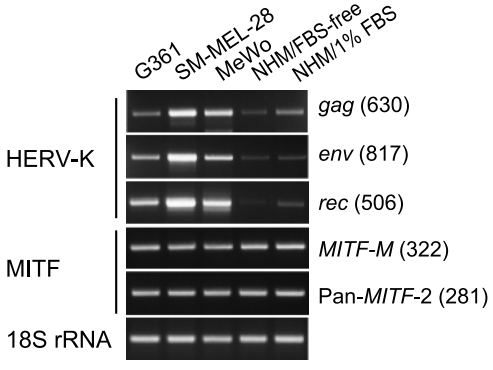

Figure 9. Expression of MITF and HERV-K mRNAs in normal and transformed melanocytes. Semiquantitative RT-PCR was performed with RNA purified from normal human melanocytes (NHM/FBS–free) and melanoma lines, G361, SK-MEL-28, and MeWo. NHM cultured with 1% FBS for 9 hours (NHM/1% FBS) was also analyzed. *MITF-M*, total *MITF*, and HERV-K mRNAs (*gag*, *env*, and *rec*), and 18S rRNA were amplified with primers shown in Figure 6.

Our results do not rule out the possibility that other MITF isoforms are involved in HERV-K LTR activation by a direct or indirect interaction with MITF-M. Moreover, the TFE family proteins including TFE3, TFEb, and TFE are closely related to the MITF proteins and also interact with the canonical E-box motif, CA(C/T)GTG, and related sequences [29,49]. *MITF* and melanocyte-specific genes are activated in breast cancers [50], which might contribute to the HERV-K expression in the cell type [21,51].

Retroviruses trigger oncogenesis by varied mechanisms including 1) virus-encoded oncogene(s) and 2) promoter/enhancer insertion upstream of tumor-inducing genes [48]. The *rec*/*np9* genes of HERV-K are expressed through double splicing as found in the *tax* gene, an oncogene of human T-cell leukemia virus [48]. The Rec/Np9 proteins also seemed to exert an oncogenic function through the interaction with cellular zinc-finger proteins [11–13]. In addition, MITF-M activates not only the 5′ LTR that controls the viral gene expression but also the 3′ LTR, which could activate cell growth/survival genes downstream of the HERV-K loci. Melanoma metastasis involves varied mechanisms including integrin α_4 expression [52], chromosomal instability [53], microRNA expression [54], and loss of AP-2 [55]. HERV-K may possibly influence the processes of melanoma progression through the MITF-M–mediated LTR activation and/or Rec/Np9 production. Recent studies on HERV-K also support this hypothesis [18–20].

Acknowledgments

The authors thank Shigeaki Shibahara (Tohoku University) for the MITF-M clone and Heui-Soo Kim (Pusan National University) for discussion.

References

[1] Bannert N and Kurth R (2004). Retroelements and the human genome: new perspectives on an old relation. *Proc Natl Acad Sci USA* **101**(suppl 2), 14572–14579.

[2] Tristem M (2000). Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the Human Genome Mapping Project Database. *J Virol* **74**, 3715–3730.

[3] Frendo J-L, Olivier D, Cheynet V, Blond J-L, Bouton O, Vidaud M, Rabreau M, Evain-Brion D, and Mallet F (2003). Direct involvement of HERV-W env glycoprotein in human trophoblast cell fusion and differentiation. *Mol Cell Biol* **23**, 3566–3574.

[4] Mallet F, Bouton O, Prudhomme S, Cheynet V, Oriol G, Bonnaud B, Lucotte G, Duret L, and Mandrand B (2004). The endogenous retroviral locus *ERVWE1* is a bona fide gene involved in hominoid placental physiology. *Proc Natl Acad Sci USA* **101**, 1731–1736.

[5] Antony JM, van Marle G, Opii W, Butterfield DA, Mallet F, Yong VW, Wallace JL, Deacon RM, Warren K, and Power C (2004). Human endogenous retrovirus glycoprotein–mediated induction of redox reactants causes oligodendrocyte death and demyelination. *Nat Neurosci* **7**, 1088–1095.

[6] Jha AR, Pillai SK, York VA, Sharp ER, Storm EC, Wachter DJ, Martin JN, Deeks SG, Rosenberg MG, Nixon DF, et al. (2009). Cross-sectional dating of novel haplotypes of HERV-K 113 and HERV-K 115 indicate these proviruses originated in Africa before *Homo sapiens*. *Mol Biol Evol* **26**, 2617–2626.

[7] Dewannieux M, Harper F, Richaud A, Letzelter C, Ribet D, Pierron G, and Heidmann T (2006). Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. *Genome Res* **16**, 1548–1556.

[8] Mayer J, Sauter M, Racz A, Scherer D, Mueller-Lantzsch N, and Meese E (1999). An almost-intact human endogenous retrovirus K on human chromosome 7. *Nat Genet* **21**, 257–258.

