RESEARCH ARTICLE

EnHERV: Enrichment analysis of specific human endogenous retrovirus patterns and their neighboring genes

Pumipat Tongyoo[1], Yingyos Avihingsanon[2,3], Santhitham Prom-On[4], Apiwat Mutirangura[5], Wuttichai Mhuantong[6], Nattiya Hirankarn[2,7] *

1 Inter-Department Program of Biomedical Sciences, Faculty of Graduate School, Chulalongkorn University, Bangkok, Thailand, 2 Center of Excellence in Immunology and Immune Mediated Diseases, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand, 3 Division of Nephrology, Department of Medicine, Chulalongkorn University and King Chulalongkorn Memorial Hospital, Bangkok, Thailand, 4 Computer Engineering Department, Faculty of Engineering, King Mongkut's University of Technology Thonburi, Bangmod, Thungkhru, Bangkok, Thailand, 5 Center of Excellence of Molecular Genetics of Cancer and Human Diseases, Department of Anatomy, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand, 6 Enzyme Technology Laboratory, Microbial Biotechnology and Biochemicals Research Unit, National Center for Genetic Engineering and Biotechnology (BIOTEC), Khlong Luang, Pathum Thani, 7 Department of Microbiology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

* Nattiya.H@gmail.com

Data Availability Statement: The tool is available at http://sysbio.chula.ac.th/enherv/ HERV neighboring gene according to specific characteristic. HERV LTRs that used in this paper were listed in Table 2 in the publication. Four solo LTR characteristics were tested; Sense solo LTR, Anti-sense solo LTR, Intragenic solo LTR, Intergenic solo LTR Up- and down-regulated genes in various disease condition (List in S1 Table) were retrieved from GEO. They were also provided as example gene lists in the download section. The download link is

Abstract

Human endogenous retroviruses (HERVs) are flanked by long terminal repeats (LTRs), which contain the regulation part of the retrovirus. Remaining HERVs constitute 7% to 8% of the present day human genome, and most have been identified as solo LTRs. The HERV sequences have been associated with several molecular functions as well as certain diseases in human, but their roles in human diseases are yet to be established. We designed EnHERV to make accessible the identified endogenous retrovirus repetitive sequences from Repbase Update (a database of eukaryotic repetitive elements) that are present in the human genome. Defragmentation process was done to improve the Repeat-Masker annotation output. The defragmented elements were used as core database in EnHERV. EnHERV is available at http://sysbio.chula.ac.th/enherv and can be searched using either gene lists of user interest or HERV characteristics. Besides the search function, EnHERV also provides an enrichment analysis function that allows users to perform enrichment analysis between selected HERV characteristics and user-input gene lists, especially genes with the expression profile of a certain disease. EnHERV will facilitate exploratory studies of specific HERV characteristics that control gene expression patterns related to various disease conditions. Here we analyzed 25 selected HERV groups/names from all four HERV superfamilies, using the sense and anti-sense directions of the HERV and gene expression profiles from 49 specific tissue and disease conditions. We found that intragenic HERVs were associated with down-regulated genes in most cancer conditions and in psoriatic skin tissues and associated with up-regulated genes in immune cells particularly from systemic lupus erythematosus (SLE) patients. EnHERV allowed the analysis of how different types of LTRs were differentially associated with specific gene expression

profiles in particular disease conditions for further studies into their mechanisms and functions.

Introduction

The human genome carries virus genetic content and is therefore part virus, similarly in various eukaryote genomes [1]. They have been known as interspersed repetitive sequences (IRSs) or transposable elements (TEs) because they can be copied or cut and then placed in other regions of the human genome. TEs are found in approximately 45% of the human genome. The ability of TEs to move within genomes impact host genome evolution [2]. Many recent studies have researched abilities of these element and how they contribute in their host gene regulation activities [3]. TEs can be classified as DNA transposons or retroelements and they encompass about 2.8% and 42.2% of the human genome, respectively [4]. Retroelements can be divided into two groups based on the presence or absence of long terminal repeats (LTRs). There are two types of non-LTR retroelements, short interspersed nuclear elements (SINEs, e.g. ALU) and long interspersed nuclear elements (LINEs), which are present in high numbers. The majority of LTR retroelements are derived from human endogenous retroviruses (HERVs), and about 8% of the human genome is made of LTR retrotransposons. While a variety of LTR retrotransposons have been identified, only vertebrate-specific endogenous retroviruses (ERVs) are known to be active in mammalian genomes [5]. Among them, HERVs have similar genomic structures to proviruses but contain a large number of mutations that have accumulated over evolution, especially in internal genes [1,2]. Some remaining HERV elements in a host genome are still active in their host genome.

