



Minireview

Z-DNA–Containing Long Terminal Repeats of Human Endogenous Retrovirus Families Provide Alternative Promoters for Human Functional Genes

Du Hyeong Lee^{1,4,5}, Woo Hyeon Bae^{1,4,5}, Hongseok Ha^{2,5}, Eun Gyung Park^{1,4}, Yun Ju Lee^{1,4}, Woo Ryung Kim^{1,4}, and Heui-Soo Kim^{3,4,*}

¹Department of Integrated Biological Sciences, Pusan National University, Busan 46241, Korea, ²Division of Life Sciences, Korea University, Seoul 02841, Korea, ³Department of Biological Sciences, College of Natural Sciences, Pusan National University, Busan 46231, Korea, ⁴Institute of Systems Biology, Pusan National University, Busan 46241, Korea, ⁵These authors contributed equally to this work.

*Correspondence: khs307@pusan.ac.kr

<https://doi.org/10.14348/molcells.2022.0060>

www.molcells.org

Transposable elements (TEs) account for approximately 45% of the human genome. TEs have proliferated randomly and integrated into functional genes during hominoid radiation. They appear as right-handed B-DNA double helices and slightly elongated left-handed Z-DNAs. Human endogenous retrovirus (HERV) families are widely distributed in human chromosomes at a ratio of 8%. They contain a 5′-long terminal repeat (LTR)-gag-pol-env-3′-LTR structure. LTRs contain the U3 enhancer and promoter region, transcribed R region, and U5 region. LTRs can influence host gene expression by acting as regulatory elements. In this review, we describe the alternative promoters derived from LTR elements that overlap Z-DNA by comparing Z-hunt and DeepZ data for human functional genes. We also present evidence showing the regulatory activity of LTR elements containing Z-DNA in *GSDML*. Taken together, the regulatory activity of LTR elements with Z-DNA allows us to understand gene function in relation to various human diseases.

Keywords: gene function, human diseases, human endogenous retrovirus, long terminal repeat elements, Z-DNA

INTRODUCTION

The human genome contains several transposable elements (TEs) introduced by exogenous retroviral infection in the germline cells of our ancestors (Lower et al., 1996). Human endogenous retroviruses (HERVs) with autonomous retroelements are the most well-known retrotransposons (Havecker et al., 2004). HERV insertions comprise two long terminal repeats (LTRs) flanking an internal region that encodes protein-coding genes (gag, pol, env) necessary for retroviral replication and propagation (Jern and Coffin, 2008). Using reverse transcriptase (RTase) encoded by the pol gene of HERVs, the genes randomly integrated into the human genome (Supplementary Fig. S1A) (Kim et al., 2004; Yi et al., 2004). They then processed multiple duplication events during hominoid radiation. HERVs comprise up to 8% of the human genome and are dispersed throughout the genome (Medstrand and Mager, 1998). They cause many mutation events, including deletion of subgenomes, insertions of other transposons (Alu, LINE, ERV, and DNA transposons), and homologous recombination between the 5′-LTR and 3′-LTR of HERV sequences. Both LTRs of one HERV show more similarity than with the LTR sequence of any other HERV; therefore,

Received 11 April, 2022; revised 16 May, 2022; accepted 31 May, 2022; published online 5 August, 2022

eISSN: 0219-1032

©The Korean Society for Molecular and Cellular Biology.

©This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>.

they allow the production of a solitary LTR element (Huh et al., 2006; Medstrand and Mager, 1998; Thomas et al., 2018). The structure of LTR elements includes a hormone responsive element, enhancer, promoter TATA box (located within the U3 region), polyadenylation signal AATAAA (located within the R region), and the U5 region (Supplementary Fig. S1B) (Sverdlov, 2000). LTRs can influence host gene expression by acting as regulatory elements (promoters or enhancers) (Durnaoglu et al., 2021; Montension et al., 2018; Ruda et al., 2004).

A left-handed double-helical Z-DNA fragment was identified using X-ray diffraction analysis (Dickerson et al., 1982; Drew et al., 1980; Wang et al., 1981). Purine-pyrimidine alternating sequences, such as poly(dT-dG)-poly(dC-dA), have been shown to adopt the Z-DNA conformation in the presence of high CsCl concentrations and in ethanolic solutions (Zimmer et al., 1982). Stretches of the dC-dG alternating sequence [Z(C-G) element] were found to be moderately repetitive in human, mouse, and salmon genomes (Hamada et al., 1982). The abundant occurrence and evolutionary conservation of the Z(T-G) and Z(C-G) elements could have important biological implications as they could be involved in regulating gene expression and act as hotspots for gene recombination or rearrangement (Hamada et al., 1982). Computer programs (Z-hunt and Z-hunt-II) have been developed to search for genomic sequences in regions most likely to adopt the Z-conformation (Ho et al., 1986; Schroth et al., 1992). The recently developed deep-learning approach, DeepZ, aggregates information from genome-wide maps of epigenetic markers, transcription factors, RNA polymerase-binding sites, and chromosome accessibility maps (Beknazarov et al., 2020).

Z-DNA has been found to form in actively transcribed regions of the genome and has been confirmed using ChIP-Seq, indicating that Z-DNA formation depends on chromatin structure as well as sequence composition and is associated with active transcription in human cells (Shin et al., 2016). Potential Z-DNA-forming sequences (ZFS) are abundant near the transcriptional start sites of genes (Li et al., 2009; Schroth et al., 1992). This suggests that Z-DNA plays a biological role in transcriptional regulation and that RNA polymerase II accumulates local negative supercoiling, creating a suitable environment for Z-DNA formation (Herbert and Rich, 2001; Liu and Wang, 1987). For instance, the highly conserved negative regulatory element (NRE) at the 5'-UTR of the human *ADAM12* gene acts as a transcriptional repressor. The NRE contains a stretch of a dinucleotide-repeat sequence that can adopt a Z-DNA conformation. ZFS negatively regulate *ADAM-12* expression in normal cells (Ray et al., 2011). Further, hypoxia-inducible factor 1 (*HIF-1*) regulates allelic variation in *SLC11A1* expression by binding directly to Z-DNA-forming microsatellites during macrophage activation due to infection or inflammation (Bayele et al., 2007). In this review, we summarize the functional genes containing alternative promoters derived from LTR elements that overlap Z-DNA prediction sites. We also indicate the location of alternative promoters, LTR elements, and Z-DNA prediction sites analyzed using the Z-hunt and DeepZ programs. Finally, we discuss variant isoforms introduced by alternative splicing as biomarkers for the detection of human

diseases associated with LTR elements and Z-DNA.

