

RESEARCH

Open Access

# HERV-K(HML-2) *rec* and *np9* transcripts not restricted to disease but present in many normal human tissues

Katja Schmitt<sup>1,4</sup>, Kristina Heyne<sup>2</sup>, Klaus Roemer<sup>2</sup>, Eckart Meese<sup>1</sup> and Jens Mayer<sup>1,3\*</sup>

## Abstract

**Background:** Human endogenous retroviruses of the HERV-K(HML-2) group have been associated with the development of tumor diseases. Various HERV-K(HML-2) loci encode retrovirus-like proteins, and expression of such proteins is upregulated in certain tumor types. HERV-K(HML-2)-encoded Rec and Np9 proteins interact with functionally important cellular proteins and may contribute to tumor development. Though, the biological role of HERV-K(HML-2) transcription and encoded proteins in health and disease is less understood. We therefore investigated transcription specifically of HERV-K(HML-2) *rec* and *np9* mRNAs in a panel of normal human tissues.

**Results:** We obtained evidence for *rec* and *np9* mRNA being present in all examined 16 normal tissue types. A total of 18 different HERV-K(HML-2) loci were identified as generating *rec* or *np9* mRNA, among them loci not present in the human reference genome and several of the loci harboring open reading frames for Rec or Np9 proteins. Our analysis identified additional alternative splicing events of HERV-K(HML-2) transcripts, some of them encoding variant Rec/Np9 proteins. We also identified a second HERV-K(HML-2) locus formed by L1-mediated retrotransposition that is likewise transcribed in various human tissues.

**Conclusions:** HERV-K(HML-2) *rec* and *np9* transcripts from different HERV-K(HML-2) loci appear to be present in various normal human tissues. It is conceivable that Rec and Np9 proteins and variants of those proteins are part of the proteome of normal human tissues and thus various cell types. Transcription of HERV-K(HML-2) may thus also have functional relevance in normal human cell physiology.

**Keywords:** Human endogenous retrovirus, Provirus, Transcription, Splicing, HERV-K Rec protein, HERV-K Np9 protein, Retrotransposition, L1 element

## Background

Human endogenous retroviruses (HERVs) stem from ancient germ line infections by exogenous retroviruses. About 8% of the human genome mass consists of retroviral sequences *in sensu stricto* and sequences with retroviral portions. There are about 40 phylogenetically distinct HERV groups documenting germ line integration, that is, provirus formations by different ancient exogenous retroviruses millions of years ago. Re-infections and intracellular amplifications often increased numbers

of proviruses per HERV group for limited evolutionary time periods following initial integration events. Most HERV groups no longer encode former retroviral proteins due to long time presence in the genome and thus accumulation of nonsense mutations including smaller and larger indels. Some retroviral proteins, in particular Envelope (Env), have been conserved during evolution to contribute important Env-mediated functions such as fusion of cell membranes [1-4].

The so-called HERV-K(HML-2) group (in short, HML-2) includes a number of evolutionarily young proviruses, some of which formed in the human lineage after the evolutionary split of human from chimpanzee about 6 million years ago. Especially the young HML-2 loci often harbor open reading frames (ORFs) for retroviral proteins such as Gag, Protease, Polymerase, and

\* Correspondence: jens.mayer@uks.eu

<sup>1</sup>Institute of Human Genetics, Center of Human and Molecular Biology, Medical Faculty, University of Saarland, 66424 Homburg/Saar, Germany

<sup>3</sup>Center of Human and Molecular Biology, University of Saarland, 66424 Homburg/Saar, Germany

Full list of author information is available at the end of the article

Envelope. Analyses of HML-2 proviral transcripts had identified typical retroviral splicing events generating an *env* mRNA and a sub-spliced *env* mRNA, originally named *cORF* and later re-named *rec*, with most of the envelope coding sequence removed. Historically, HML-2 proviruses have been divided into type 2 loci, the transcripts of which can be sub-spliced to *rec* mRNA, and type 1 loci that lack a characteristic 292-bp sequence located about 50 bp into the *env* coding sequence [5,6]. Lack of the 292-bp sequence in type 1 loci impairs sub-splicing of *env* mRNA to *rec* mRNA because of lack of the *rec* splice donor (SD) site located within the deleted region. Instead, a SD site just upstream of the 292-bp deletion is now employed in combination with a splice acceptor (SA) site located at the 3' end of *env* that is the same SA for splicing of transcripts from type 1 and type 2 loci. Such spliced transcripts derived from HML-2 type 1 loci have been named *np9* [7,8] (Figure 1).

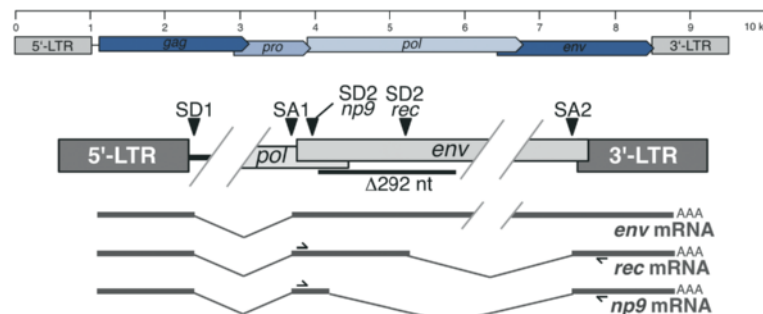
Clinical relevance of HML-2 transcription and proteins has been investigated in the context of various human diseases. Especially germ cell tumors (GCT) display strongly upregulated HML-2 transcription and expression of HML-2 proteins already in early stages of tumor development. GCT patients display strong antibody titers against HML-2 Gag and Env proteins at the time of tumor detection (reviewed in ref. [2]).

Both *rec* and *np9* mRNA can encode proteins with potentially important cellular functions that may be relevant to disease development. HML-2 Rec protein is basically a functional homologue of HIV<sub>Rev</sub> protein [9-13]. Nude mice transgenic for Rec protein develop lesions reminiscent of testicular carcinoma *in situ* [14]. Rec protein was shown to interact with several functionally relevant cellular proteins such as promyelocytic zinc finger protein (PLZF), testicular zinc finger protein (TZFP), Staufen-1, and human small glutamine-rich

tetratricopeptide repeat protein (hSGT). Np9 protein was shown to interact with PLZF and ligand of Numb protein X (LNX). All of those interactions may have important cellular consequences depending on cellular context [15-20].

Several recent studies have identified a number of HML-2 loci transcribed in various disease as well as normal conditions by assigning specifically generated HML-2 cDNA sequences to genomic HML-2 loci employing characteristic sequence differences between HML-2 loci, with transcription patterns varying considerably between conditions (for instance, see [21-26]). Several of the transcribed HML-2 loci can, in principle, encode *rec* or *np9* mRNA. We have previously analyzed HML-2 loci specifically for coding capacity for *rec* mRNA and protein by analysis of HML-2 locus sequences for features required for *rec* mRNA splicing and presence of a Rec ORF within predicted mRNA sequences. We also had identified a number of HML-2 loci generating *rec* mRNA by means of cDNA sequence assignments to genomic HML-2 loci [21,27].

The role(s) of HML-2 Rec and Np9 proteins in human biology is still little understood. As various HML-2 loci are also transcribed in normal human tissue types, it is conceivable that HML-2 proteins also exert biological functions apart from potential roles in disease development. To contribute to a better understanding of a potential biological relevance of HML-2 Rec and Np9, we investigated presence of *rec* and *np9* mRNA in a collection of normal human tissue types and identified HML-2 loci generating *rec* and *np9* mRNA. We also identified additional HML-2 loci not present in the human reference genome sequence and additional splicing variants of HML-2 transcripts potentially encoding HML-2 protein variants in the course of our studies.



**Figure 1** Schematic of HERV-K(HML-2) provirus and splicing of *rec* and *np9* mRNAs. Irrelevant proviral regions are omitted. *rec* mRNA is generated by a second splicing event of *env* mRNA removing most of the *env* gene region. Both *rec* and *np9* mRNA utilize the same splice acceptor site just upstream of the 3'-LTR. Because SD2 for *rec* mRNA is located within the 292-bp sequence missing in HML-2 type 1 proviruses, an alternative *np9*-specific SD2 is used instead also resulting in a translation frameshift when spliced to exon 3. Location of PCR primers used for amplification of *rec* and *np9* mRNA/cDNA is indicated. Note that the lower provirus map is not drawn to scale. LTR, long terminal repeat.