The abundance and distribution of HERVs have been well characterized [6,7]. They have been known as junk DNA for a long time but more and more studies support their important regulatory function in the human genome e.g., the long terminal repeats of HERVH function as enhancers and a nuclear long noncoding RNA required to maintain hESC identity [8] and their contribution into the core regulatory network of embryonic stem cells [9]. The HERV fragments that remain in the human genome still have the ability to produce functional retroviral proteins e.g., the HERVW GAG protein detected in human brain, HERV-W7q ENV (Syncitin) expressed in placenta, and HERV-K (HML-2) loci encode retrovirus-like proteins expressed in tumor [10–12]. Moreover, their LTRs were shown to function as parts of regulatory sequences e.g., HERV-K LTR found as coexpression with MITF-M in malignant melanomas [3], MER21A/ERV1 acted as a primary promoter of *HSD17B1* in ovary and placenta [6], expression of *ZNF80* zinc-finger gene was driven by a solitary LTR of ERV9 [11]. A genome-wide screen identified more than 20,000 candidate regulatory regions derived from retrotransposons in the human genome and more than 2,000 examples of bidirectional transcription, emphasizing the regulatory role of retrotransposon in the mammalian genome [13]. Bioinformatic profiling and high-throughput experiments help speed up the discovery and one recent study identified ~110,000 regulatory active HERV elements that might impact the molecular function in human cells [14].

Generally, a complete HERV element is composed of two LTRs flanking a set of internal retroviral genes and can be represented as LTR1-Internal-LTR2. HERVs are frequently incorrectly annotated as complete elements because of the massive accumulation of insertions and deletions in HERV sequences. As a result, HERV annotations are often displayed as a number of fragments of an element rather than as a unified sequence with gaps [15]. Solitary LTR is

the most abundant HERV annotation in the human genome because the recombination event between the 5′ and 3′ LTRs of a full-length provirus results in the loss of the internal sequence. The structure and distribution of TEs relative to other genes in the genome may help detect genomic elements that contribute to the development of phenotypic differences between disease and healthy individuals. The role of HERVs in human disease has been discussed especially in cancer and numerous autoimmune, neurological and infectious diseases. The expression level of HERV-E gag (group antigen) was found to be increased in peripheral blood mononuclear cells (PBMCs) of systemic lupus erythematosus (SLE) patients and increased HERV-K gag gene expression was reported in rheumatoid arthritis patients [16]. Furthermore, HERV-E gag transcription correlated with blood plasma concentrations of anti-U1 ribonucleoprotein (RNP) and anti-Sm antibodies in SLE patients. A HERV element was shown to participate in splicing of pre-mRNA to mRNA in SLE patients [17]. Nakkuntod and colleagues [18] examined the methylation status of two HERV-E and HERV-K sequences in lymphocytes from patients with SLE and found that hypomethylation of specific HERVs was a feature for SLE patients. One hypothesis is that lower methylation levels allow for expression of HERV genes, which may have some biological consequences. For example, 1) aberrant HERV transcripts and their protein products might lead to the production of auto-antibodies due to molecular mimicry, 2) HERV mRNAs might serve as foreign nucleic acids and stimulate an abnormal immune response via endogenous immune receptors, or 3) regulatory regions, such as LTRs, in the HERVs can affect neighboring gene expression. Investigating the possible relations between a gene set and HERVs is important in identifying novel disease pathogenesis.