[9] Buscher K, Trefzer U, Hofmann M, Sterry W, Kurth R, and Denner J (2005). Expression of human endogenous retrovirus K in melanomas and melanoma cell lines. *Cancer Res* **65**, 4172–4180.

[10] Armbruester V, Sauter M, Krautkraemer E, Meese E, Kleiman A, Best B, Roemer K, and Mueller-Lantzsch N (2002). A novel gene from the human endogenous retrovirus K expressed in transformed cells. *Clin Cancer Res* **8**, 1800–1807.

[11] Denne M, Sauter M, Armbruester V, Licht JD, Roemer K, and Mueller-Lantzsch N (2007). Physical and functional interactions of human endogenous retrovirus proteins Np9 and rec with the promyelocytic leukemia zinc finger protein. *J Virol* **81**, 5607–5616.

[12] Kaufmann S, Sauter M, Schmitt M, Baumert B, Best B, Boese A, Roemer K, and Mueller-Lantzsch N (2007). Human endogenous retrovirus protein Rec interacts with the testicular zinc-finger protein and androgen receptor. *J Gen Virol* **91**, 1494–1502.

[13] Boese A, Sauter M, Galli U, Best B, Herbst H, Mayer J, Kremmer E, Roemer K, and Mueller-Lantzsch N (2000). Human endogenous retrovirus protein cORF supports cell transformation and associates with the promyelocytic leukemia zinc finger protein. *Oncogene* **19**, 4328–4336.

[14] Muster T, Waltenberger A, Grassauer A, Hirschl S, Caucig P, Romirer I, Fodinger D, Seppele H, Schanab O, Magin-Lachmann C, et al. (2003). An endogenous retrovirus derived from human melanoma cells. *Cancer Res* **63**, 8735–8741.

[15] Lower R, Lower J, Tondera-Koch C, and Kurth R (1993). A general method for the identification of transcribed retrovirus sequences (R-U5 PCR) reveals the expression of the human endogenous retrovirus loci HERV-H and HERV-K in teratocarcinoma cells. *Virology* **192**, 501–511.

[16] Patience C, Simpson GR, Colletta AA, Welch HM, Weiss RA, and Boyd MT (1996). Human endogenous retrovirus expression and reverse transcriptase activity in the T47D mammary carcinoma cell line. *J Virol* **70**, 2654–2657.

[17] Bieda K, Hoffmann A, and Boller K (2001). Phenotypic heterogeneity of human endogenous retrovirus particles produced by teratocarcinoma cell lines. *J Gen Virol* **82**, 591–596.

[18] Reiche J, Pauli G, and Ellerbrok H (2010). Differential expression of human endogenous retrovirus K transcripts in primary human melanocytes and melanoma cell lines after UV irradiation. *Melanoma Res* **20**, 435–440.

[19] Schanab O, Humer J, Gleiss A, Mikula M, Sturlan S, Grunt S, Okamoto I, Muster T, Pehamberger H, and Waltenberger A (2011). Expression of human endogenous retrovirus K is stimulated by ultraviolet radiation in melanoma. *Pigment Cell Melanoma Res* **24**, 656–665.

[20] Serafino A, Balestrieri E, Pierimarchi P, Matteucci C, Moroni G, Oricchio E, Rasi G, Mastino A, Spadafora C, Garaci E, et al. (2009). The activation of human endogenous retrovirus K (HERV-K) is implicated in melanoma cell malignant transformation. *Exp Cell Res* **315**, 849–862.

[21] Ono M, Kawakami M, and Ushikubo H (1987). 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.

[22] Golan M, Hizi A, Resau JH, Yaal-Hahoshen N, Reichman H, Keydar I, and Tsarfaty I (2008). Human endogenous retrovirus (HERV-K) reverse transcriptase as a breast cancer prognostic marker. *Neoplasia* **10**, 521–533.

[23] Lavie L, Kitova M, Maldener E, Meese E, and Mayer J (2005). CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2). *J Virol* **79**, 876–883.

[24] Fuchs NV, Kraft M, Tondera C, Hanschmann K-M, Lower J, and Lower R (2011). Expression of the human endogenous retrovirus group HML-2/HERV-K does not depend on canonical promoter elements but is regulated by the transcription factors Sp1 and Sp3. *J Virol* **85**, 3436–3448.

[25] Cheli Y, Ohanna M, Ballotti R, and Bertolotto C (2009). Fifteen-year quest for microphthalmia-associated transcription factor target genes. *Pigment Cell Melanoma Res* **23**, 27–40.

[26] Levy C, Khaled M, Robinson KC, Veguilla RA, Chen PH, Yokoyama S, Makino E, Lu J, Larue L, Beermann F, et al. (2010). Lineage-specific transcriptional regulation of DICER by MITF in melanocytes. *Cell* **141**, 994–1005.