## Results

### Identification of HERV-K(HML-2) *rec* and *np9* mRNA in normal human tissues

Recent findings indicated transcription of HERV-K (HML-2) loci in various human cell and tissue types. Several proteins encoded by some HML-2 loci are considered to be involved in the development of some diseases, among them HML-2 Rec and Np9 proteins in the development of certain tumor types. Carrying on the identification of transcribed HERV and especially HML-2 loci in various disease and normal conditions, we were now interested in whether there are *rec* and *np9* mRNAs in normal human tissues and, if so, from which HML-2 loci those *rec* and *np9* mRNAs were generated.

To identify *rec* and *np9* mRNA, we made use of a multiple tissue cDNA panel that included cDNAs from 16 different tissue types (15 actual tissues and peripheral blood leukocytes (PBL), henceforth all designated as 'tissues' for the sake of simplicity). We amplified *rec* and *np9* mRNA-derived cDNA by using PCR primers located within exons 2 and 3 of HML-2 proviral full-length transcripts (Figure 1) and considering sequence variations between HML-2 loci within PCR primer binding regions to compensate for potentially suboptimal amplification of respective cDNAs.

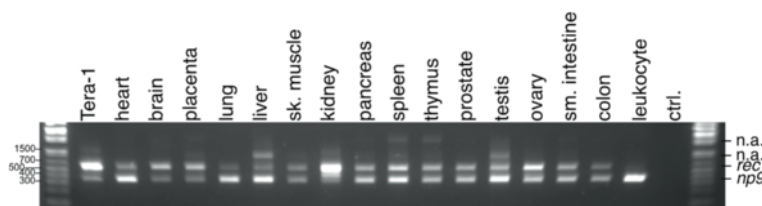
PCR products of *ca.* 580 bp indicative of *rec* mRNA could be amplified from all 16 tissue cDNAs, with amplification from cDNA from PBL resulting in only a faint band after gel electrophoresis. PCR products of *ca.* 360 bp indicative of *np9* mRNA could be amplified from all 16 tissue cDNAs as well, with amplification from cDNA from PBL producing a relatively strong PCR product (Figure 2). PCR products of *ca.* 1 kb amplified from liver and testis cDNAs and *ca.* 2.2 kb amplified from spleen and thymus cDNAs were not further regarded in this study. Taken together, *rec* and *np9* mRNA appeared to be present in all tissue types examined in this study.

### Identification of *rec* and *np9* mRNA encoding HERV-K (HML-2) loci

We then identified HML-2 loci having generated those *rec* and *np9* mRNAs. To do so, we cloned *rec* and *np9* mRNA representing PCR products and sequenced inserts from randomly selected plasmid clones. We then assigned resulting cDNA sequences, on average 41 (min. 29, max. 46) per tissue type, to specific HML-2 type 1 and type 2 loci in the human reference genome sequence by means of characteristic sequence differences between the various HML-2 loci. Despite the rather short-sized PCR products, thus short cDNA sequences (excluding primer regions), there was a sufficient number of sequence differences between relevant exon regions of HML-2 loci for unambiguously assigning generated *rec* and *np9* cDNA sequences to loci. For *rec* mRNA-derived cDNAs, only two HML-2 loci in chromosome 1 were identical in sequence for the regarded exon regions, and respective *rec* transcripts could thus, in principle, not be assigned to either one of them (Additional file 1: Figure S1).

We identified *rec* transcripts originating from, in total, nine different genomic HML-2 loci. Some tissue types (lung and colon) appeared to contain *rec* mRNA from up to five different HML-2 loci, while in kidney tissue, *rec* transcripts originated from only one HML-2 locus. Other tissue types displayed intermediate numbers of transcribed *rec* mRNA coding loci. *rec* transcripts from two HML-2 loci in chromosome 2q32.1 and 5q15 were found in 15 and 13, respectively, of the examined tissues (Table 1). As we will describe below, several of those loci are special with regard to *rec* mRNA.

We identified *np9* transcripts originating from, in total, seven different HML-2 loci present in the human reference genome sequence. As for *rec* mRNA, *np9* mRNA originated from variable numbers of HML-2 loci depending on tissue type. *np9* mRNA transcripts from HML-2 loci in chromosomes 1q22 and 3q12.3 were found in 14 and 13, respectively, of the examined tissues (Table 1). Two additional *np9* mRNA encoding HML-2



**Figure 2** Gel electrophoretic separation of *rec*- and *np9*-specific PCR products amplified from a panel of cDNAs. Generated from 15 normal human tissues and peripheral blood leukocytes. Expected product sizes were *ca.* 580 and 360 bp for *rec* and *np9*, respectively. cDNA generated from GCT-derived Tera-1 cell line, known to strongly express HML-2 both on the RNA and protein level (for instance, see ref. [28]), served as a positive control. 'ctrl.' indicates a PCR control reaction without template DNA. A faint band representing *rec*-specific PCR product from PBL is not properly reproduced. 'n.a.' indicates additional PCR products of approx. 1 kb in the liver and testis and approx. 2.2 kb in the spleen and the thymus not further investigated in this study.

**Table 1 Summary of absolute numbers of *np9* and *rec* representing cDNA sequences assignable to HERV-K(HML-2) loci<sup>a</sup>**

Type 1/2 loci	HGNC	Band	Heart	Heart tot. RNA	Brain	Brain tot. RNA	Placenta	Lung	Liver	Skeletal muscle	Kidney
chr1:75615359-75621731	<i>ERVK-1</i>	1p31.1					1				
chr1:153863081-153872260	<i>ERVK-7</i>	1q22	1	7			11	15	21	2	1
chr1:158927199-158936430	<i>ERVK-18</i>	1q23.3									
chr3:102893427-102902549	<i>ERVK-5</i>	3q12.3	1	10	6	1		7	3	1	2
chr3:114225814-114234972	<i>ERVK-3</i>	3q13.2									
chr4:166136289-166143518	-	4q32.3									
chr22:17306187-17315361	<i>ERVK-24</i>	22q11.21			1						
HERV-K111			16	6	10	5	2		2	5	6
Venter locus			1	13	1		6	2		3	4
Not assignable <i>np9</i> -like			15	25	7	3	4		2	12	4
chr2:187093879-187095344	<i>ERVK-30</i>	2q32.1	10	13	2	2	6	5	4	4	
chr3:127091992-127101129	<i>ERVK-4</i>	3q21.2				1					
chr5:30522516-30531962	-	5p13.3									
chr5:92818136-92819668	<i>ERVK-31</i>	5q15	2	8	13	89	10	11	1		
chr6:78483381-78492802	<i>ERVK-9</i>	6q14.1						1		15	
chr7:4588583-4606557	<i>ERVK-6</i>	7p22.1						1			
chr10:101570559-101577735	<i>ERVK-17</i>	10q24.2					1		8		
chr12:57007509-57016965	<i>ERVK-21</i>	12q14.1		3			2	1			20
chr19:32820338-32829201	<i>ERVK-27</i>	19q12			2						
Total # of sequences			46	85	42	101	43	43	41	42	37

Type 1/2 loci	Pancreas	Spleen	Thymus	Prostate	Testis	Ovary	Small intestine	Colon	Colon tot. RNA	Leukocyte	ORF	mRNA
chr1:75615359-75621731	1										Y	
chr1:153863081-153872260	1	22	11	4	7		10	8	6	24	Y	
chr1:158927199-158936430			3			1		1		2		
chr3:102893427-102902549	1	1	2	4	7		3	5	12		Y	
chr3:114225814-114234972				1							Y	<i>np9</i>
chr4:166136289-166143518					1							
chr22:17306187-17315361											Y	
HERV-K111	1		2	2		1	3	4	11			
Venter locus	15			5		13	5	3				
Not assignable <i>np9</i> -like	5	2	4	12	3	8	5	4	18			
chr2:187093879-187095344	13	13	12	3	1	5	4	8	28	17		
chr3:127091992-127101129							1		1		Y	
chr5:30522516-30531962				1							Y	
chr5:92818136-92819668	4	7	1	11	7	10	9	7	12			<i>rec</i>
chr6:78483381-78492802											Y	
chr7:4588583-4606557				1				2			Y	
chr10:101570559-101577735	1				3			1			Y	
chr12:57007509-57016965			1			1	1	1			Y	
chr19:32820338-32829201											Y	
Total # of sequences	42	45	36	44	29	39	41	44	88	43		

<sup>a</sup>Normal tissue types from which cDNA sequences were generated are indicated on the top. Values in italics indicate results from experiments starting from total RNA instead of pre-made cDNA specifically for heart, brain, and colon tissue (see paper text). HML-2 loci including chromosomal coordinates in the hg18 human reference genome sequence are indicated in the left-most column, followed by HUGO Gene Nomenclature Committee (HGNC)-assigned locus designations [29] and location in chromosomal bands. Total numbers of analyzed cDNA sequences per tissue are indicated at the bottom. *np9*-like cDNA sequences unambiguously assignable to recently described HERV-K111 and Venter locus sequences not present in the human reference genome sequence (see text for details) are given separately. *np9*-like cDNA sequences not assignable to any of the listed or other HML-2 type 1 locus sequences are also given separately. Presence of ORFs for Np9 or Rec protein ('Y') in predicted mRNA sequences derived from the various HML-2 loci is indicated. See Figure 5 for actual protein sequences.

loci, from which transcripts were identified in many tissues and which are not present in the human reference genome sequence, are described below.