IDENTIFICATION AND CHROMOSOMAL LOCATION OF HERV LTR ELEMENTS AND Z-DNA

Retrotransposon activity is linked with Z-DNA-forming sites that overlap with recombination hotspots (Blaho and Wells, 1989; Wahls et al., 1990). A large portion of ZFS are enriched in promoter regions and contain sequences with high potential to form Z-DNA. The Z-DNA-forming sites identified using ChIP-Seq are associated with actively transcribed regions (Shin et al., 2016). ZFSs are also abundant in transposable elements (Alu) (Herbert, 2019). Alternative splicing and Z-formation appear in genes with Alu repeats and dsRNA editing of transcripts. Homologous recombination between the 5'-LTR and 3'-LTR of HERVs results in excision of structural genes (*gag-pol-env*), leaving a solitary LTR element (Kim, 2012; Ruda et al., 2004; Thomas et al., 2018). LTRs have regulatory potential to host protein-coding genes because of their highly enriched transcription factor-binding sites (Ito et al., 2017; Yu et al., 2013). Global information about ZFS positioning could provide useful information for further understanding Z-DNA structure-dependent transcriptional regulation. Elucidation of ZFS in HERV LTR elements has revealed variant isoforms of functional genes in relation to alternative promoters. Analysis of the genomic position between LTRs and ZFS is needed to determine whether LTR elements contain ZFSs. The positions of TEs including LTR class were obtained from the group "Repeats" (RepeatMasker) on the table browser of UCSC genome browser from the human genome (hg19) (Kent et al., 2002). A dataset (<https://github.com/Nazar1997/DeepZ/tree/master/annotation>) annotated in a previous study was adopted to obtain the genomic positions of ZFS (Beknazarov et al., 2020). This dataset included experimental Z-DNA regions and putative regions generated using Z-hunt (Schroth et al., 1992) or DeepZ (Beknazarov et al., 2020). The IntersectBed module from Bedtools was used to identify TE coordinates that overlapped with the ZFS position. The output files were modified using an in-house Python code. In this processing step, only the LTR elements belonging to the LTR repeat class were considered for downstream analyses. The modified file was used as an input for the web-based PhenoGram (<http://visualization.ritchieab.org/pheno-grams/plot>) to visualize the coordinates of LTR elements with ZFS on human chromosomes. Regarding the chromosomal region of the overlaps, an average of 1.11% of the LTR elements in each chromosome (7,823 out of 708,332 in entire chromosomes) had ZFSs (Supplementary Table S1). Among these, LTR/ERV1 accounted for the largest proportion, with 34.33% (2,686 of 7,823 in LTR containing ZFS), followed by LTR/ERVL-MaLR with 30.68% (2,400 out of 7,823). However, LTR/ERVK had the highest density with 5.89% (618 out of 10,490). The density indicates the ratio of each LTR class/family to ZFS (the number of LTR fragments containing ZFS/the number of LTR fragments in the human genome). In our previous study, we reported the chromosomal distribution and copy numbers of the HERV family in humans and great apes, indicating that HERV-K/solitary LTR elements were the most abundant (Kim, 2012). HERV-K family members also

proliferated continuously during hominid evolution (Supplementary Fig. S2) (Anderssen et al., 1997; Di Cristofano et al., 1995; Hervé et al., 2004; Kjellman et al., 1999; Lavie et al., 2004; Lee and Kim, 2006; Yi et al., 2007a; 2007b; 2007c). Human-specific HERV-K activity has contributed to genomic divergence between humans and chimpanzees, as well as within the human population (Shin et al., 2016). Multiple copy numbers of solitary LTR elements belonging to the HERV-K family (GenBank accession No. AC002350, AC002400, AC002508, L47334, U47924, Z80898, and AL034407) have been identified as being unique to humans (Akopov et al., 1998; Buzdin et al., 2002; Medstrand and Mager, 1998). Solitary LTR elements were formed because of an equal homologous recombination excision event. Several evolutionary processes have occurred throughout the chromosomes during primate evolution. HERV-K LTR elements are the youngest retrovirus family in the human genome and

are the only group of endogenous retroviruses that have polymorphic members in human populations (Macfarlane and Simmonds, 2004). As shown in Fig. 1, HERV-K/solitary LTR elements containing Z-DNA are present in all chromosomes except chromosomes 15, 20, 21, and 22, suggesting that they are still active in the human genome as regulatory members. Therefore, we investigated functional genes with alternative promoters derived from LTR elements containing Z-DNA.

ALTERNATIVE PROMOTERS DERIVED FROM LTR ELEMENTS OVERLAPPING Z-DNA

Promoters regulate the transcription of exons located in downstream positions. Over half of human genes contain more than one promoter, which are collectively described as alternative promoters. Alternative promoters provide tran-