[27] Levy C, Khaled M, and Fisher DE (2006). MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol Med* **12**, 406–414.

[28] Hemesath TJ, Price ER, Takemoto C, Badalian T, and Fisher DE (1998). MAP kinase links the transcription factor microphthalmia to c-Kit signalling in melanocytes. *Nature* **391**, 298–301.

[29] Shibahara S, Takeda K, Yasumoto K, Udono T, Watanabe K, Saito H, and Takahashi K (2001). Microphthalmia-associated transcription factor (MITF): multiplicity in structure, function, and regulation. *J Investig Dermatol Symp Proc* **6**, 99–104.

[30] Bertolotto C, Bille K, Ortonne JP, and Ballotti R (1996). Regulation of tyrosinase gene expression by cAMP in B16 melanoma cells involves two CATGTG motifs surrounding the TATA box: implication of the microphthalmia gene product. *J Cell Biol* **134**, 747–755.

[31] Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, Beroukhim R, Milner DA, Granter SR, Du J, et al. (2005). Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117–122.

[32] Hershey CL and Fisher DE (2005). Genomic analysis of the microphthalmia locus and identification of the MITF-J/Mitf-J isoform. *Gene* **347**, 73–82.

[33] Okuyama T, Kurata S, Tomimori Y, Fukunishi N, Sato S, Osada M, Tsukinoki K, Jin HF, Yamashita A, Ito M, et al. (2008). p63(TP63) elicits strong trans-activation of the *MFG-E8/lactadherin/BA46* gene through interactions between the TA and DeltaN isoforms. *Oncogene* **27**, 308–317.

[34] Fukunishi N, Katoh I, Tomimori Y, Tsukinoki K, Hata R, Nakao A, Ikawa Y, and Kurata S (2010). Induction of DeltaNp63 by the newly identified keratinocyte-specific transforming growth factor beta signaling pathway with Smad2 and IκB kinase α in squamous cell carcinoma. *Neoplasia* **12**, 969–979.

[35] Reus K, Mayer J, Sauter M, Scherer D, Muller-Lantzsch N, and Meese E (2001). Genomic organization of the human endogenous retrovirus HERV-K (HML-2.HOM) (ERVK6) on chromosome 7. *Genomics* **72**, 314–320.

[36] Watson JD, Baker TA, Bell SP, Gann A, Levine M, and Losick R (2004). Molecular Biology of the Gene (5th ed). Benjamin Cummings (Cold Spring Harbor Laboratory Press), San Francisco, CA. Chapter 2: Mechanisms of Transcription. pp. 347–377.

[37] Patikoglou GA, Kim JL, Sun L, Yang SH, Kodadek T, and Burley SK (1999). TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev* **13**, 3217–3230.

[38] Humer J, Waltenberger A, Grassauer A, Kurz M, Valencak J, Rapberger R, Hahn S, Lower R, Wolff K, Bergmann M, et al. (2006). Identification of a melanoma marker derived from melanoma-associated endogenous retroviruses. *Cancer Res* **66**, 1658–1663.

[39] Wang Y, Radfar S, Liu S, Riker A, and Khong H (2010). Mitf-Mdel, a novel melanocyte/melanoma–specific isoform of microphthalmia-associated transcription factor-M, as a candidate biomarker for melanoma. *BMC Med* **8**, 14.

[40] Davis IJ, Kim JJ, Ozsolak F, Widlund HR, Rozenblatt-Rosen O, Granter SR, Du J, Fletcher JA, Denny CT, Lessnick SL, et al. (2006). Oncogenic MITF dysregulation in clear cell sarcoma: defining the MiT family of human cancers. *Cancer Cell* **9**, 473–484.

[41] Lo K and Smale ST (1996). Generality of a functional initiator consensus sequence. *Gene* **182**, 13–22.

[42] Javahery R, Khachi A, Lo K, Zenzie-Gregory B, and Smale ST (1994). DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol Cell Biol* **14**, 116–127.

[43] Leung CS and Srigley JR (1995). Distribution of lipochrome pigment in the prostate gland: biological and diagnostic implications. *Hum Pathol* **26**, 1302–1307.

[44] Das G, Choi Y, Sicinski P, and Levine E (2009). Cyclin D1 fine-tunes the neurogenic output of embryonic retinal progenitor cells. *Neural Dev* **4**, 15.

[45] Shah M, Bhoumik A, Goel V, Dewing A, Breitwieser W, Kluger H, Krajewski S, Krajewska M, Dehart J, Lau E, et al. (2010). A role for ATF2 in regulating MITF and melanoma development. *PLoS Genet* **6**, e1001258.