Taken together, our results indicate that *rec* or *np9* mRNA originated from at least 18 different HML-2 loci in various normal human tissue types.

#### **An additional HML-2 locus formed by L1-mediated retrotransposition of *rec* mRNA**

We have previously reported a HML-2 locus located in chromosome 2q32.1 that was formed by L1-mediated retrotransposition of a *rec* mRNA [21]. In the present study, that locus was found to be transcribed in almost all examined normal human tissues (Table 1). We now identified an additional HML-2 locus located in chromosome 5q15 (hg18; chr5: 92818136–92819668), also transcribed in almost all of the investigated tissues (Table 1), that very likely was also formed by L1-mediated retrotransposition of a *rec* mRNA. The 1.5-kb-long locus is flanked by target site duplications (5'-TTAAAAATGT-3') typical of an L1 target site consensus sequence [30] with a poly-A tail and a poly-A signal located (more) upstream of the locus' 3' end. Apart from an approximately 250-bp 5' truncation of the retrotransposed mRNA, portions typically missing from a full-length proviral HML-2 sequence are those likewise not present in a *rec* mRNA and boundaries of missing sequence portions coincide with known splice donor and acceptor sites of *rec* mRNA. Also, those sites are basically identical with the ones of the retrotransposed locus in chromosome 2q32.1. The locus in chromosome 5q15 is evolutionarily old as it is also present in the homologous genome regions of chimpanzee, gorilla, orangutan, and gibbon, but missing in rhesus, baboon (data not shown), and the common marmoset (Figure 3). The latter, as a new world primate, is lacking HML-2 homologous sequences entirely. The locus in chromosome 5q15 does not encode a Rec(-like) protein due to a stop codon 20 triplets into the coding sequence.

#### **True transcription of two retrotransposed HML-2 loci**

We employed for amplification of *rec* and *np9* transcripts a pre-made panel of cDNAs. As opposed to true splicing events, the two retrotransposed HML-2 loci in chromosomes 2q32.1 and 5q15 would produce identically sized PCR products when amplified from mRNA/cDNA or genomic DNA. We were therefore concerned that amplified PCR products assignable to those two loci were due to traces of genomic DNA present in the pre-made cDNAs, thus conceivably a false indication of those two HML-2 loci being transcribed in the examined tissues. To investigate this further, we generated cDNA from commercially available total RNA from three normal human tissues, specifically heart, brain, and colon, following our previously established protocols for

rigorous DNA removal and including strict controls for DNA contamination [21]. We assigned, on average, 91 cDNA sequences from each of the three tissue RNAs to genomic HML-2 loci. Overall, when taking higher numbers of cDNA sequences generated per tissue into account, we obtained very similar numbers regarding transcribed HML-2 loci when compared to the pre-made cDNA panel, especially regarding cDNA sequences assignable to the loci in chromosomes 2q32.1 and 5q15 (Table 1). It thus appears that the two retrotransposed HML-2 loci are truly transcribed in quite a number of normal human tissue types.

In accord with an HUGO Gene Nomenclature Committee initiative [29], the two HML-2 loci in chromosomes 2q32.1 and 5q15 were designated *ERVK-30* and *ERVK-31*, respectively.

#### **Transcription of HML-2 loci not present in the human reference genome sequence**

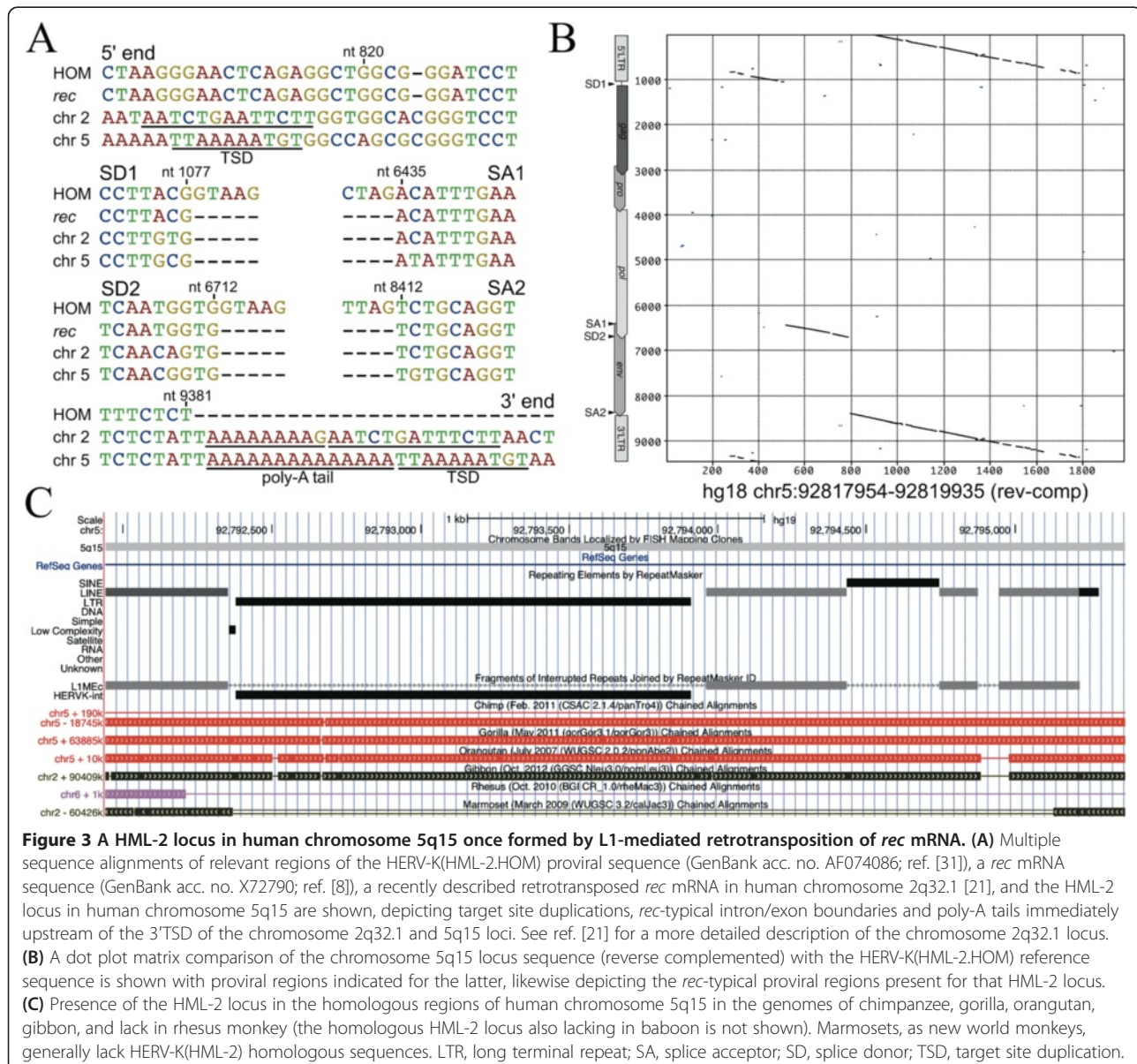
Generated cDNA sequences regularly included *np9* mRNA-like sequences that could not be assigned to HML-2 type 1 loci in the human reference genome sequence. At first glance, one population of sequences was most similar to the *ERVK-5* locus in chromosome 3q12.3, yet almost all those sequences uniformly displayed eight different nucleotide positions to that locus. Further analysis provided evidence that about half of those cDNA sequences were very likely transcribed from recently reported HERV-K(HML-2) loci not present in the human reference genome sequence. Specifically, out of 280 cDNA sequences deemed unassignable to HML-2 loci in the human reference genome, a subset of 76 cDNA sequences were identical to the recently reported HERV-K111 sequence (GenBank acc. no. GU476554; ref. [32]) along the comparable proviral regions and thus were presumably transcribed from the HERV-K111 locus. Another subset of 71 cDNA sequences were identical to a recently reported, 4214-bp-long sequence entry consisting of a partial HERV-K(HML-2) type 1 locus (GenBank acc. no. ABBA01159463; ref. [33]) (DNA donor: J. Craig Venter; henceforth named 'Venter locus') and were thus most likely transcribed from the Venter locus (Additional file 1: Figure S2). Another subset of 133 unassignable sequences was neither identical to HERV-K111 nor the Venter locus.