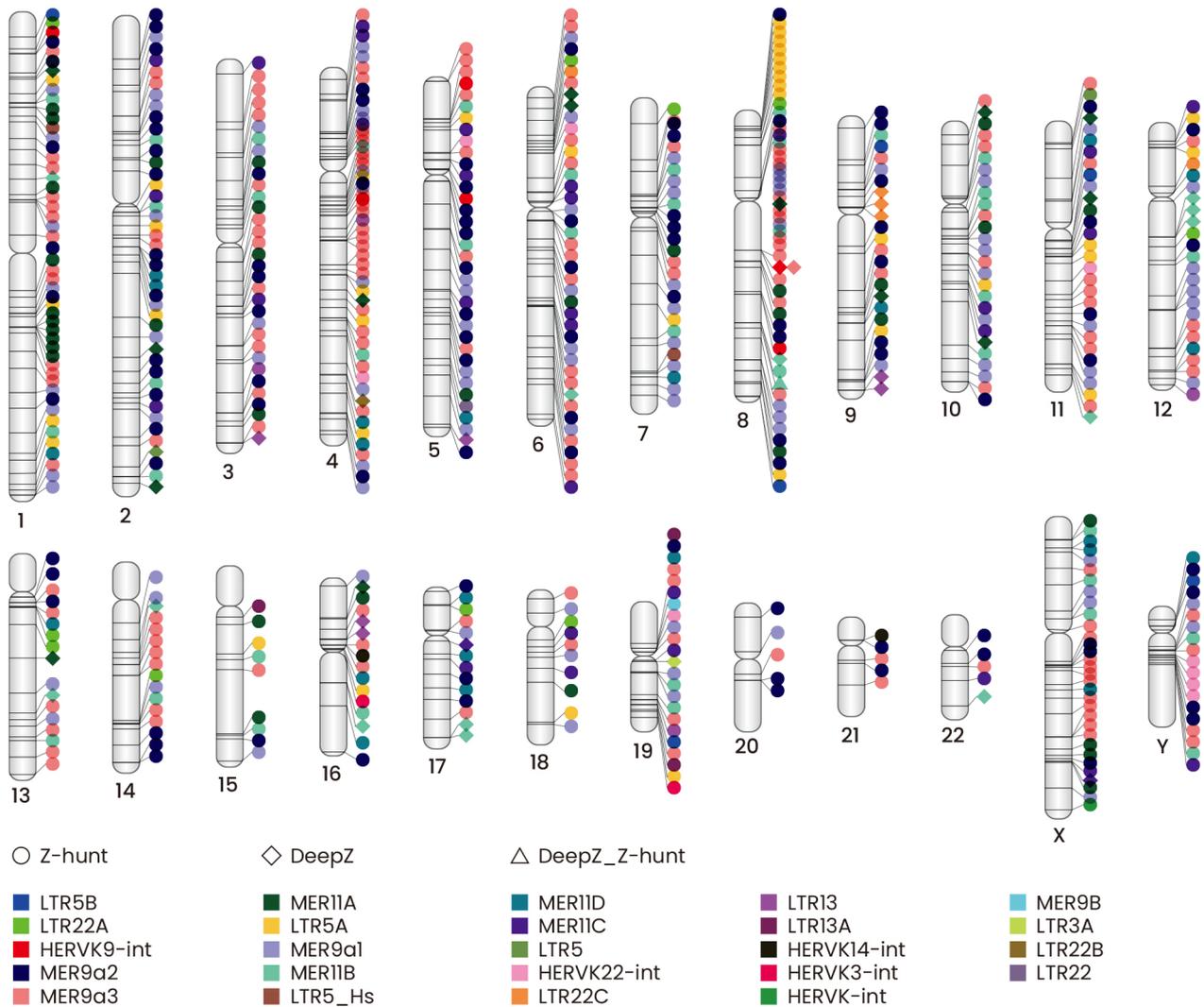


Fig. 1. The chromosome ideogram showing the LTR classes of HERV-K families. Each diagram indicates the putative Z-DNA location detected using Z-hunt and DeepZ programs. Different LTR classes are distinguished by different colors.

Table 1. Functional genes containing alternative promoter derived from LTR elements overlapping Z-DNA prediction site

Genes	NCBI Gene ID	Loci	TE types
<i>MAN1C1</i>	57134	1p36.11	MER52A-ERV1
<i>ZNF80</i>	7634	3q13.31	LTR12C-ERV1
<i>FIS(C5orf27)</i>	202299	5q15	LTR12C-ERV1
<i>ZNF323(ZSCAN31)</i>	64288	6p22.1	HERV18 int-ERV1
<i>LOC223075(CCDC129)</i>	223075	7p14.3	LTR12D-ERV1
<i>FLJ45974</i>	401337	7p12.1	MLT2B3-ERV1
<i>HHLA1</i>	10086	8q24.22	HERV-H int-ERV1
<i>LY6K</i>	54742	8q24.3	LTR43-ERV1
<i>INSL4</i>	3641	9p24.1	LTR22B/HML-5
<i>NOV1(C11orf40)</i>	143501	11p15.4	LTR18B-ERV1
<i>PRDM10</i>	56980	11q24.3	LTR52-ERV1
<i>LINC00615</i>	439916	12q21.33	LTR60-ERV1
<i>MSLN</i>	10232	16p13.3	MER54B-ERV1
<i>OTOA</i>	146183	16p12.2	LTR45B-ERV1
<i>CCL15</i>	6359	17q12	MER50-ERV1
<i>FLJ10260(SLFN12)</i>	55106	17q12	MER51A-ERV1
<i>GSDML(GSDMB)</i>	55876	17q21.1	LTR7B-ERV1
<i>APOC1</i>	341	19q13.32	LTR2/HERV-E
<i>HSPC072(LINC00652)</i>	29075	20p11.23	MLTF-ERV1

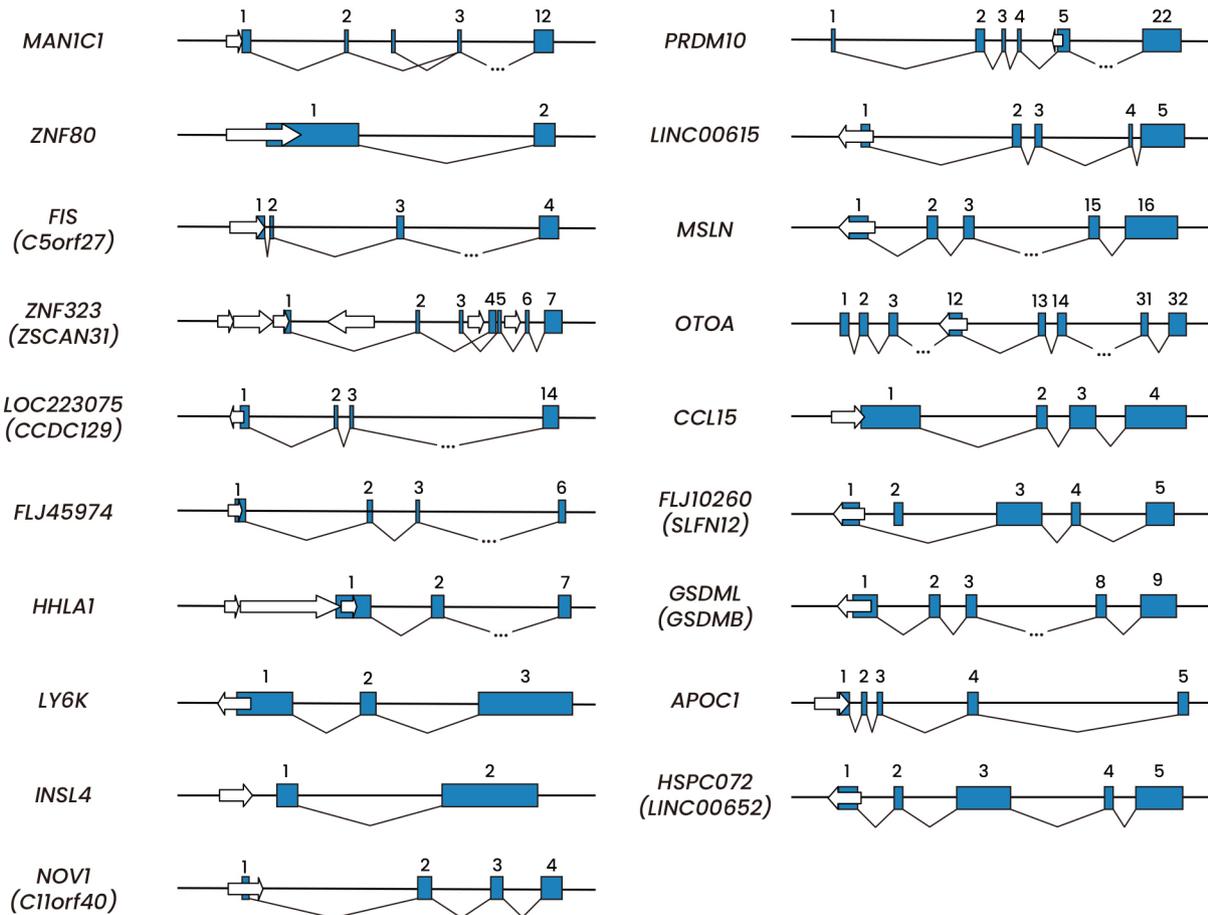


Fig. 2. The genomic structure of functional genes containing alternative promoters derived from LTR elements overlapping with Z-DNA.

script diversity and confer dimensional complexity to cells (Landry et al., 2003). They also have different tissue specificities, developmental activities, and expression levels (Medstrand et al., 2001; Schon et al., 2009). Alternative promoters contribute to expression diversity as they create mRNA isoforms by expanding the choice of transcription initiation sites in a gene (Jacox et al., 2010).