Although HERVs have been known for more than two decades, there is only a limited number of databases that facilitate finding HERVs in the human genome. Existing databases including 1) HERVd [19], designed to search for the location of HERV elements in the genome but information in the tool are quite outdated. RepeatMasker in HERVd is based on the 09/20/2000 version. 2.) Transpogene [20], which allows users to search for intragenic transposable elements in human transcripts. This version of human genome is based on NCBI build 36.1 (UCSC hg18). None provide an enrichment analysis function for gene of interest. Moreover, the location of HERVs in the human genome relative to coding exons can affect their function as well; therefore, we designed EnHERV to provide HERV neighboring gene information as intergenic and intragenic HERV elements. Not only does it provide HERV location, it also makes available the orientation of HERV elements and types of truncation patterns which result from the defragmentation process. Furthermore, EnHERV allows the user to define a distance of intergenic HERV elements from their neighboring genes from ranges of 1 to 100 kbs. Besides the search function in EnHERV, the enrichment analysis function provides association analysis between designed HERV characteristics and their neighboring genes in certain gene expression conditions.

Results and discussion

HERV identification

Only data of the main chromosomes (chromosome 1–22, X, and Y) were included in the analysis because comprehensive genome annotation information is only available for the main chromosomes. We retrieved 687,420 HERV elements from a total of 5,298,130 repeat sequence records in the human hg19/GRCh37 genome (12.97% of the repeat sequences). Most of UCSC cross-reference sequences in the 24 main chromosomes (94.91% of total UCSC known genes) contained HERV elements. The cross-reference sequence annotation (kgXref table) was used to convert UCSC known genes to HGNC official gene names. We investigated the association

Table 1. Solo LTRs used in the enrichment analysis in EnHERV.

Superfamily	Family	Name/Group
ERV1	ERV9	LTR12, LTR12C
	HERVH	LTR7
	HERVW	LTR2, LTR2B, LTR2C
	HUERSP1	LTR8
	LOR1	LOR1a
	MER39	MER39
	MER4	MER4C
	MER52	MER52A
	MER57	MER57B1
ERVK	HERVK10/HERVK (HML-2)	LTR5_Hs
	HERVK11/HERVK (HML-8)	MER11C
ERVL	HERV16	LTR16A1
	HERVL33	LTR33
	HERVL52	LTR52
	HERVL66	LTR66
ERVL-MaLR	MLT	MLT1D, MLT2B3
	MST	MSTD
	THE1	THE1A, THE1B, THE1C, THE1D

https://doi.org/10.1371/journal.pone.0177119.t001

of HERV under various disease conditions in this study. A list of HERV superfamilies and families is shown in Table 1. The association was done on three levels. For superfamilies, all four superfamilies were investigated. While some HERV families and individual HERVs were selected to represent their members in the group as listed in Table 1. The proportion of HERV fragments in the annotation data is shown in Fig 1. ERVL-MaLR is the most abundant HERV in the human genome. While only a little ERVK are present in our genome.

In a post-processing step, we used REannotate [21] to defragment the HERV annotations before using them in EnHERV. Defragmentation was based on distance and orientation between fragments and HERV families. Components of HERV elements were separated into three parts of the complete genomic structure, which can be symbolized as LTR1-Internal-LTR2, where LTR1, Internal, and LTR2 represent an upstream LTR, an internal sequence, and a downstream LTR, respectively. Intactness ratios of the elements, which indicate how complete an element is, were also provided by REannotate. Typically, this value was calculated from the fraction of the reference sequence that matched in the query sequence.