A fourth cDNA sequence population was most similar to the *ERVK-18* locus in human chromosome 1q23.3, yet uniformly displayed 25 differing nucleotides to that locus. That sequence population harbored an additional 245 nt compared to the amplified *np9* cDNA-derived PCR product. The difference in length was due to a SD signal located 252 nt downstream within the *env* gene region and a different SA2 located 7 nt downstream from the canonical *rec/np9* SA2 (see below). Also, those sequences

lacked the 292-bp sequence discriminating HML-2 type 1 and type 2 loci, so that they cannot be interpreted as *rec*-like mRNA transcribed from a HML-2 type 2 locus (Figure 4). The sequences displayed between 2- and 5-nt differences along the comparable 567 nt of cDNA sequence to the Venter locus and to HERV-K111. It is conceivable that those 245-bp longer cDNA sequences, compared to *np9* mRNA, represent alternatively spliced transcripts from HML-2 type 1 loci, potentially from sequence alleles of HERV-K111, several of which have been reported recently [34], or an allele of the Venter locus, or other hitherto unknown sequence or presence/absence alleles of HML-2 loci, several of which were partially described recently [35] (see also the Discussion section).

All of the abovementioned sequences appeared to have been spliced differently from *np9* mRNA. Specifically, the SA site of intron 2 (removing most of the envelope coding region) was located 7 bp downstream from the canonical *rec* and *np9* mRNA SA2 site [7,8]. This was most likely due to both HERV-K111 and the Venter locus harboring a mutated SA2 (5'-TGTTAGTCTG-3' → 5'-TGTTGGTCTG-3') and a SA signal located 7 bp downstream (5'-CTGCAGGTGT-3') being used instead (Figure 4).

Taken together, based on detected sequence similarities, our analysis provided evidence for HERV-K111 and the Venter locus being transcribed and encoding *np9*-like mRNA in various normal human tissue types. It was



unclear whether one or several alleles of those two loci or another HML-2 type 1 locus encodes an alternatively spliced, 245-bp longer mRNA.

#### Coding capacity of *rec* and *np9* mRNAs

We analyzed whether transcribed *rec* and *np9*(-like) mRNAs also have potential to encode Rec and Np9 proteins. We have previously analyzed the capacity of genomic HML-2 type 2 loci to encode Rec protein and have identified a number of HML-2 loci encoding *rec* mRNA [27]. We now analyzed in a similar fashion the capacity of HML-2 type 1 loci to potentially encode Np9 protein by (i) presence of canonical SD and SA sites, (ii) and an ORF for Np9 protein within the predicted mRNA sequence. We identified a total of 12 HML-2 type 1 loci to potentially produce a spliced mRNA and to harbor an ORF for Np9 protein as previously reported in size [7]. Several of the resulting Np9 proteins displayed amino acid differences compared among each other (Figure 5). For instance, a Np9 protein potentially encoded by a locus on chromosome 1 (hg18: 205875079–205879259) would harbor a deletion of three amino acids and additional amino acid differences overlapping with previously reported nuclear localization and LNX protein interaction domains [27]. Other HML-2 type 1 loci could potentially only encode Np9-like proteins about ten or more amino acids shorter than full-length Np9, or being similar to Np9 only within the N-terminal third of a (shorter) protein. This is also the case for proteins potentially encoded by HERV-K111 and the Venter locus that are identical with the canonical Np9 protein sequence only for the N-terminal 15 aa (Figure 5). Notably, several of the potentially protein encoding HML-2 type 1 loci were found transcribed in various normal human tissues.

Similar to our study from a decade ago [27], we re-analyzed for the present study HML-2 type 2 loci potentially encoding Rec protein by examining sequence features required for the splicing of *rec* mRNA and the translation of a Rec protein. Our re-analyses identified ORFs for Rec protein presumably in up to 19 HML-2 type 2 loci, among them the two polymorphic HERV-K113 and HERV-K115 loci [36]. Two other HML-2 loci on human chromosome 4 (hg18: 4029946–4039536 and 9268677–9278272) may potentially encode much longer, Rec-like proteins if they were transcribed. Several other loci only harbor (very) short Rec protein ORFs. The potential full-length Rec proteins display various amino acid differences when compared among each other (Figure 5). As for Np9, several of the potentially Rec protein encoding loci were found transcribed in various normal human tissues.

#### Discussion

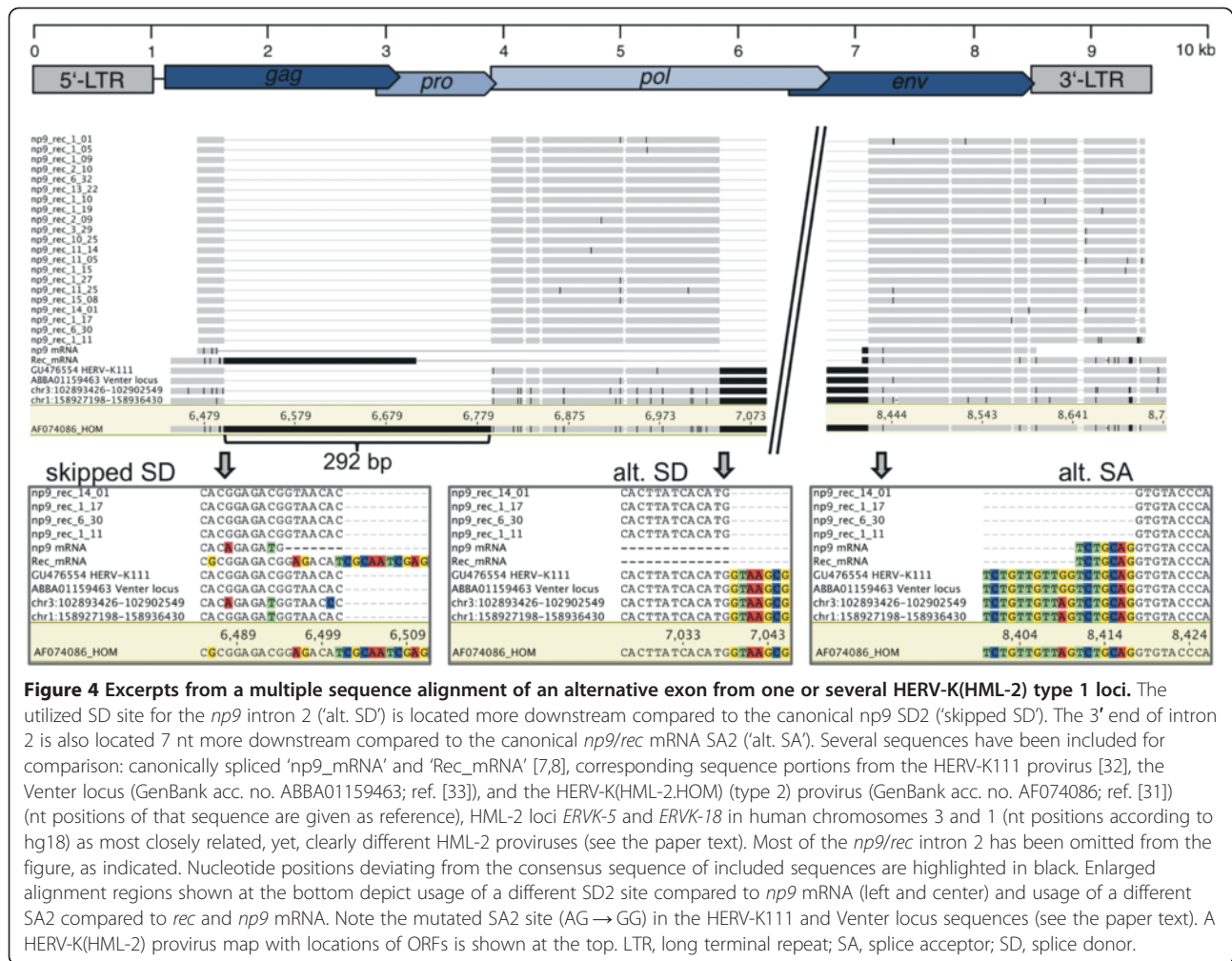
Several lines of evidence suggest biological significance of HERV-K(HML-2)-encoded proteins Rec and Np9 in