[46] Sonnenblick A, Levy C, and Razin E (2004). Interplay between MITF, PIAS3, and STAT3 in mast cells and melanocytes. *Mol Cell Biol* **24**, 10584–10592.

[47] Ling J, Pi W, Bollag R, Zeng S, Keskintepe M, Saliman H, Krantz S, Whitney B, and Tuan D (2002). The solitary long terminal repeats of ERV-9 endogenous retrovirus are conserved during primate evolution and possess enhancer activities in embryonic and hematopoietic cells. *J Virol* **76**, 2410–2423.

[48] Coffin JM, Hughes SH, and Varmus HE (1997). *Retroviruses*. Cold Spring Harbor Laboratory Press, New York, NY.

[49] Verastegui C, Bertolotto C, Bille K, Abbe P, Ortonne JP, and Ballotti R (2000). TFE3, a transcription factor homologous to microphthalmia, is a potential transcriptional activator of *tyrosinase* and *TyrpI* genes. *Mol Endocrinol* **14**, 449–456.

[50] Montel V, Suzuki M, Galloy C, Mose ES, and Tarin D (2009). Expression of melanocyte-related genes in human breast cancer and its implications. *Differentiation* **78**, 283–291.

[51] Wang-Johanning F, Radvanyi L, Rycaj K, Plummer JB, Yan P, Sastry KJ, Piyathilake CJ, Hunt KK, and Johanning GL (2008). Human endogenous retrovirus K triggers an antigen-specific immune response in breast cancer patients. *Cancer Res* **68**, 5869–5877.

[52] Rebhun RB, Cheng H, Gershenwald JE, Fan D, Fidler IJ, and Langley RR (2010). Constitutive expression of the α4 integrin correlates with tumorigenicity and lymph node metastasis of the B16 murine melanoma. *Neoplasia* **12**, 173–182.

[53] Silva AG, Graves HA, Guffei A, Ricca TI, Mortara RA, Jasiulionis MG, and Mai S (2010). Telomere-centromere–driven genomic instability contributes to karyotype evolution in a mouse model of melanoma. *Neoplasia* **12**, 11–19.

[54] Penna E, Orso F, Cimino D, Tenaglia E, Lembo A, Quaglino E, Poliseno L, Haimovic A, Osella-Abate S, De Pitta C, et al. (2011). microRNA-214 contributes to melanoma tumour progression through suppression of TFAP2C. *EMBO J* **30**, 1990–2007.

[55] Bar-Eli M (1999). Role of AP-2 in tumor growth and metastasis of human melanoma. *Cancer Metastasis Rev* **18**, 377–385.

Figure W1. Nucleotide sequences of the 5′ RACE clones obtained from MeWo cells. (A) Clones indicating transcriptional initiation at nucleotide no. 796 are aligned with the LTR-leader sequences (HERVK-LTR-Leader) of HML-2.HOM (GenBank AF074086.2). Nucleotides identical to the HERVK-LTR-leader sequences are in blue, and mismatches in magenta. The 3′ end of the LTR is marked by a line. (B) Clones corresponding to the transcripts starting at no. 826 are shown. The first 6 clones of 11 are shown. (C) Alignment of the 5′ RACE products indicating transcriptional initiation by a non-LTR promoter at 7q22. Corresponding chromosomal sequences (NW_001839071) are also aligned. Regions unrelated to HERV-K are in magenta.

Figure W2. Nucleotide sequences of the 5′ RACE products obtained from HEK293 cells. (A) The 5′ RACE clones starting at no. 460 are aligned with the HERV-K LTR-leader sequences. These clones showed 44 nucleotide mismatches (only 92% sequence identity) compared with the 570 nucleotides of HML-2.HOM. A BLAST search identified the template sequences at a unique locus on chromosome 3 (NT_005612). Nucleotides identical to the HERVK-LTR-leader sequences (GenBank AF074086.2) are in blue and mismatches are in magenta. The 3′ end of the LTR is marked by a line. (B) Clones starting at no. 813 or no. 815 are aligned. They show approximately 99% sequence identity to the HERV-K LTR-Leader and thought to represent the transcripts from HML-2.HOM and related HERV-K loci.