HERV LTR elements have a potential evolutionary role in enhancing the coding capacity and regulatory versatility of the genome without compromising its integrity (Sorek, 2007). Moreover, they increase genome plasticity and provide beneficial effects for the species by providing alternative promoters (Akopov et al., 1998; Sverdlov, 2000). Most protein-coding genes in humans are regulated by multiple distinct promoters, suggesting that promoter choice is as important as the level of transcriptional activity. Transcriptome diversity is the key to cellular identity. Although most HERV elements appear inactive, some are still transcribed and translated in specific human tissues (Lower et al., 1996). In our previous study, we examined the LTR10A element located upstream of the original promoter sequence of *NOS3*. Expression analysis

using RT-PCR and reporter gene assays in HCT116 and COS7 cells have demonstrated that placenta-specific expression of *NOS3* is driven by the LTR10A-derived alternative promoter (Huh et al., 2008). Alternative transcripts (*FPR3-1* and *FPR3-2*) generated by the LTR54 element have been reported to show tissue-specific patterns with strong expression in the human lung or uterus, whereas the *FPR3-1* transcript in rhesus macaque is broadly expressed in various tissues (Ha et al., 2011). Bioinformatics analyses have revealed that the LTR12C element has multiple transcription factor Y (*NF-Y*) and that the promoter activity of LTR12C is significantly decreased after *NF-Y* knockdown (Jung et al., 2017). Twelve alternative transcripts of *PCDH11X/Y* in relation to TEs have also been identified by *in silico* analysis, indicating that dominant expression patterns are present in several tissues (Tx1-fetal liver, Tx3-adult brain, Tx4-adult brain and kidney, Tx5-bone marrow, Ty1-fetal brain, and Ty2-adrenal glands). Tx4 transcripts show specific expression patterns in olfactory tissues (Ahn et al., 2010). The expression of HERV LTRs varies significantly in various cell lines and shows strict cell-type specificity in some cases

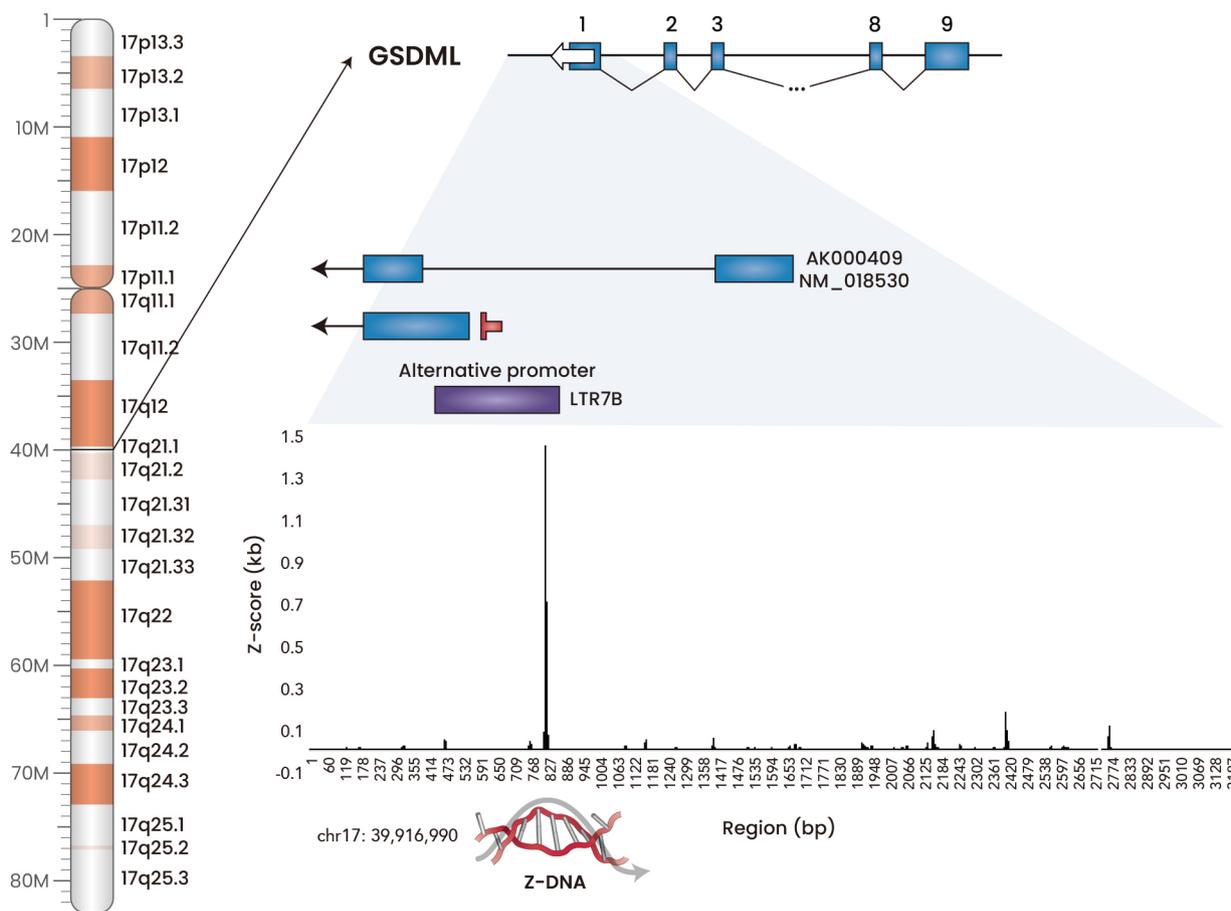


Fig. 3. Expression of the transcript variant (NM_018530) of GSDML regulated by integration of the LTR7B element as an alternative promoter. Z-hunt detected a high Z-score band of 1.5 within the LTR7B element, which overlapped with Z-DNA.