the development of tumor diseases, for instance, germ cell tumors and melanoma (for reviews, see [2,37]). However, biological roles of those proteins still need to be investigated in much more detail. Since transcripts from several HML-2 loci have been identified in non-tumor and even normal tissues, it is also conceivable that Rec and Np9 exert biological roles in normal human tissues. We therefore investigated in this study whether there are *rec* and *np9* mRNA transcripts also in normal human tissues. Our strategy for identification of such transcripts involved the amplification of a PCR product encompassing the second and third exons of *rec* and *np9* mRNA. Sequence differences within primer binding regions were also considered to potentially demonstrate transcripts from sequence-diverged HML-2 loci (see also ref. [21]). Contrary to previous studies (for instance, see [21,38]), this study does not allow for estimating relative transcript levels for different HML-2 loci as *rec* and *np9* mRNA representing PCR products were isolated and cloned in a combined fashion thus potentially falsifying relative frequencies of *rec* and *np9* mRNA-derived sequences and transcribed loci. It is also conceivable that full-length transcripts from some HML-2 loci are spliced more efficiently to *rec* or *np9* mRNA than those from other loci. Nevertheless, higher numbers of cDNA sequence derived from particular loci may hint towards higher transcript levels or more efficient splicing of full-length transcripts from those loci. Also, amplified cDNA sequences do not fully document the structure of the actual *rec* and *np9* mRNA sequences as we amplified exons 2 and 3 encompassing intron 2, disregarding exon 1 and intron 1 of *rec* and *np9* mRNAs. It is therefore, in principle, possible that for some HML-2 loci transcript, regions outside of the examined regions are spliced in some non-canonical way, though Rec and Np9 protein coding regions appear to be spliced properly (see also below).

Independent of that, *rec* and *np9* mRNA appear to be present in quite a number of human tissue types as indicated by respective PCR products amplified from 16 normal human tissues investigated in this study.

Our assignments of *rec* and *np9* cDNA sequences to HML-2 loci also indicate that at least 18 different HML-2 loci can be transcribed in normal human tissues and very likely encode *rec* or *np9* mRNAs. Additional transcribed loci encoding *rec* or *np9* may be identified, and tissue-specific patterns of such *rec* or *np9* mRNA encoding loci may be identified when much higher numbers of cDNA sequences than in this study are assigned to HML-2 loci.

Our assignment of cDNA sequences to HML-2 loci also lends support to peripheral blood leukocytes not being the sole source of *rec* and *np9* mRNAs in the various tissues as all tissues would then have displayed a PBL-typical basic pattern of transcribed HML-2 loci. However, HML-2 loci identified as transcribed and



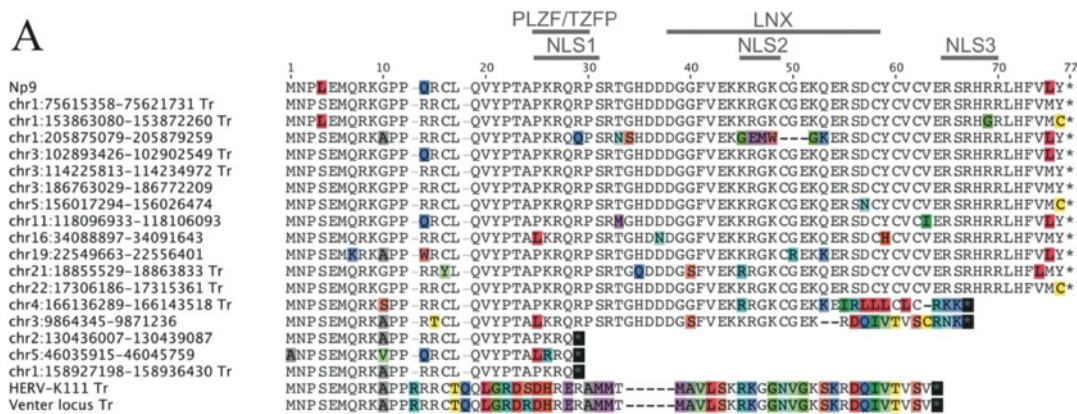
encoding *rec* and *np9* mRNAs in PBL are quite different from locus patterns observed for the various tissues. As tissues are always composed of various cell types, it remains to be investigated which cell types in the regarded tissues actually produce *rec* and *np9* mRNAs. Different scenarios are conceivable. For instance, some HML-2 loci may be transcribed and produce a subset of *rec* and/or *np9* mRNAs in some cell types within a particular tissue, while other cell types within that tissue contribute a different mRNA subset because of other HML-2 loci being transcribed in those cells. The proportions of HML-2 loci transcribed in different cell types could differ considerably between tissue types. Eventually, cell type-specific transcription patterns will have to be established.

As for *Rec* and *Np9* protein levels, it is currently not known how efficiently *rec* and *np9* mRNAs are translated into respective proteins, how stable those proteins are in normal human cells, and whether there is protein translated in one or the other tissue/cell type at all.

Cell culture-based experiments demonstrated a relatively stable HML-2 *Rec* protein with a half-life >8 h [39]. *Np9* protein seems much shorter-lived in cell culture experiments [18,19]. Nevertheless, little appears to be known about the expression of *Rec* and *Np9* proteins in normal and diseased conditions. *Rec* protein expression was reported in some melanoma tissue samples, but not in melanocytes or normal lymph nodes [40,41] and in normal synovial, rheumatoid arthritis, and osteoarthritis specimens [42]. *Np9* protein was identified in EBV-positive Raji cells that are derived from a Burkitt's lymphoma, and in an EBV-transformed human lymphoblastoid cell line, IB4 [43]. The detection of *Rec* and *Np9* proteins was accompanied in those studies by detection of *rec* and *np9* mRNAs. Therefore, *rec* and *np9* mRNAs present in normal human tissues imply presence of *Rec* and *Np9* protein in normal human tissues. However, specially designed studies on *Rec* and *Np9* protein levels, half-life, cellular distribution, and so on, in normal human cells will be required. Well-suited *Rec*- and *Np9*-specific antibodies