The maximum distance is an important REannotate parameter, which is used to set the greatest distance allowed between two HERV fragments for them to be joined into the same element. REannotate was run several times using different distance parameters to determine the sensitivity of this value as mentioned in materials and methods. Defragmented elements are HERV elements that originated from combining more than one fragment together; the non-defragmented elements were composed of only one fragment; and the total number of all elements that resulted from defragmentation was the sum of the numbers of the defragmented and non-defragmented elements. Rate changes of all elements resulting from defragmentation using different distance parameter values are illustrated in Fig 2. Numbers of all elements and non-defragmented elements tended to decrease when the distance parameter was increased because more single fragments had to be used in the joining events, which resulted in more defragmented elements.
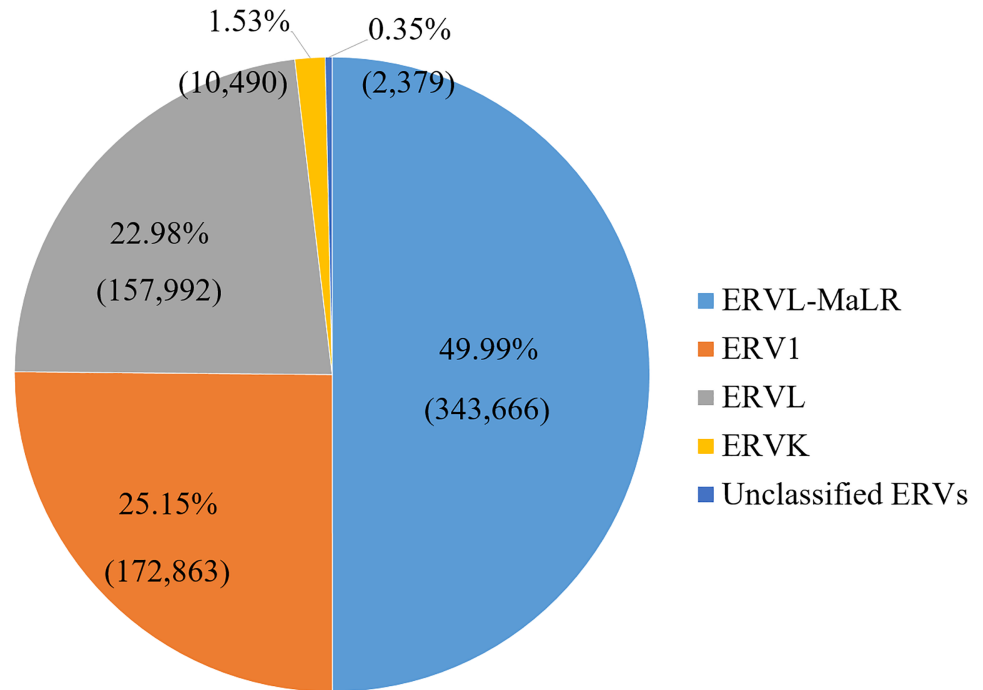
Fig 1. HERV superfamily distribution. Percentage and number of copy of each HERV superfamily in hg19/GRCh37 genome. 687,420 HERV elements were distributed into 5 groups including 4 superfamilies and one unclassified group. Almost 50% of HERV elements in the human genome belongs to the ERVL-MaLR superfamily.

https://doi.org/10.1371/journal.pone.0177119.g001
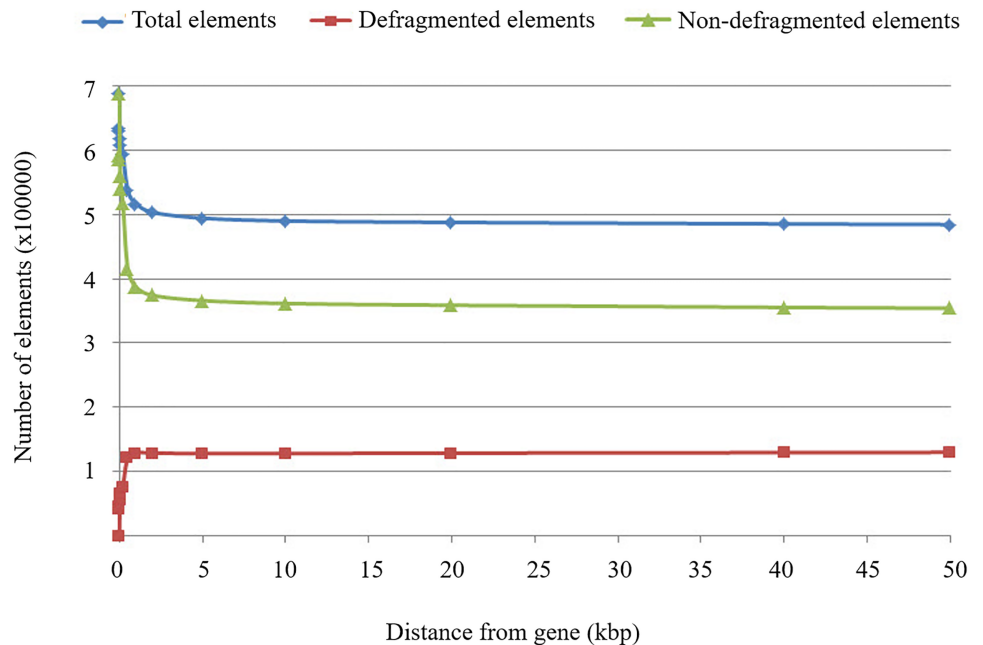
Fig 2. Number of defragmented and non-defragmented elements. Graph shows a number of defragmented and non-defragmented elements in relation to distance parameters.

https://doi.org/10.1371/journal.pone.0177119.g002

Table 2. The number of defragmented elements, resulting from HERV defragmentation using REannotate with different values of distance parameters, and their coverage percentages.