(Schon et al., 2001). We thus summarized functional genes containing alternative promoters derived from HERV and solitary LTR elements overlapping Z-DNA prediction sites (Table 1, Fig. 2). Among 72 known LTR-derived gene promoters, 19 (26.39%) show ZFS in the Z-hunt analysis. Placenta-specific expression of insulin-like 4 (*INSL4*) is mediated by the 3' LTR of the HERV element, and the latter may play a major role in *INSL4* upregulation during human cytotrophoblast differentiation into syncytiotrophoblasts (Bieche et al., 2003). *GSDML* (gasdermin-like protein), located on human chromosome 17q21.1, is an oncogenic recombination hotspot. We previously identified the LTR element of HERV-H with reverse orientation, which acts as an alternative promoter of *GSDML*, and analyzed its expression pattern in human tissues and cancer cells. The transcripts of this LTR7B-derived promoter were found to be widely distributed in various human tissues and cancer cells, whereas transcripts of the cellular promoter were found only in stomach tissues. A reporter gene assay for the promoter activity of LTR7B on the *GSDML* in HCT-116, HeLa, and Cos7 cells revealed that the LTR7B promoter with reverse orientation had a stronger promoter activity compared with that of the forward promoter (Sin et al., 2006). These findings suggest that a new transcript variant of *GSDML* was formed by integrating the antisense-oriented HERV-H LTR element (possibly forming Z-DNA) during hominoid evolution (Huh et al., 2008; Sin et al., 2006). HERV-H LTR sequences were found to positively regulate the transcriptional activity of *GSDML*. In a transient transfection assay, deletion of the U5 region resulted in a significant decrease in the transcriptional activity of *GSDML* (Huh et al., 2008). As shown in Fig. 3, a transcript variant (NM 018530) appeared upon integration of the LTR7B element. Within the LTR7B element, a high Z-score 1.5 band was determined using Z-hunt, overlapped with Z-DNA. Taken together, genomic integration by antisense-oriented HERV and solitary LTR elements results in Z-DNA, that

acts as a regulatory element, such as a promoter or enhancer. This kind of alternative promoter or enhancer can play an important biological role in human cells, including recruitment of transcription factors, regulation of gene expression, and control of genome instability, resulting in biodiversity.

HERV AND SOLITARY LTR ELEMENTS ARE IMPLICATED IN VARIOUS HUMAN DISEASES

HERV and solitary LTR elements can cause several human diseases such as azoospermia, multiple sclerosis, schizophrenia, diabetes, and cancer (Conrad et al., 1997; Kamp et al., 2000; Karlsson et al., 2001; Kim et al., 1999a; 1999b; Li et al., 2020; Patzke et al., 2002; Perron et al., 1997; Xiao and Xu, 2021). Apolipoprotein C1 (*APOC1*) appears to be an independent prognostic factor in patients with clear cell renal cell carcinoma (ccRCC). *APOC1* could be a potential therapeutic target for ccRCC as it regulates cell growth and metastasis pathways (Xiao and Xu, 2021). *APOC1* promotes metastasis of ccRCC cells by activating *STAT3*. Moreover, the metastatic potential of ccRCC cells driven by *APOC1* is suppressed by DPP-4 inhibition (Li et al., 2020). Human transcripts containing HERV-E LTRs are fused to the *APOC1* and endothelin B receptor (EBR) genes (Medstrand et al., 2001). Alternative transcripts of *APOC1* and EBR are initiated and promoted by LTRs. In contrast to the LTR at the *APOC1* locus, a significant proportion of EBR transcripts is derived from the LTR promoter in the placenta. This type of LTR element seems to have a dual role, acting both as a promoter and an enhancer for the expression of neighboring functional genes in specific tissues (Fig. 4).

LY6K-AS long noncoding RNA (lncRNA) is an anti-sense transcript, known as a prognostic biomarker, which shows elevated expression in patients with lung adenocarcinoma (LUAD); higher expression of *LY6K-AS* in LUAD predicts poor

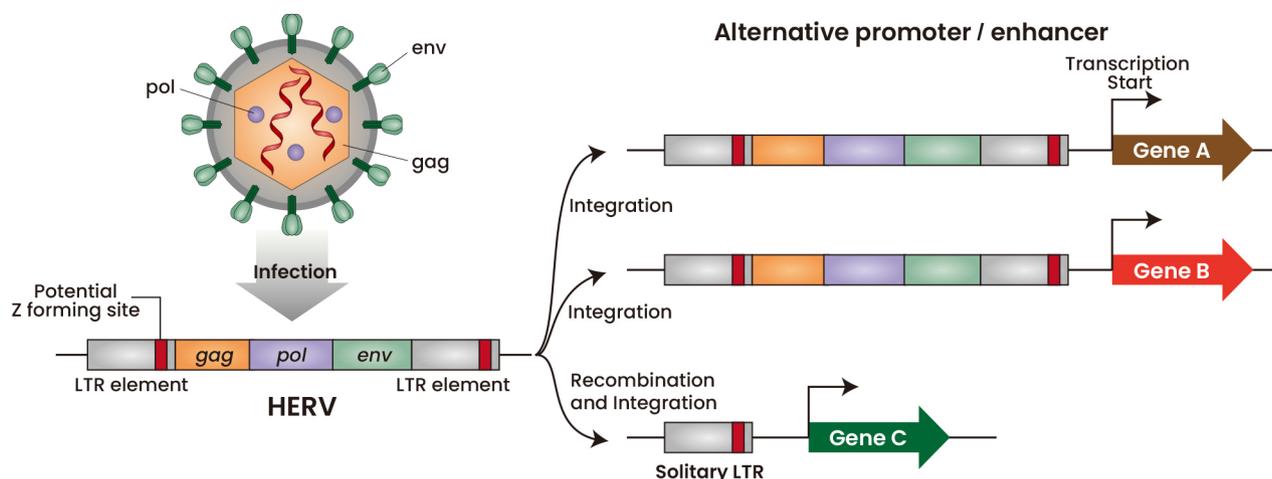


Fig. 4. Schematic illustration of the mechanism that LTRs containing ZFS act as regulatory elements. Some of the HERVs and solitary LTRs caused by exogenous retrovirus infection could have potential ZFS, and be integrated into the neighboring region of functional genes. This integration could result in a Z-DNA conformation, and LTRs containing ZFS might act as alternative promoters or enhancers of a functional gene.

survival outcomes, indicating that *LY6K-AS* silencing is a promising therapeutic option, which inhibits oncogenic mitotic progression in LUAD (Ali et al., 2021). Osteoarthritis (OA) is the most prevalent articulating joint disease in humans and frequently results in joint pain, movement limitations, inflammation, and progressive degradation of the articular cartilage. LncRNAs are involved in multiple cellular and biological processes. Moreover, numerous lncRNAs are differentially expressed in human OA cartilage (Abbasifard et al., 2020). LncRNA-mRNA co-expression analysis has revealed a remarkable relationship, wherein *OTOA* may play a critical role in the differential mechanisms of OA progression between Tibetan and Han Chinese populations (Luo et al., 2021). LncRNA-derived LINC00652 can exert biological functions by co-expression with prognostic genes (*INSL3*, *SNAP91*, and *REN*) and lipid metabolism-related genes (*MIA2*, *APOA1*). Accordingly, this lncRNA-mRNA-based classifier might be clinically useful for predicting the recurrence and prognosis of childhood acute lymphoblastic leukemia (ALL) (Qi et al., 2021).