A



B

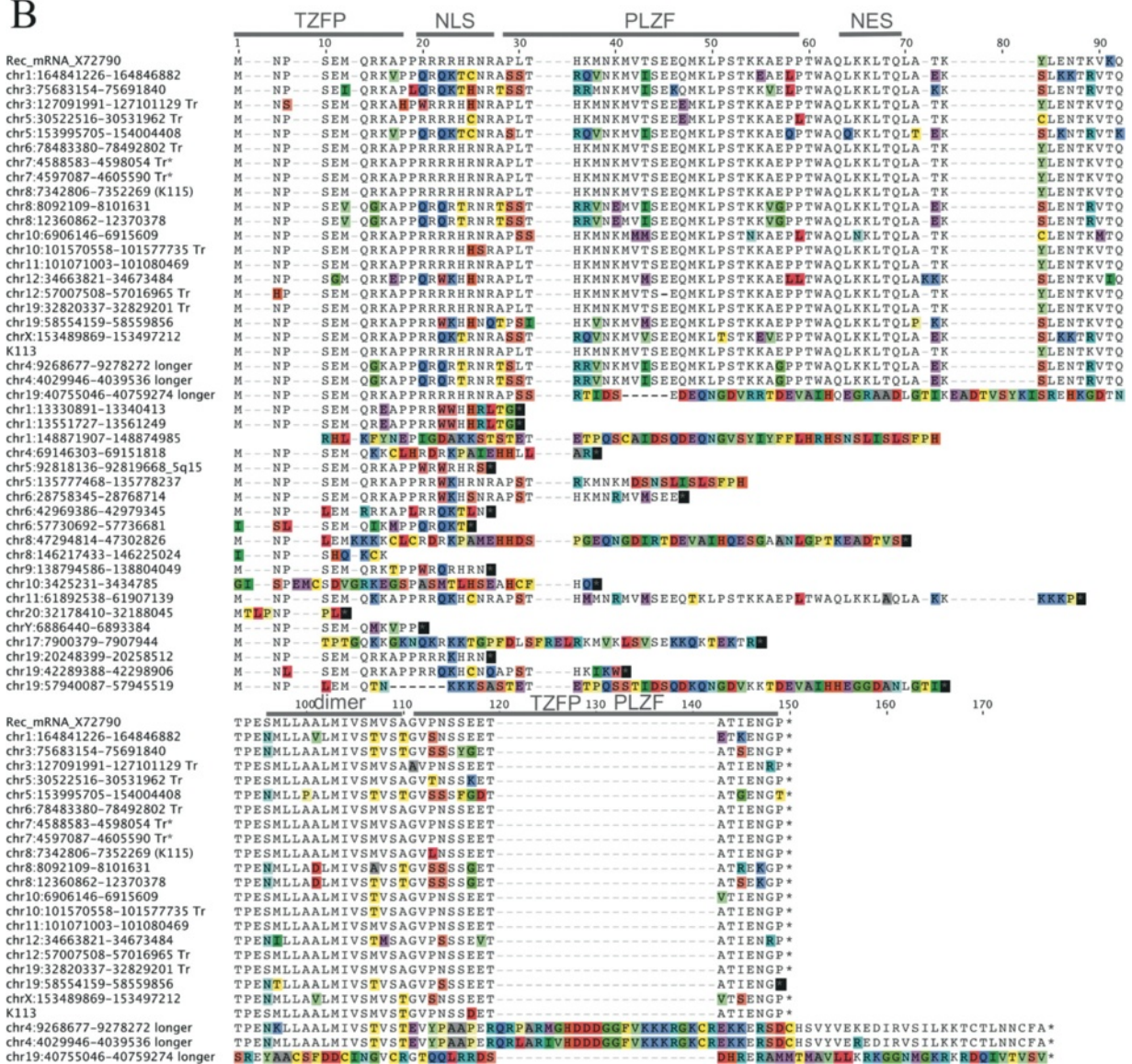


Figure 5 (See legend on next page.)

(See figure on previous page.)

**Figure 5 Multiple sequence alignments of conceptual translations of HML-2 type 1 and type 2 Np9 and Rec ORFs.** HML-2 loci potentially encoding Np9 (A) or Rec (B) protein and loci encoding only shorter proteins are shown. Chromosomal positions are given for each locus with respect to the human hg18 reference genome sequence. 'Tr' indicates that the locus was identified as transcribed in the present study. Protein sequences of a previously identified Np9-encoding mRNA [7] and a Rec-encoding mRNA (GenBank acc. no. X72790; ref. [8]) are given on the top of each alignment. Previously reported Np9 and Rec functional domains are indicated in each alignment. Full-length protein-encoding HML-2 loci are given in chromosomal order followed by shorter Np9-like and Rec-like ORFs. Proteins encoded by type 1 loci HERV-K111 and the Venter locus (see text) are included in the Np9 alignment. Amino acid positions different among each other are highlighted. NES, nuclear export signal; NLS, nuclear localization sequence; PLZF, promyelocytic zinc finger protein; TZFP, testicular zinc finger protein.

appear crucial for such protein studies including Western blot and immunohistochemistry and immunocytochemistry for examination of tissue and cellular distributions, respectively. It seems unclear whether current Rec and Np9 antibodies will be fully suited for such studies especially when considering that several HML-2 protein variants with amino acid sequences and protein sizes very similar to Rec and Np9 proteins have been described recently (for instance, see [21,39]) and in this study.

Our analysis of transcribed HML-2 loci furthermore identified two loci (designated *ERVK-30* and *ERVK-31*), both once formed by retrotransposition of *rec* mRNA by L1 machinery, as transcribed in normal human tissues. Locus *ERVK-30*, located in chromosome 2q32.1, has already been described before [21]. The present study identified another such locus, *ERVK-31*, located in chromosome 5q15, that is due to L1-mediated retrotransposition as it displays typical hallmarks of that process and is identical regarding exon-intron junctions compared to the *ERVK-30* locus in 2q32.1 and *rec* mRNA. The *ERVK-31* locus in chromosome 5q15, located central within a ~57-kb intron of the *NR2F1 antisense RNA 1 (NR2F1-AS1)* gene producing a non-coding RNA, is about as evolutionarily old as the *ERVK-30* locus in 2q32.1; none of the two loci is present in the homologous regions of the rhesus monkey and baboon genomes but both homologous loci are present in the genomes of subsequent primate species.

Evidence for both loci being transcribed was not due to artifactual amplification from contaminating DNA potentially still present in employed cDNA tissue panels, as demonstrated by our control experiments from total RNA from three selected normal tissues. Sequence data from, for instance, the ENCODE project provide additional support for the retrotransposed *rec* mRNA loci *ERVK-30/2q32.1* and *ERVK-31/5q15* being transcribed. Numerous single-pass and paired-read RNA-seq reads generated from various cell lines and normal cell types were mapped to the two loci's sequence portions (data not shown; ref. [44]). We thus describe here the second instance of a HML-2 *rec* mRNA that was retrotransposed by L1 machinery and is now transcribed by a hitherto unknown promoter active in many human tissues. Contrary to the retrotransposed *rec* mRNA

locus in chromosome 2q32.1, the locus in chromosome 5q15 does not appear to encode a Rec-like protein as it harbors a stop mutation about 20 triplets and a frameshift about 63 triplets into the Rec coding sequence.

Our study presumably identified in various human tissues transcripts from two recently described HML-2 loci that are not present in the human reference genome sequence. One population of *np9* cDNA sequences was identical to the recently described HERV-K111 provirus [32]; another population of *np9* cDNA sequences was identical to sequence portions in GenBank acc. no. ABBA01159463, a sequence identified in the genome of J. Craig Venter, that consists of HML-2 *pol*, *env*, and 3' long terminal repeat (LTR) sequence portions, starting at *ca.* nt 5000 relative to the HERV-K(HML-2.HOM) proviral sequence (GenBank acc. no. AF074086.2; ref. [31]). Notably, transcripts from both loci employed a SA2 site located 7 nt downstream from the canonical SA2 site due to a mutation within that site. Transcripts assignable to the HERV-K111 provirus were identified in our study in all but four of the investigated normal tissue types including lack of detection of HERV-K111 transcript in PBL. The HERV-K111 provirus was previously reported to be specifically active during HIV infection [32].

Transcripts from the 'Venter locus' were identified in 11 of the investigated tissues. Since several of the employed tissue cDNA panels were pools from higher numbers of donors, it seems less likely that lack of transcripts in several tissues is due to a polymorphic presence/absence status of the HERV-K111 and the Venter locus in respective tissues. In any case, more detailed analyses will be required to characterize especially the status, genome location, and exact sequence of the Venter locus. We note in this context that numerous sequence variants of HERV-K111 were recently reported [34]. The Venter locus sequence displays 35-nt differences to HERV-K111 along 4.178 kb of comparable sequence and only 1-nt difference along the amplified cDNA sequence portion. While the sequence of the Venter locus is not identical to any of the reported HERV-K111 variants (data not shown), it is nevertheless

conceivable that the Venter locus is, in fact, an undescribed variant of HERV-K111.

Additional analyses will also be required to identify the HML-2 source locus (or loci) producing an alternatively spliced, HML-2 type 1 locus-derived mRNA that utilizes a splice donor site located 252 nt downstream from the canonical *np9* mRNA's SD2 and thus resulting in a longer mRNA. The isolated cDNA sequences display an open reading frame of 408 nt, starting at nt 41 of the cDNA sequences (nt 6805 relative to the HERV-K (HML-2.HOM) sequence, GenBank acc. no. AF074086), encoding a 135 aa long protein identical to HML-2 Env along the N-terminal 80 aa. Since our employed forward PCR primer overlapped with the start codon of the Env, Rec, and Np9 ORFs, we currently do not have information as to whether the source locus of the alternative splice variant has a start codon identical to the start codon of Env/Rec/Np9 proteins. That is, the ORF may extend further upstream from the start codon at nt 41 of the available cDNA sequence up to the Env/Rec/Np9 canonical start codon.

In this context, we also obtained cDNA sequence evidence for splicing of transcripts from a previously characterized HML-2 locus in human chromosome 10 very likely formed by reverse transcription and integration of a HML-2 transcript lacking *env* gene portions [21]. Testis- and colon-derived cDNA sequences included one sequence each lacking a 334-nt sequence compared to full-length cDNA sequences from that chromosome 10 locus. The missing sequence portion was compatible with that sequence being a spliced out intron. The resulting protein from that splice variant would encode a chimeric protein consisting of Rec and Np9 portions (Additional file 1: Figure S3).