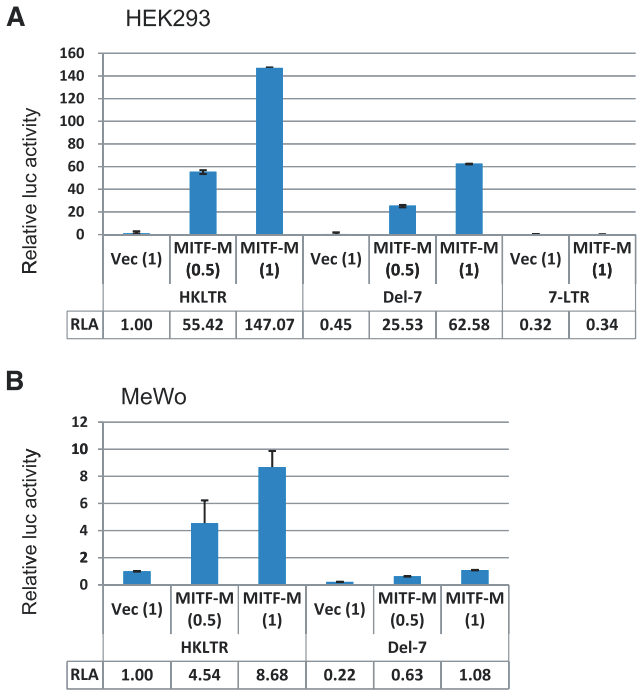

Figure W3. Transactivation of HERV-K LTR by MITF-M. Luc expression assay with HKLTR-1, Del-7, and 7-LTR in combination with the MITF-M expression vector. Relative luciferase activity (RLA) (mean ± SD) is presented in relation to the control assay (1.00) with HKLTR-1 and the empty vector (Vec). The ratio (0.5, 1) of the activator plasmid amount to the Luc plasmid amount is given in parentheses. Experiments were performed two times in triplicate.

Table W1. Human Genomic LTRs with Less Than 95% Sequence Identity to the HERV-K LTR (HML-2.HOM).

Chromo-some	Genomic Contig	Position	Inr460	MITF-3	MITF-2	MITF-1	793TATA	Inr826	Identity (≧95%)	Gaps	Score (bits)
7	Template HKLTR-1	#1	GGAATGC	CACATG	CTTGTG	CACATC	AATAAATA	GGATCCT			
7	NT_079592.2	4636386	GGAATGC	CACATG	CTTGTG	CACATC	AATAAATA	GGATCCT	944/945 (99%)	1/945 (0%)	1696
		4627883	GGAATGC	CACATG	CTTGTG	CACATC	AGTAAATA	GGATCCT	942/945 (99%)	1/945 (0%)	1687
		4619379	GGAATGC	CACATG	CTTGTG	CACATC	AATAAATA	GGATCCT	938/945 (99%)	1/945 (0%)	1669
		16240452	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	934/945 (98%)	1/945 (0%)	1651
		23081303	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	933/945 (98%)	1/945 (0%)	1647
12	NT_029419.11	20865515	GGAATGC	CACATG	CTTGTG	CACATC	AATAAATA	GGATCCT	936/945 (99%)	1/945 (0%)	1660
		20874004	GGAATGT	CACATG	CTTGTG	CACATC	AATAAATA	GGATCCT	932/945 (98%)	1/945 (0%)	1642
		13992344	GGAATGT	CACGTG	CTTGTG	CACATC	AATAAATA	GGCTCCT	929/945 (98%)	1/945 (0%)	1629
		17870521	GGAATGT	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	929/946 (98%)	1/945 (0%)	1622
		18938415	GGAATGT	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	925/945 (97%)	1/945 (0%)	1609
7	NT_007741.13	2904345	GGAATGC	CACATG	CTTGTG	CACATC	AATAAATA	GGATCCT	944/945 (99%)	1/945 (0%)	1660
7	NT_079596.2	57285861	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	935/945 (98%)	1/945 (0%)	1656
		24287145	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	930/945 (98%)	1/945 (0%)	1633
		22844882	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	929/945 (98%)	1/945 (0%)	1629
		25232346	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	926/946 (97%)	2/946 (0%)	1609
		3784161	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	921/946 (97%)	3/946 (0%)	1582
		188057	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	906/946 (95%),	2/946 (0%)	1519
		12386954	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	906/946 (95%)	11/946 (1%)	1517
8	NT_007995.14	7372243	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	934/945 (98%)	1/945 (0%)	1651
		13916038	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	912/945 (96%)	1/945 (0%)	1552
		12972195	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	903/947 (95%)	3/947 (0%)	1499
Y	NT_011903.12	1088725	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	934/945 (98%)	1/945 (0%)	1651
		2722474	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	934/945 (98%)	1/945 (0%)	1651
		3337067	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	934/945 (98%)	1/945 (0%)	1651
19	NT_011109.15	10625035	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	934/945 (98%)	1/945 (0%)	1651
		17365636	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	926/945 (97%)	1/945 (0%)	1615
		10090511	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	931/959 (97%)	15/959 (1%)	1608
		397710	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	930/972 (95%)	28/972 (2%)	1581
		24814575	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	914/946 (96%)	2/946 (0%)	1555
		10390031	GGAATGC	CACGTG	CTTGTG	CATATC	AATAAATA	AGATCCT	909/945 (96%)	7/945 (0%)	1539
		21661083	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	905/945 (95%)	6/945 (0%)	1521
		1449405	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	AGATCCT	904/946 (95%)	2/946 (0%)	1510
		24680600	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	902/945 (95%)	3/945 (0%)	1499
10	NT_008705.15	9157724	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	933/945 (98%)	1/945 (0%)	1647
16	NT_037887.4	5745357	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	933/945 (98%)	1/945 (0%)	1647
		8169731	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	925/945 (97%)	1/945 (0%)	1611
8	NT_008183.18	6798560	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	933/945 (98%)	1/945 (0%)	1647
		9966476	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	911/945 (96%)	11/945 (1%)	1543
4	NT_006216.14	1502918	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	933/945 (98%)	1/945 (0%)	1647
		6812371	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	933/945 (98%)	1/945 (0%)	1647
		9985448	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	911/945 (96%)	11/945 (1%)	1543
16	NT_010498.15	28448457	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	932/945 (98%)	1/945 (0%)	1642
		1512569	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	928/945 (98%)	1/945 (0%)	1624
1	NT_004487.18	6639369	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	932/945 (98%)	1/945 (0%)	1642