DYX1C1 is a candidate gene for developmental dyslexia. One of its transcripts has been directly associated with an HERV-H LTR element, and alternatively spliced transcript variants of *DYX1C1* have been demonstrated as potential biomarkers to detect colorectal cancer (Kim et al., 2009; 2011). Moreover, *DYX1C1* expression in breast cancer is associated with several clinicopathological parameters, whereas loss of *DYX1C1* is correlated with a more aggressive disease (Rosin et al., 2012). In our previous study, we detected transcript variants (a and b) of the choroideremia (*CHM*) gene in human cancer cells and tissues. High expression levels of *CHM* isoform b caused by an LTR12C element offering an alternative splicing site were detected in colon and lung cancer cell lines and in tissues of patients with colon cancer (Jung et al., 2011). Identification of alternative spliced variants as biomarkers to distinguish between normal and cancer cells could thus enhance the existing understanding of tumorigenesis. Compared with the adjacent normal tissues, high expression levels of HERV-K were noted in testis tumor tissues, HERV-R in liver and lung tumor tissues, HERV-H in liver, lung, and testis tumor tissues, and HERV-P in colon and liver tumor tissues (Ahn and Kim, 2009). Human *HHLA1* (HERV-H LTR-associating 1) and *OC90* (otoconin-90) are normally expressed independently of different promoters but are expressed from the LTR promoter and are spliced together in teratocarcinoma cells, indicating that the strong activity of the LTR promoter in this cell type could induce transcriptional fusion of these two genes (Kowalski et al., 1999). The *PLA2L* (phospholipase A2-like domain) 5'-HERV-H sequence functions as an abnormally long and complex 5'-UTR, resulting in suppressed translation in both teratocarcinoma cell lines and full-length cDNA transfectants (Kowalski and Mager, 1998). Taken together, HERV and solitary LTR elements have been randomly integrated into neighboring functional genes during primate evolution and have evolved as regulatory elements, such as promoters or enhancers. Moreover, they provide alternative splicing and binding sites for transcription factors that control tissue-specific gene expression and transcript variants in relation to various human diseases.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

ACKNOWLEDGMENTS

This work was supported by a two-year research grant from Pusan National University.

AUTHOR CONTRIBUTIONS

D.H.L., W.H.B., and H.H. conceived the research and performed Z-DNA analyses. E.G.P., Y.J.L., and W.R.K. provided comments. H.S.K. wrote and revised the manuscript.

CONFLICT OF INTEREST

The authors have no potential conflicts of interest to disclose.

ORCID

Du Hyeong Lee	https://orcid.org/0000-0001-8403-4395
Woo Hyeon Bae	https://orcid.org/0000-0003-0126-8271
Hongseok Ha	https://orcid.org/0000-0002-0923-9187
Eun Gyung Park	https://orcid.org/0000-0002-9575-7043
Yun Ju Lee	https://orcid.org/0000-0002-4578-3323
Woo Ryung Kim	https://orcid.org/0000-0003-1840-4444
Heui-Soo Kim	https://orcid.org/0000-0002-5226-6594

REFERENCES

- Abbasifard, M., Kamiab, Z., Bagheri-Hosseiniabadi, Z., and Sadeghi, I. (2020). The role and function of long non-coding RNAs in osteoarthritis. *Exp. Mol. Pathol.* 114, 104407.
- Ahn, K., Huh, J.W., Kim, D.S., Ha, H.S., Kim, Y.J., Lee, J.R., and Kim, H.S. (2010). Quantitative analysis of alternative transcripts of human PCDH11X/Y genes. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 153B, 736-744.
- Ahn, K. and Kim, H.S. (2009). Structural and quantitative expression analyses of HERV gene family in human tissues. *Mol. Cells* 28, 99-103.
- Ali, M.M., Marco, M.D., Mahale, S., Jachimowicz, D., Kosalari, S.T., Reischl, S., Statello, L., Mishra, K., Darnfors, C., Kanduri, M., et al. (2021). LY6K-AS lncRNA is a lung adenocarcinoma prognostic biomarker and regulator of mitotic progression. *Oncogene* 40, 2463-2478.
- Anderssen, S., Sjøttem, E., Svineng, G., and Johansen, T. (1997). Comparative analyses of LTRs of the ERV-H family of primate-specific retrovirus-like elements isolated from marmoset, African green monkey, and man. *Virology* 234, 14-30.
- Akopov, S.B., Nikolaev, L.G., Khil, P.P., Lebedev, Y.B., and Sverdlov, E.D. (1998). Long terminal repeats of repeats of human endogenous retrovirus K family (HERV-K) specifically bind host cell nuclear proteins. *FEBS Lett.* 421, 229-233.
- Bayele, H.K., Peyssonnaud, C., Giatromanolaki, A., Arrais-Silva, W.W., Mohamed, H.S., Collins, H., Giorgio, S., Koukourakis, M., Hohnson, R.S., Blackwell, J.M., et al. (2007). HIF-1 regulates heritable variation and allele expression phenotypes of the macrophage immune response gene *SLC11A1* from a Z-DNA-forming microsatellite. *Blood* 110, 3039-3048.
- Beknazarov, N., Jin, S., and Poptsova, M. (2020). Deep learning approach for predicting functional Z-DNA regions using omics data. *Sci. Rep.* 10, 19134.
- Bieche, I., Laurent, A., Laurendeau, I., Duret, L., Giovannardi, Y., Frendo, J.L., Olivi, M., Fausser, J.L., Evain-Brion, D., and Vidaud, M. (2003). Placenta-specific *INSL4* expression is mediated by a human endogenous retrovirus element. *Biol. Reprod.* 68, 1422-1429.
- Blaho, J.A. and Wells, R.D. (1989). Left-handed Z-DNA and genetic recombination. *Prog. Nucleic Acid Res. Mol. Biol.* 37, 107-126.