The maximum distances (bp)	The number of defragmented elements	Coverage percentages
10	42,489	32.67%
20	45,678	35.12%
50	57,284	44.04%
100	66,500	51.13%
200	77,070	59.26%
300	91,343	70.23%
400	111,250	85.53%
500	122,437	94.14%
1,000	129,607	99.65%
2,000	129,007	99.19%

The number of all elements tended to change rapidly at smaller distances. This may suggest that the optimal distance should not be too small because many defragmented elements would be ignored. We also measured the rates of change of the number of all elements by varying distance parameters to determine the distance at which there were no changes in the number of all elements. Although, no standard value was available for the distance parameter in defragmentation, suggestions from previous studies have varied between 500 bps to 30 kb [22,23]. Based on these results in Table 2, we set the appropriate value of the distance parameter as 500 bps to cover 122,437 defragmented elements (94.14% of the maximum number of defragmented elements). In total, we obtained 537,061 HERV elements from the defragmentation process, which were 10.26% of repeat sequences in UCSC hg19 main chromosomes. Of the total number of HERVs, 86% were annotated as LTRs and the remaining 14% were annotated as internal genes.

A HERV family and group manipulation process was performed to avoid redundant family names in repeat annotations. According to the four superfamilies, we obtained a total of 413 groups of HERVs, which were classified into 133 HERV families. The full annotation is listed in S4 Table. Notably, most of the existing HERVs were rarely found as complete elements due to the accumulations of insertions and deletions in their sequences over time.

According to the 5′-LTR1-Internal-LTR2-3′ structure of the HERVs, five classification types of the truncation patterns were detected: 1) complete, 2) 5′-truncated, 3) 3′-truncated, 4) both 5′- and 3′-truncated elements, and 5) solitary or solo LTRs. The proportion of each truncation type is shown in Fig 3. The majority of truncation patterns were solo LTRs. Since LTRs may drive the transcription of adjacent host genomic sequences [24], we developed EnHERV to analyze various HERVs patterns that may be associated with gene expression patterns in certain disease conditions.

The manipulated HERV defragmented elements from the REannotate output and the selected gene annotations were mapped together. HERV defragmented elements located between 100,000 bps upstream from the transcription start site and 100,000 bps downstream from the transcription termination site of the neighboring gene were considered as HERV neighboring genes as they have been known to act upon genes up to 70–100 kb away [25–27]. The results of this integration of HERV elements and human genes showed that 382,662 HERV elements (71.24% of the total 537,061 elements) were identified in 73,645 gene isoforms (99.98% of the total 73,660 UCSC gene isoforms), which may imply that most gene isoforms in the human genome contain HERV elements near or within the genes.
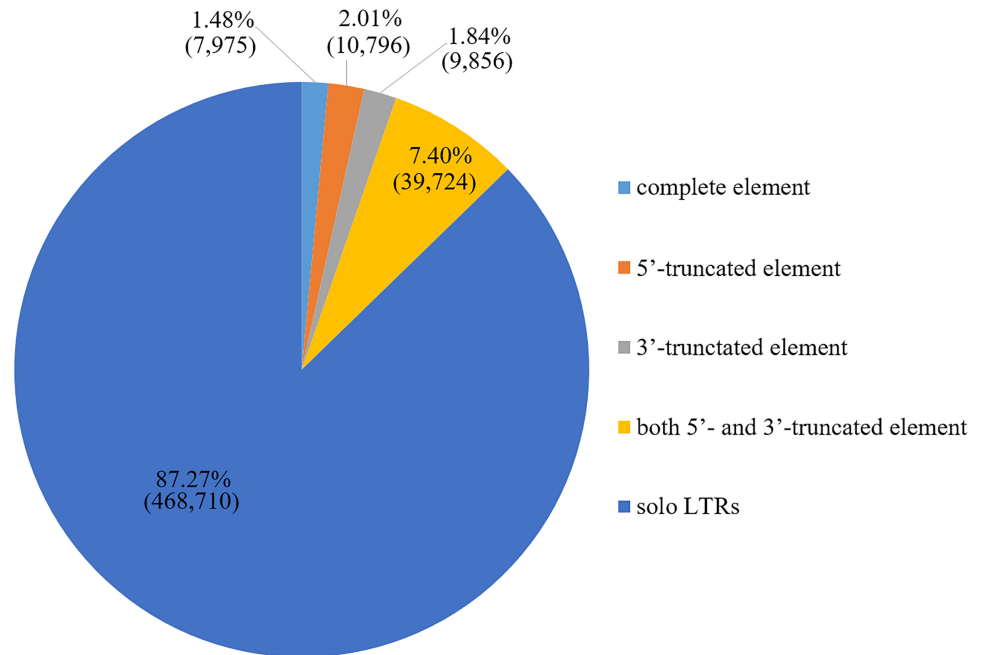
Fig 3. HERV truncation pattern distribution. Percentage and number of each HERV truncation pattern in the HERV elements resulting from the defragmentation process.