Table W1. (*continued*).

		6087779	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	930/945 (98%)	1/945 (0%)	1633
		6095991	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	930/945 (98%)	1/945 (0%)	1633
		6060651	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	906/946 (95%)	2/946 (0%)	1519
		10227941	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	AGATCCT	903/945 (95%)	1/945 (0%)	1512
2	NT_022184.14	6498779	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	932/945 (98%)	1/945 (0%)	1642
		16268394	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	931/945 (98%)	1/945 (0%)	1638
		9653238	GGAATGC	CACATG	CTTGTG	CACATC	AATAAATA	GGATCCT	930/945 (98%)	1/945 (0%)	1633
		65304362	GGCATGC	CACGTG	GGCATG	CACATC	AATAAATA	GGATCCT	899/946 (95%)	2/946 (0%)	1487
X	NT_011786.15	9124856	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	931/945 (98%)	1/945 (0%)	1638
4	NT_016354.18	86128085	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	931/945 (98%)	1/945 (0%)	1638
		44812799	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	929/945 (98%)	3/945 (0%)	1626
11	NT_033899.7	5136708	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	931/945 (98%)	1/945 (0%)	1626
		5128210	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	929/945 (98%)	12/945 (1%)	1629
		22163299	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	909/946 (96%)	3/946 (0%)	1528
		22155106	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	908/946 (95%)	4/946 (0%)	1521
8	NT_023736.16	7343774	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	931/945 (98%)	1/945 (0%)	1638
		7352269	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	923/945 (97%)	9/945 (0%)	1604
3	NT_029928.12	275097	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	931/945 (98%)	1/945 (0%)	1638
3	NT_022517.17	47241028	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	930/945 (98%)	1/945 (0%)	1633
		14073653	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	928/945 (98%)	1/945 (0%)	1624
		53952548	GGAATGC	CACATG	CTTGTG	CACATC	AATAAATA	GGATCCT	919/945 (97%)	1/945 (0%)	1584
		50497327	GGAATGC	CACGTG	CTTGTG	CACATC	CATAAATA	GGATCCT	909/945 (96%)	12/945 (1%)	1535
		23527120	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	900/946 (95%)	8/946 (0%)	1489
7	NT_007933.14	50045767	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	930/945 (98%)	1/945 (0%)	1633
		48604326	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	928/945 (98%)	1/945 (0%)	1624
		50991790	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	926/945 (97%)	1/945 (0%)	1615
		29571945	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	923/945 (97%)	2/945 (0%)	1599
		25977162	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	906/946 (95%)	2/945 (0%)	1519
		38166108	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	906/946 (95%)	11/946 (1%)	1517
17	NT_010799.14	3765019	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	930/945 (98%)	1/945 (0%)	1633
		9204418	GGCATGC	CACATG	CTTGTG	CACATC	AATAAATA	GGATCCT	913/949 (96%)	5/949 (0%)	1544
9	NT_008470.18	31512589	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	930/945 (98%)	1/945 (0%)	1633
		18678098	GGAATGT	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	919/945 (97%)	9/945 (0%)	1586
6	NT_007299.12	26911478	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	930/945 (98%)	1/945 (0%)	1633
		17388676	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	927/943 (98%)	1/943 (0%)	1626
		16256254	GGAATGC	CACATG	CTTGTG	CACATC	AATAAATA	GGATCCT	925/945 (97%)	9/945 (0%)	1613
		16247792	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	924/945 (97%)	9/945 (0%)	1609
		31703257	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	920/944 (97%)	3/944 (0%)	1588
20	NT_011362.9	5652451	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	930/945 (98%)	1/945 (0%)	1633
12	NT_009775.16	1577352	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	929/945 (98%)	1/945 (0%)	1629
		9113298	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	909/946 (96%)	11/946 (1%)	1532
14	NT_026437.11	46446026	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	929/945 (98%)	1/945 (0%)	1629
		36491018	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	910/945 (96%)	15/945 (1%)	1522
		59261730	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	902/946 (95%)	10/946 (1%)	1503
		19587055	GGCATGC	CACGTG	CTTGTG	CACATC	AGTAAATA	GGATCCT	902/946 (95%)	2/946 (0%)	1501
		4194701	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	900/945 (95%)	1/945 (0%)	1498
3	NT_005612.15	91776453	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	929/945 (98%)	1/945 (0%)	1629
		91784665	**GGCATGC**	**CACGTG**	**CTTGTG**	**CACATC**	**AATAAATA**	**GGATCCT**	928/945 (98%)	1/945 (0%)	**1624**
		19239229	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	927/945 (98%)	9/945 (0%)	1622