## Conclusions

Our study demonstrates HERV-K(HML-2) *rec* and *np9* transcripts from various HML-2 loci in various normal human tissues. Among them are Rec or Np9 protein coding loci and it is thus conceivable that Rec and Np9 proteins might be present in normal human tissues. Rec/Np9/Env-like proteins potentially encoded by retrotransposed HERV-K(HML-2) loci identified recently and in this study may also be present in various normal tissues. Besides potential roles in disease development, it seems worthwhile hypothesizing that various HERV-K(HML-2)-encoded proteins exert biological functions also in normal human tissues. Better knowledge of specific HERV-K(HML-2) loci encoding *rec* and *np9* mRNAs in the various human tissue-composing cell types and knowledge of amounts of Rec and Np9 protein present in those cell types will likely contribute to a better understanding of those proteins' functions under normal cellular conditions.

## Methods

### Multiple tissue cDNA panel and tissue total RNAs

We utilized the Human MTC™ Panel I and II (Clontech/Takara Bio, Otsu, Japan). The two panels included normalized, first-strand cDNA preparations from RNA from, in total, 15 different normal human tissues and peripheral blood leukocytes. cDNAs for each tissue consisted of pools from 3 to 98 Caucasian individuals and 550 male/female Caucasians in the case of peripheral blood leukocytes. Lung and liver derived cDNAs were from one male Caucasian each.

We also utilized commercially available total RNA from normal human heart, brain, and colon tissues (Clontech/Takara Bio; catalogue numbers 636532; 636530; 636553).

### Amplification of *np9* and *rec* PCR products, cloning of PCR products, plasmid preparation, sequencing

We amplified *rec* and *np9* mRNA representing PCR products from tissue cDNA panels employing forward primers *rec-np9-for-1*: 5'-ATG AAC CCA TCA GAG ATG CAA-3'; *rec-np9-for-2*: 5'-ATG AAT CCA TCA GAG ATG CAA-3'; *rec-np9-for-3*: 5'-GCG AAC CCT TCA GAG ATG CAA-3'; *rec-np9-for-4*: 5'-ATG AAC CCA TCG GAG ATG AAA-3'; that were combined in a ratio of 85/5/5/5, and reverse primers *rec-np9-rev-1*: 5'-AGC ATC TGT TTA ACA AAG CA-3'; *rec-np9-rev-2*: 5'-AGC ATG TTT AAC AAA GCA-3' 5% combined in a ratio of 95/5. The various primer variants considered sequence differences of HERV-K(HML-2) loci within primer binding regions. PCR products were amplified using standard conditions with AmpliTaq Gold (Applied Biosystems/Life Technologies, Carlsbad, CA, USA) DNA polymerase and the following PCR program: 12 min 95°C; 35 cycles: 50 s 95°C, 50 s 58°C, 30 s 72°C, and final elongation 10 min 72°C. PCR products were separated by agarose gel electrophoresis; *np9* and *rec* representing PCR products were purified from gels using NucleoSpin Gel and PCR Clean-Up Kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany). Products were cloned into pCR II-TOPO (Invitrogen/Life Technologies) and transformed into *Escherichia coli* DH-5α cells. Plasmid DNA from randomly selected bacterial colonies was purified and subjected to Sanger sequencing (see below).

We amplified *rec* and *np9* mRNA representing cDNA from total RNA from heart, brain, and colon tissues by RT-PCR following a previously established procedure [21]. PCR products were cloned into pGEM T-Easy (Promega GmbH, Mannheim, Germany). Plasmid DNA from randomly selected bacterial colonies was prepared as described before [32].

cDNA inserts were sequenced using vector-specific T7 primer and an Applied Biosystems 3730 DNA-Analyzer (Seq-IT GmbH, Kaiserslautern, Germany). Sequence

qualities were verified by eye, and poor quality sequence reads were excluded from further analysis.

### Chromosomal assignment of cDNA sequences

We assigned *rec* and *np9* representing cDNA sequences to specific HML-2 loci by sequence comparisons essentially as described before [21,31]. Sequences that could be unambiguously assigned to a HML-2 locus in the human reference genome sequence with less than three mismatches were considered for analysis. Sequences not matching a locus in the human reference genome sequence are described in the main text. For assignment of *np9* cDNA sequences, we omitted the rather short (23 nt when excluding primer binding region) and thus little informative exon 2 of *np9* mRNA (SA1 - SD2) because the remaining sequence portions provided a sufficient number of sequence differences between (type 1) loci (see Additional file 1: Figure S1).

### Analysis of HML-2 loci for *np9* mRNA coding capacity

Similar to a recent analysis of HML-2 *rec* coding capacity [27], we examined in a multiple alignment of genomic HML-2 sequences, plus polymorphic HML-2 proviruses not present in the human reference genome sequence, features required for *np9* mRNA splicing and Np9 protein coding capacity, specifically presence of 5' and 3' LTRs, splice donor and acceptor sites, and open reading frames as previously described [7].

### Availability of supporting data

LN624403 and LN624404 are accession numbers of cDNA sequences assignable to two HML-2 loci in human chromosomes 2q32.1 and 5q15 reported in this study and previously [21]. Accession numbers LN680257 to LN680271 are 245 bp longer, *np9*-like cDNA sequences. Sanger sequence reads generated in the course of this study have been deposited at the European Nucleotide Archive (study accession number PRJEB8273).

### Additional file

**Additional file 1: Additional analysis of HERV-K(HML-2) sequences.**  
**Figure S1.** Sequence differences between *rec* and *np9* cDNA sequences.  
**Figure S2.** Structure of 'Venter locus'. **Figure S3.** A spliced transcript encoded by a HERV-K(HML-2) locus in human chromosome 10.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

KS, KH, and JM performed the experiments and the additional analyses. KR, EM, and JM conceived the study. JM wrote the paper. All authors read and approved the final manuscript.

### Acknowledgements

This study was supported by grant number Ma2298/8-1 provided by the Deutsche Forschungsgemeinschaft to JM.

### Author details

<sup>1</sup>Institute of Human Genetics, Center of Human and Molecular Biology, Medical Faculty, University of Saarland, 66424 Homburg/Saar, Germany. <sup>2</sup>José Carreras Research Center, Medical Faculty, University of Saarland, 66424 Homburg/Saar, Germany. <sup>3</sup>Center of Human and Molecular Biology, University of Saarland, 66424 Homburg/Saar, Germany. <sup>4</sup>Sanofi-Aventis Deutschland GmbH, Industriepark Hoechst, K703, Elisabeth Kuhn Street, Frankfurt/Main 65926, Germany.