https://doi.org/10.1371/journal.pone.0177119.g003

HERV neighboring gene profiles

We identified 382,662 HERV elements (71.24%) as neighboring loci of 73,645 gene isoforms (99.98%) in the human genome. There was a slight bias in the anti-sense orientation of HERV elements and their neighboring genes comparing to sense orientation which is 206,637 elements (54.05%) and 176,025 elements (45.95%), respectively. 5′ upstream and 3′ downstream regions of the gene which are 160,718 elements (42%) and 168,371 elements (44%), respectively. Most of the remaining intragenic HERVs were located in introns. Solo LTRs were the most abundant HERV structure.

The EnHERV database

Users can access EnHERV at http://sysbio.chula.ac.th/enherv. EnHERV provides two search functions: 1) Search by gene(s) and 2) Search by HERV characteristics, including HERV super-family, family, group/name, location in genome, distance from gene (which user defines distance as an option), orientation, and structure completeness. Search results are displayed in table format. EnHERV provides a link to the UCSC genome browser for visualizing the region of the genome structure for the search results. The EnHERV database also allows users to download results for downstream analysis. In the search by gene name option, EnHERV will try to auto-complete a user-input gene name that contains or is surrounded by HERVs. Moreover, EnHERV allows users to download all the records from the database for further customized analyses.

In addition to a searching function, EnHERV also provides an enrichment analysis function that allows users to perform an enrichment analysis between genes with user-specified HERV characteristics and a user-defined gene list. EnHERV will calculate Fisher's p-values and odds ratios for analysis results as mentioned in the methods section. Genes containing the selected

Fig 4. The enrichment analysis page. EnHERV's enrichment analysis page. EnHERV gives the Fisher's exact, P-value and odd ratio of the designed enrichment analysis. EnHERV also provides parallel analysis for sub-selected HERV superfamily/family.

https://doi.org/10.1371/journal.pone.0177119.g004

HERV characteristics will be displayed in a result table, which users can download for further investigation. Furthermore, EnHERV also allows users to perform enrichment analyses for all the members of the selected HERV superfamily/family at once. Users can then save the enrichment analysis output as illustrated in Fig 4.

Analysis of solo LTRs in cancer and autoimmune diseases

Hypomethylated HERVs have been found to be active in cancer and autoimmune diseases [18,28]. Distinct isoforms or gene silencing due to global hypomethylation of HERVs has been reported in various diseases, particularly in cancers [29,30] and SLE [31]. Because HERVs can control neighboring genes by either up- or down-regulating their expression, we developed a model to identify genes that were associated with HERVs in the genome and were

differentially expressed in various diseases. This information can serve as a screening tool for further studies of candidate genes that might be regulated by HERVs.