Table W1. (*continued*).

		19247428	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	927/945 (98%)	9/945 (0%)	1622
		82119455	GGCATGC	CACATG	CTTGTG	CACATC	AATAAATA	GGATCCT	915/945 (96%)	1/945 (0%)	1566
		93106159	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	910/945 (96%),	2/945 (0%)	1539
		93100773	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	905/946 (95%)	10/946 (1%)	1517
		7918791	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	904/946 (95%)	2/946 (0%	1510
		7914037	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	903/947 (95%	4/947 (0%)	1496
1	NT_004559.13	729687	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	929/945 (98%)	1/945 (0%)	1629
		4259163	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	901/946 (95%)	2/946 (0%)	1496
1	NT_032977.8	45814690	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	929/945 (98%)	1/945 (0%)	1629
		45820095	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	929/945 (98%)	1/945 (0%)	1629
		63715225	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	928/945 (98%)	1/945 (0%)	1624
		22443836	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	920/945 (97%)	2/945 (0%)	1584
		36863376	GGAATGC	CATGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	917/945 (97%)	9/945 (0%)	1577
		15965693	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	907/943 (96%)	2/943 (0%)	1532
		59505425	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	903/946 (95%)	10/946 (1%)	1508
5	NT_023133.12	895242	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	929/945 (98%)	1/945 (0%)	1629
		903454	GGAATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	929/945 (98%)	1/945 (0%)	1629
		23751517	GGCATGC	CATGTG	CTTGTG	CACATC	AATAAATG	GGATCCT	920/945 (97%)	1/945 (0%)	1588
		14234396	GGAATGT	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	917/945 (97%)	9/945 (0%)	1577
		25063855	GGCATGC	CACGTG	CTTGTG	CACATC	AATAAATA	GGATCCT	898/945 (95%)	5/945 (0%)	1483

Sequences identical to the corresponding positions of the HERV-K LTR (HML-2.HOM) are in black. Altered sequences that seemed to preserve the proposed functions are in blue. Nucleotide substitutions expected to disrupt the function are in red.

Table W2. Rhesus Monkey (*Macaca mulatta*) LTRs with Less Than 80% Sequence Identity to the Human-type HERV-K LTR.