- Buzdin, A., Khodosevich, K., Mamedov, I., Vinogradova, T., Lebedev, Y., Hunsmann, G., and Sverdlov, E. (2002). A technique for genome-wide identification of differences in the interspersed repeats integrations between closely related genomes and its application to detection of human-specific integration of HERV-K LTRs. *Genomics* 79, 413-422.
- Conrad, B., Weissmahr, R.N., Boni, J., Arcari, R., Schubach, J., and Mach, B. (1997). A human endogenous retroviral superantigen as candidate autoimmune gene in type I diabetes. *Cell* 90, 303-313.
- Dickerson, R.E., Drew, H.R., Conner, B.C., Wing, R.M., Fratini, A.V., and Kopka, M.L. (1982). The anatomy of A-, B-, and Z-DNA. *Science* 216, 475-485.
- Di Cristofano, A., Strazzullo, M., Longo, L., and La Mantia, G. (1995). Characterization and genomic mapping of the ZNF80 locus: expression of this zinc-finger gene is driven by a solitary LTR of ERV9 endogenous retroviral family. *Nucleic Acids Res.* 23, 2823-2830.
- Drew, H., Takano, T., Tanaka, S., Itakura, K., and Dickerson, R.E. (1980). High-salt d(CpGpCpG), a left-handed Z-DNA double helix. *Nature* 286, 567-573.
- Durnaoglu, S., Lee, S.K., and Ahn, J. (2021). Human endogenous retroviruses as gene expression regulators: insights from animal models into human diseases. *Mol. Cells* 44, 861-878.
- Ha, H.S., Huh, J.W., Gim, J.A., Han, K., and Kim, H.S. (2011). Transcriptional variations mediated by an alternative promoter of the FPR3 gene. *Mamm. Genome* 22, 621-633.
- Hamada, H., Petrino, M.G., and Kakunaga, T. (1982). A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc. Natl. Acad. Sci. U. S. A.* 79, 6465-6469.
- Havecker, E.R., Gao, X., and Voytas, D.F. (2004). The diversity of LTR retrotransposons. *Genome Biol.* 5, 225.
- Herbert, A. (2019). Z-DNA and Z-RNA in human disease. *Commun. Biol.* 2, 7.
- Herbert, A. and Rich, A. (2001). The role of binding domains for dsRNA and Z-DNA in the in vivo editing of minimal substrates by ADAR1. *Proc. Natl. Acad. Sci. U. S. A.* 98, 12132-12137.
- Hervé, C.A., Forrest, G., Löwer, R., Griffiths, D.J., and Venables, P.J.W. (2004). Conservation and loss of the ERV3 open reading frame in primates. *Genomics* 83, 940-943.
- Ho, P.S., Ellison, M.J., Quigley, G.J., and Rich, A. (1986). A computer aided thermodynamic approach for predicting the formation of Z-DNA in naturally occurring sequences. *EMBO J.* 5, 2737-2744.
- Huh, J.W., Kim, D.S., Kang, D.W., Ha, H.S., Ahn, K., Noh, Y.N., Min, D.S., Chang, K.T., and Kim, H.S. (2008). Transcriptional regulation of GSDML gene by antisense-oriented HERV-H LTR element. *Arch. Virol.* 153, 1201-1205.
- Huh, J.W., Kim, D.S., Ha, H.S., Kim, T.H., Kim, W., and Kim, H.S. (2006). Formation of a new solo-LTR of the human endogenous retrovirus H family in human chromosome 21. *Mol. Cells* 22, 360-363.
- Ito, J., Sugimoto, R., Nakaoka, H., Yamada, S., Kimura, T., Hayano, T., and Inoue, I. (2017). Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. *PLoS Genet.* 13, e1006883.
- Jacox, E., Gotea, V., Ovcharenko, I., and Elnitski, L. (2010). Tissue-specific and ubiquitous expression patterns from alternative promoters of human genes. *PLoS One* 5, e12274.
- Jern, P. and Coffin, J.M. (2008). Effects of retroviruses on host genome function. *Annu. Rev. Genet.* 42, 709-732.
- Jung, Y.D., Huh, J.W., Kim, D.S., Kim, Y.J., Ahn, K., Ha, H.S., Lee, J.R., Yi, J.M., Moon, J.W., Kim, T.O., et al. (2011). Quantitative analysis of transcript variants of CHM gene containing LTR12C element in humans. *Gene* 489, 1-5.
- Jung, Y.D., Lee, H.E., Jo, A., Hiroo, I., Cha, H.J., and Kim, H.S. (2017). Activity analysis of LTR12C as an effective regulatory element of the RAE1 gene. *Gene* 634, 22-28.
- Kamp, C., Hirschmann, P., Voss, H., Huellen, K., and Vogt, P.H. (2000). Two long homologous retroviral sequence blocks in proximal Yq11 cause AZFa microdeletions as a result of intrachromosomal recombination events. *Hum. Mol. Genet.* 9, 2563-2572.
- Karlsson, H., Bachmann, S., Schroder, J., McArthur, J., Torrey, E.F., and Yolken, R.H. (2001). Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4634-4639.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* 12, 996-1006.
- Kim, H.S. (2012). Genomic impact, chromosomal distribution and transcriptional regulation of HERV elements. *Mol. Cells* 33, 539-544.
- Kim, H.S., Takenaka, O., and Crow, T.J. (1999a). Isolation and phylogeny of endogenous retrovirus sequences belonging to the HERV-W family in primates. *J. Gen. Virol.* 80, 2613-2619.
- Kim, H.S., Wadekar, R.V., Takenaka, O., Winstanley, C., Mitsunaga, F., Kageyama, T., Hyun, B.H., and Crow, T.J. (1999b). SINE-R.C2 (a Homo sapiens specific retroposon) is homologous to cDNA from postmortem brain in schizophrenia and to two loci in the Xq21.3/Yp block linked to handedness and psychosis. *Am. J. Med. Genet.* 88, 560-566.
- Kim, T.H., Jeon, Y.J., Yi, J.M., Kim, D.S., Huh, J.W., Hur, C.G., and Kim, H.S. (2004). The distribution and expression of HERV families in the human genome. *Mol. Cells* 18, 87-93.
- Kim, Y.J., Huh, J.W., Kim, D.S., Bae, M.I., Lee, J.R., Ha, H.S., Ahn, K., Kim, T.O., Song, G.A., and Kim, H.S. (2009). Molecular characterization of the DYX1C1 gene and its application as a cancer biomarker. *J. Cancer Res. Clin. Oncol.* 135, 265-270.
- Kim, Y.J., Huh, J.W., Kim, D.S., Han, K., Kim, H.M., and Kim, H.S. (2011). Evolutionary diversification of DYX1C1 transcripts via an HERV-H LTR integration event. *Genes Genet. Syst.* 86, 277-284.
- Kjellman, C., Sjögren, H.O., and Widegren, B. (1999). HERV-F, a new group of human endogenous retrovirus sequences. *J. Gen. Virol.* 80, 2383-2392.
- Kowalski, P.E., Freeman, J.D., and Mager, D.L. (1999). Intergenic splicing between a HERV-H endogenous retrovirus and two adjacent human genes. *Genomics* 57, 371-379.
- Kowalski, P.E. and Mager, D.L. (1998). A human endogenous retrovirus suppresses translation of an associated fusion transcript, PLA2L. *J. Virol.* 72, 6164-6168.
- Landry, J.R., Mager, D.L., and Wilhelm, B.T. (2003). Complex controls: the role of alternative promoters in mammalian genomes. *Trends Genet.* 19, 640-648.
- Lavie, L., Medstrand, P., Schempp, W., Meese, E., and Mayer, J. (2004). Human endogenous retrovirus family HERV-K (HML-5): status, evolution, and reconstruction of an ancient betaretrovirus in the human genome. *J. Virol.* 78, 8788-8798.
- Lee, J.W. and Kim, H.S. (2006). Endogenous retrovirus HERV-I LTR family in primates: sequences, phylogeny, and evolution. *Arch. Virol.* 151, 1651-1658.
- Li, H., Xiao, J., Li, J., Lu, L., Feng, S., and Droge, P. (2009). Human genomic Z-DNA segments probed by the Z alpha domain of ADAR1. *Nucleic Acids Res.* 37, 2737-2746.
- Li, Y.L., Wu, L.W., Zeng, L.H., Zhang, Z.Y., Wang, W., Zhang, C., and Lin, N.M. (2020). ApoC1 promotes the metastasis of clear cell renal cell carcinoma via activation of STAT3. *Oncogene* 39, 6203-6217.
- Liu, L.F. and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. U. S. A.* 84, 7024-7027.
- Lower, R., Lower, J., and Kurth, R. (1996). The viruses in all of us: characteristics and biological significance of human endogenous