Received: 10 December 2014 Accepted: 12 February 2015

Published online: 05 March 2015

### References

- Kurth R, Bannert N. Beneficial and detrimental effects of human endogenous retroviruses. *Int J Cancer*. 2010;126(2):306–14.
- Ruprecht K, Mayer J, Sauter M, Roemer K, Mueller-Lantzsch N. Endogenous retroviruses: endogenous retroviruses and cancer. *Cell Mol Life Sci*. 2008;65(21):3366–82.
- Stoye JP. Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nat Rev Microbiol*. 2012;10(6):395–406.
- Dupressoir A, Lavialle C, Heidmann T. From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. *Placenta*. 2012;33(9):663–71.
- Ono M. Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. *J Virol*. 1986;58(3):937–44.
- Ono M, Yasunaga T, Miyata T, Ushikubo H. Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. *J Virol*. 1986;60(2):589–98.
- Armbruster V, Sauter M, Krautkraemer E, Meese E, Kleiman A, Best B. A novel gene from the human endogenous retrovirus K expressed in transformed cells. *Clin Cancer Res*. 2002;8(6):1800–7.
- Löwer R, Tönjes RR, Korbmacher C, Kurth R, Löwer J. Identification of a Rev-related protein by analysis of spliced transcripts of the human endogenous retroviruses HTDV/HERV-K. *J Virol*. 1995;69(1):141–9.
- Magin-Lachmann C, Hahn S, Strobel H, Held U, Löwer J, Löwer R. Rec (formerly Corf) function requires interaction with a complex, folded RNA structure within its responsive element rather than binding to a discrete specific binding site. *J Virol*. 2001;75(21):10359–71.
- Yang J, Bogerd H, Le SY, Cullen BR. The human endogenous retrovirus K Rev response element coincides with a predicted RNA folding region. *RNA*. 2000;6(11):1551–64.
- Magin C, Hesse J, Löwer J, Löwer R. Corf, the Rev/Rex homologue of HTDV/HERV-K, encodes an arginine-rich nuclear localization signal that exerts a trans-dominant phenotype when mutated. *Virology*. 2000;274(1):11–6.
- Yang J, Bogerd HP, Peng S, Wiegand H, Truant R, Cullen BR. An ancient family of human endogenous retroviruses encodes a functional homologue of the HIV-1 Rev protein. *Proc Natl Acad Sci U S A*. 1999;96(23):13404–8.
- Magin C, Löwer R, Löwer J. cORF and RcRE, the Rev/Rex and RRE/RxRE homologues of the human endogenous retrovirus family HTDV/HERV-K. *J Virol*. 1999;73(11):9496–507.
- Galli UM, Sauter M, Lecher B, Maurer S, Herbst H, Roemer K, et al. Human endogenous retrovirus rec interferes with germ cell development in mice and may cause carcinoma *in situ*, the predecessor lesion of germ cell tumors. *Oncogene*. 2005;24(19):3223–8.
- Hanke K, Chudak C, Kurth R, Bannert N. The Rec protein of HERV-K(HML-2) upregulates androgen receptor activity by binding to the human small glutamine-rich tetratricopeptide repeat protein (hSGT). *Int J Cancer*. 2013;132(3):556–67.
- Hanke K, Hohn O, Liedgens L, Fiedke K, Wamara J, Kurth R, et al. Staufen-1 interacts with the human endogenous retrovirus family HERV-K(HML-2) rec and gag proteins and increases virion production. *J Virol*. 2013;87(20):11019–30.
- Kaufmann S, Sauter M, Schmitt M, Baumert B, Best B, Boese A, et al. Human endogenous retrovirus protein Rec interacts with the testicular zinc-finger protein and androgen receptor. *J Gen Virol*. 2010;91(Pt 6):1494–502.
- Denne M, Sauter M, Armbruster V, Licht JD, Roemer K, Mueller-Lantzsch N. Physical and functional interactions of human endogenous retrovirus

- proteins Np9 and rec with the promyelocytic leukemia zinc finger protein. *J Virol.* 2007;81(11):5607–16.
19. Armbruster V, Sauter M, Roemer K, Best B, Hahn S, Nty A, et al. Np9 protein of human endogenous retrovirus K interacts with ligand of numb protein X. *J Virol.* 2004;78(19):10310–9.
  20. Boese A, Sauter M, Galli U, Best B, Herbst H, Mayer J, et al. Human endogenous retrovirus protein cORF supports cell transformation and associates with the promyelocytic leukemia zinc finger protein. *Oncogene.* 2000;19(38):4328–36.
  21. Schmitt K, Reichrath J, Roesch A, Meese E, Mayer J. Transcriptional profiling of human endogenous retrovirus group HERV-K(HML-2) loci in melanoma. *Genome Biol Evol.* 2013;5(2):307–28.
  22. Ruprecht K, Ferreira H, Flockerzi A, Wahl S, Sauter M, Mayer J, et al. Human endogenous retrovirus family HERV-K(HML-2) RNA transcripts are selectively packaged into retroviral particles produced by the human germ cell tumor line Tera-1 and originate mainly from a provirus on chromosome 22q11.21. *J Virol.* 2008;82(20):10008–16.
  23. Flockerzi A, Ruggieri A, Frank O, Sauter M, Maldener E, Kopper B, et al. Expression patterns of transcribed human endogenous retrovirus HERV-K(HML-2) loci in human tissues and the need for a HERV Transcriptome Project. *BMC Genomics.* 2008;9(1):354.
  24. Agoni L, Guha C, Lenz J. Detection of human endogenous retrovirus K (HERV-K) transcripts in human prostate cancer cell lines. *Front Oncol.* 2013;3:180.
  25. Fuchs NV, Loewer S, Daley GQ, Izsvak Z, Lower J, Lower R. Human endogenous retrovirus K (HML-2) RNA and protein expression is a marker for human embryonic and induced pluripotent stem cells. *Retrovirology.* 2013;10:115.
  26. Sugimoto J, Matsuura N, Kinjo Y, Takasu N, Oda T, Jinno Y. Transcriptionally active HERV-K genes: identification, isolation, and chromosomal mapping. *Genomics.* 2001;72(2):137–44.
  27. Mayer J, Ehlhardt S, Seifert M, Sauter M, Muller-Lantzsch N, Mehraein Y, et al. Human endogenous retrovirus HERV-K(HML-2) proviruses with Rec protein coding capacity and transcriptional activity. *Virology.* 2004;322(1):190–8.
  28. Sauter M, Schommer S, Kremmer E, Remberger K, Dölken G, Lemm I, et al. Human endogenous retrovirus K10: expression of Gag protein and detection of antibodies in patients with seminomas. *J Virol.* 1995;69(1):414–21.
  29. Mayer J, Blomberg J, Seal RL. A revised nomenclature for transcribed human endogenous retroviral loci. *Mob DNA.* 2011;2(1):7.
  30. Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc Natl Acad Sci U S A.* 1997;94(5):1872–7.
  31. Mayer J, Sauter M, Racz A, Scherer D, Mueller-Lantzsch N, Meese E. An almost-intact human endogenous retrovirus K on human chromosome 7. *Nat Genet.* 1999;21(3):257–8.
  32. Contreras-Galindo R, Kaplan MH, Contreras-Galindo AC, Gonzalez-Hernandez MJ, Ferlenghi I, Giusti F, et al. Characterization of human endogenous retroviral elements in the blood of HIV-1-infected individuals. *J Virol.* 2012;86(1):262–76.
  33. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The diploid genome sequence of an individual human. *PLoS Biol.* 2007;5(10):e254.
  34. Contreras-Galindo R, Kaplan MH, He S, Contreras-Galindo AC, Gonzalez-Hernandez MJ, Kappes F, et al. HIV infection reveals widespread expansion of novel centromeric human endogenous retroviruses. *Genome Res.* 2013;23(9):1505–13.
  35. Marchi E, Kanapin A, Magiorkinis G, Belshaw R. Unfixed endogenous retroviral insertions in the human population. *J Virol.* 2014;88(17):9529–37.
  36. Turner G, Barbulescu M, Su M, Jensen-Seaman MI, Kidd KK, Lenz J. Insertional polymorphisms of full-length endogenous retroviruses in humans. *Curr Biol.* 2001;11(19):1531–5.
  37. Hohn O, Hanke K, Bannert N. HERV-K(HML-2), the best preserved family of HERVs: endogenization, expression, and implications in health and disease. *Front Oncol.* 2013;3:246.
  38. Schmitt K, Richter C, Backes C, Meese E, Ruprecht K, Mayer J. Comprehensive analysis of human endogenous retrovirus group HERV-W locus transcription in multiple sclerosis brain lesions by high-throughput amplicon sequencing. *J Virol.* 2013;87(24):13837–52.
  39. Ruggieri A, Maldener E, Sauter M, Mueller-Lantzsch N, Meese E, Fackler OT, et al. Human endogenous retrovirus HERV-K(HML-2) encodes a stable signal peptide with biological properties distinct from Rec. *Retrovirology.* 2009;6:17.
  40. Büscher K, Hahn S, Hofmann M, Trefzer U, Ozel M, Sterry W, et al. Expression of the human endogenous retrovirus-K transmembrane envelope, Rec and Np9 proteins in melanomas and melanoma cell lines. *Melanoma Res.* 2006;16(3):223–34.
  41. Muster T, Waltenberger A, Grassauer A, Hirschl S, Caucig P, Romirer I, et al. An endogenous retrovirus derived from human melanoma cells. *Cancer Res.* 2003;63(24):8735–41.
  42. Ehlhardt S, Seifert M, Schneider J, Ojak A, Zang KD, Mehraein Y. Human endogenous retrovirus HERV-K(HML-2) Rec expression and transcriptional activities in normal and rheumatoid arthritis synovia. *J Rheumatol.* 2006;33(1):16–23.
  43. Gross H, Barth S, Pfuhl T, Willnecker V, Spurk A, Gurtsevitch V, et al. The NP9 protein encoded by the human endogenous retrovirus HERV-K(HML-2) negatively regulates gene activation of the Epstein-Barr virus nuclear antigen 2 (EBNA2). *Int J Cancer.* 2011;129(5):1105–15.
  44. Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, et al. ENCODE data in the UCSC genome browser: year 5 update. *Nucleic Acids Res.* 2013;41(Database issue):D56–63.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