		Inr460	MITF-3	MITF-2	MITF-1	793TATA	Inr826				
Template HKLTR-1	#1	GGAATGC	CACATG	CTTGTG	CACATC	AATAAATA	Insertion of Inr consensus, PyPyAN(T/A)PyPy	GGATCCT			
M. mulatta genomic reference	Position								Identity (≧80%)	Gaps	Score (bits)
AC210649.4	68871	G——GC	CACATG	CCTATT	CACATT	AATAAATA	CCATTGC	GGGTCCT	800/957 (84%)	30/957 (3%)	996
AC206102.3	78780	GG--TCT	CACATT	CCTATT	CACATC	AATAAATA		GGGTCCT	749/922 (82%)	27/922 (2%)	836
AC213330.3	173369	GG--TCT	CACATT	CCTATT	CACATC	AATAAATA		GGGTCCT	749/922 (82%)	27/922 (2%)	836
AC191812.4	161342	AGAAGGC	CACATG	CCTATT	CACTTC	AATAAATA		GGGTCCT	774/967 (81%),	47/967 (4%)	836
AC197502.3	156174	G——GC	TACATG		TTCATC	AACAAATA		CCATCCT	775/962 (81%)	50/962 (5%)	834
	165784	G——TC	CACATG	CCTATT	CACTTC	AGCAAATA		AGGTCCT	756/950 (80%)	46/950 (4%)	821
AC218138.3	19882	AGAAGGC	CACATG	CTATTG	CACTTC	AATAAATA		CGAGTCC	776/968 (81%)	35/968 (3%)	827
AC209790.2	102901	AGAAGGC	CCTATT	CCTATT	CACTTC	AATAAATA		CGAGTCC	776/968 (81%)	35/968 (3%)	827
AC201622.7	131721	G——GC	CACATG	CCTATT	CACATC	AATAAATA			687/831 (83%)	34/831 (4%)	827
AC148674.1	31678	GG——CC	CACATG	CCTATA	CACATC	AATAAATA		GGGTCCT	781/980 (80%)	45/980 (4%)	821
AC194579.7	213314	GG-CGGC	CACGTT	CCTATG	CACATC	AATAAATA	TCAGTGC	AGGTCCT	782/980 (80%)	48/980 (4%)	820
	195164	GG-CGGC	CACGTT	CCTATG	CACATC	AATAAATA	CCAGTGC	GG-TCCT	780/981 (80%)	53/981 (5%)	807
AC191828.3	148275	GG-CGGC	CACGTT	CCTATG	CACATC	AATAAATA	TCAGTGC	AGGTCCT	782/980 (80%)	48/980 (4%)	820
AB128049.1	2710708	AGAAGGC	CATATG	CCTATT	CACTTC	AATAAATA		GGGTCCT	771/970 (80%)	59/970 (6%)	820
	3070098	AGAAGGC	CATATG	CCTATT	CACTTC	AATAAATA		GGGTCCT	770/967 (80%)	55/967 (5%)	816
	3159304	AGAAGGC	CATATG	CCTATT	CACTTC	AATAAATA	CCAGCTT	GGGTCCT	766/967 (80%)	54/967 (5%)	801
	2817651	AGAAGGC	CATATG	CCTATT	CACTTC	AATAAATA	CCAGTGC	GGGTCCT	766/967 (80%)	57/967 (5%)	794
AC199608.6	115869	AGAAGGC	CATATG	CCTATT	CACTTC	AATACATA	CCAGTGC	GGGCCCT	769/967 (80%)	50/967 (5%)	816
AC210647.2	87803	AGAAGGC	CATATG	CCTATT	CACTTC	AATAAATA		GGGTCCT	770/969 (80%)	60/969 (6%)	814
AC148684.1	87804	GGAATGT	CATATG	CCTATT	CACTTC	AATAAATA		GGGTCCT	770/969 (80%)	60/969 (6%)	814
AC148680.1	161314	AGAAGGC	CATATG	CCTATT	CACTTC	AATAAATA		GGGTCCT	770/969 (80%)	60/969 (6%)	814
AC193058.7	104908	G——GC	CACATC	CCTATT	CACATC	AATAAATA	CTAGTGC	GGGTC—A	778/981 (80%)	62/981 (6%)	801
AC148690.1	2395	GGAATAT	CATATG	CCTATT	CACTTC	AATAAATA	CCAGCTT	GGGTCCT	766/967 (80%)	54/967 (5%)	801
	94162	AGAAGGC	CATATG	CCTATT	CACTTC	AATAAATA	CCAGCTT	GGGTCCT	766/967 (80%)	54/967 (5%)	801
AC148682.1	106447	AGAAGGC	CATATG	CCTATT	CACTTC	AATAAATA	CCAGCTT	GGGTCCT	766/967 (80%)	54/967 (5%)	801
	198215	AGAAGGC	CATATG	CCTATT	CACTTC	AATAAATA	CCAGCTT	GGGTCCT	766/967 (80%)	54/967 (5%)	801
AC225834.3	157305	AGAAGGC	CATATG	CCTATT	CACTTC	AATAAATA		GGGTCCT	765/965 (80%)	51/965 (5%)	800
AC193521.7	79951	GGCTTCT	CTCATG	CCTATA	CACATC	AGTAAATA	CCAGTCC	GGGTCCT	775/977 (80%)	46/977 (4%)	798
AC214376.3	140057	AGAAGGC	CACATG	CCTATT	CACTTC	AATAAATA		GGGTCCT	755/951 (80%)	56/951 (5%)	792
AC210118.6	102633	AGAAGCC	CATATG	CTATTA	CACTTC	AATAAATA		GGGTCCT	765/966 (80%)	49/966 (5%)	791
AC148703.1	4548	AGAAGGC	CATATG	CCTATT	CACTTC	AATAAATC		GGGTCCT	765/967 (80%)	53/967 (5%)	791
AC148694.1	141943	AGAAGGC	CATATG	CCTATT	CACTTC	AATAAATC		GGGTCCT	765/967 (80%)	53/967 (5%)	791
AC191957.7	14747	AGAAGGC	CATATG	CCTATT	CACTTC	AATAAATA		GGGTCCT	733/927 (80%)	49/927 (5%)	756

Sequences identical to the corresponding positions of the HERV-K LTR (HML-2.HOM) are in black. Altered sequences that seemed to preserve the proposed functions are in blue. Nucleotide substitutions expected to disrupt the function are in red.