- retrovirus sequences. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 5177-5184.
- Luo, J., Luo, X., Duan, Z., Bai, W., Che, X., Shan, Z., Li, X., and Peng, J. (2021). Comprehensive analysis of lncRNA and mRNA based on expression microarray profiling reveals different characteristics of osteoarthritis between Tibetan and Han patients. *J. Orthop. Surg. Res.* **16**, 133.
- Macfarlane, C. and Simmonds, P. (2004). Allelic variation of HERV-K(HML-2) endogenous retroviral elements in human populations. *J. Mol. Evol.* **59**, 642-656.
- Medstrand, P., Landry, J.R., and Mager, D.L. (2001). Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. *J. Biol. Chem.* **276**, 1896-1903.
- Medstrand, P. and Mager, D.L. (1998). Human-specific integrations of the HERV-K endogenous retrovirus family. *J. Virol.* **72**, 9782-9787.
- Montension, M., Williams, Z.H., Subramanian, R.P., Kuperwasser, C., and Coffin, J.M. (2018). Promoter expression of HERV-K (HML-2) provirus-derived sequences is related to LTR sequence variation and polymorphic transcription factor binding sites. *Retrovirology* **15**, 57.
- Patzke, S., Lindeskog, M., Munthe, E., and Aasheim, H.C. (2002). Characterization of a novel human endogenous retrovirus, HERV-H/F, expressed in human leukemia cell lines. *Virology* **303**, 164-173.
- Perron, H., Garson, J., Bedin, F., Beseme, F., Paranhos-Baccala, G., Komurian-Pradel, F., Mallet, F., Tuke, P.W., Voisset, C., Blond, J.L., et al. (1997). Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 7583-7588.
- Qi, H., Chi, L., Wang, X., Jin, X., Wang, W., and Lan, J. (2021). Identification of a seven-lncRNA-mRNA signature for recurrence and prognostic prediction in relapsed acute lymphoblastic leukemia based on WGCNA and LASSO analyses. *Anal. Cell. Pathol. (Amst.)* **2021**, 6692022.
- Ray, B.K., Dhar, S., Shakya, A., and Ray, A. (2011). Z-DNA-forming silencer in the first exon regulates human ADAM-12 gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 103-108.
- Rosin, G., Hannelius, U., Lindstrom, L., Hall, P., Bergh, J., Hartman, J., and Kere, J. (2012). The dyslexia candidate gene DYX1C1 is a potential marker of poor survival in breast cancer. *BMC Cancer* **12**, 79.
- Ruda, V.M., Akopov, S.B., Trubetskoy, D.O., Manuylov, N.L., Vetchinova, A.S., Zavalova, L.L., Nikolaev, L.G., and Sverdlov, E.D. (2004). Tissue specificity of enhancer and promoter activities of a HERV-K(HML-2) LTR. *Virus Res.* **104**, 11-16.
- Schon, U., Diem, O., Leitner, L., Gunzburg, W.H., Mager, D.L., Salmons, B., and Leib-Mosch, C. (2009). Human endogenous retroviral long terminal repeat sequences as cell type-specific promoters in retroviral vectors. *J. Virol.* **83**, 12643-12650.
- Schon, U., Seifarth, W., Baust, C., Hohenadl, C., Erfle, V., and Leib-Mosch, C. (2001). Cell type-specific expression and promoter activity of human endogenous retroviral long terminal repeats. *Virology* **279**, 280-291.
- Schroth, G.P., Chou, P.J., and Ho, P.S. (1992). Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes. *J. Biol. Chem.* **267**, 11846-11855.
- Shin, S.I., Ham, S., Park, J., Seo, S.H., Lim, C.H., Jeon, H., Huh, J., and Roh, T.Y. (2016). Z-DNA-forming sites identified by ChIP-Seq are associated with actively transcribed regions in the human genome. *DNA Res.* **23**, 477-486.
- Sin, H.S., Huh, J.W., Kim, D.S., Kang, D.W., Min, D.S., Kim, T.H., Ha, H.S., Kim, H.H., Lee, S.Y., and Kim, H.S. (2006). Transcriptional control of the HERV-H LTR element of the GSDML gene in human tissues and cancer cells. *Arch. Virol.* **151**, 1985-1994.
- Sorek, R. (2007). The birth of new exons: mechanisms and evolutionary consequences. *RNA* **13**, 1603-1608.
- Sverdlov, E.D. (2000). Retroviruses and primate evolution. *Bioessays* **22**, 161-171.
- Thomas, J., Perron, H., and Feschotte, C. (2018). Variation in proviral content among human genomes mediated by LTR recombination. *Mob. DNA* **9**, 36.
- Wahls, W.P., Wallace, L.J., and Moore, P.D. (1990). The Z-DNA motif d(TG)30 promotes reception of information during gene conversion events while stimulating homologous recombination in human cells in culture. *Mol. Cell. Biol.* **10**, 785-793.
- Wang, A.J., Quigley, G.J., Kolpak, F.J., van der Marel, G., van Boom, J.H., and Rich, A. (1981). Left-handed double helical DNA: variations in the backbone conformation. *Science* **211**, 171-176.
- Xiao, H. and Xu, Y. (2021). Overexpression of apolipoprotein C1 (APOC1) in clear cell renal cell carcinoma and its prognostic significance. *Med. Sci. Monit.* **27**, e929347.
- Yi, J.M. and Kim, H.S. (2007a). Expression and phylogenetic analyses of human endogenous retrovirus HC2 belonging to the HERV-T family in human tissues and cancer cells. *J. Hum. Genet.* **52**, 285-296.
- Yi, J.M. and Kim, H.S. (2007b). Molecular phylogenetic analysis of the human endogenous retrovirus E (HERV-E) family in human tissues and human cancers. *Genes Genet. Syst.* **82**, 89-98.
- Yi, J.M., Kim, T.H., Huh, J.W., Park, K.S., Jang, S.B., Kim, H.M., and Kim, H.S. (2004). Human endogenous retroviral elements belonging to the HERV-S family from human tissues, cancer cells, and primates: expression, structure, phylogeny and evolution. *Gene* **342**, 283-292.
- Yi, J.M., Schuebel, K., and Kim, H.S. (2007c). Molecular genetic analyses of human endogenous retroviral elements belonging to the HERV-P family in primates, human tissues, and cancer cells. *Genomics* **89**, 1-9.
- Yu, H.L., Zhao, Z.K., and Zhu, F. (2013). The role of human endogenous retroviral long terminal repeat sequences in human cancer. *Int. J. Mol. Med.* **32**, 755-762.
- Zimmer, C., Tymen, S., Marck, C., and Guschlbauer, W. (1982). Conformational transitions of poly(dA-dC) (poly(dG-dT)) induced by high salt or in ethanolic solution. *Nucleic Acids Res.* **10**, 1081-1091.