As an example, we studied differentially expressed gene patterns in various specific tissue and disease conditions. Association analyses were performed in 49 disease conditions as listed in S1 Table. Significant associations were analyzed between various gene expression conditions and the four HERV superfamilies (S2 Table) and 25 individual HERVs that represent each family (S3 Table).

First, significant associations were detected mostly with the intragenic HERVs (S2 and S3 Tables). Second, significant associations with most gene expression conditions were found (Fig 5 and S2 Table). Heatmap in Fig 5 represents the association level of HERVs. The minus

		ERV1	ERVK	ERVL	ERVL-MaLR	ERV1	ERVK	ERVL	ERVL-MaLR
GEO acc.	Gene expression condition		Down	DEG			Up	DEG	
Autoimmune									
GSE4588	SLE B cells	0.0000	-0.6048	-1.2155	-1.8113	-0.2350	-0.0961	-0.3610	-0.7224
GSE10325	SLE B cells	-1.3726	0.0000	-0.5785	-1.3044	-2.8994	0.0000	-2.1878	-1.3173
GSE4588	SLE CD4 cells	-0.6131	-1.7824	-2.2713	-0.9188	-1.3964	-1.4402	-1.0404	-1.3998
GSE10325	SLE T cells	-0.5628	-0.0804	-1.0550	-0.5833	-2.1894	0.0000	-0.6908	-0.9982
GSE13887	SLE T cells		-1.8540			8.0752	-2.7083	7.5808	5.7127
GSE10325	SLE myeloid cells	0.0000	-1.0407	-0.4463	-0.0828		-1.2304	4.4121	
GSE27427	SLE neutrophil cells	-0.9362	0.0000	-0.5002	-0.2135	-2.0741	-0.0974	-0.8054	-0.8023
GSE20864	SLE PBMC	0.0000	-1.3209	-0.9237	-0.962	8.3731	-0.6242	-1.6857	-1.6635
GSE24706	SLE pbmc ANA	0.0000	0.0000	0.0000	0.0000	-0.0459	-0.8137	-0.7236	-0.0897
GSE61635	SLE pbmc RNP	-1.7400	-1.0522	-1.6368	-1.4441	6.2128	11.8109	11.0323	
GSE52471	SLE DLE skin	-2.2788	-0.6275	-3.3074	-2.7931	-1.1145	-1.0548	-0.3311	-0.4032
GSE30153	Inactive SLE B cells	-0.5372	0.0000	-0.5031	0.0000	-0.1352	-0.7343	-0.9707	-0.9886
GSE32591	LN glomolular	-1.4327	-1.5776	-1.0240	-0.0360	-1.1702	-0.9205		-2.3268
GSE32591	LN tubulolar	-1.5138	0.0000	-1.1009	-2.1340	-1.3106	-0.1849	0.0000	-1.2061
GSE1919	RA Synovial Tissues	-1.5292	0.0000	-0.1953	-0.1558	-0.8131	-1.2675	-0.5174	-0.6813
GSE4588	RA B cella	-0.2987	-1.0512	-0.4920	-1.4986	-2.5059	-0.2018	-1.6042	-1.2613
GSE4588	RA CD4 cells	-1.9238	-0.1630	-1.0197	-1.2294	-0.1054	-0.9706	-0.1240	-1.2988
GSE10500	RA macophage	-2.0946	-1.5024	-1.5055	-2.135		-0.3476	3.9314	4.6574
GSE15573	RA PBMC	0.0000	0.0000	0.0000	0.0000	-1.3401	0.0000	-1.3300	-1.6533
GSE13355	Psorisis skin	7.1885	-2.2705	4.3505		-1.2416	-1.5720	-1.0490	-1.0601
GSE14905	Psorisis skin	14.5880		14.9970	14.4910	-2.2781	-1.9684	-1.5258	-1.2704
GSE32407	Psorisis skin	0.0000		0.0000		-0.8583	0.0000	-0.8368	-0.9485
GSE52471	Psorisis skin	5.9291		14.2140	10.8552	-1.1346	-1.9123	-1.0225	-0.3737
GSE71957	Graves CD4 cells	-2.6400		-2.3079		-0.2707	-0.8900	-1.1311	-0.1822
GSE71957	Graves CD8 cells	-1.5387	0.0000	-1.5401	-1.3649	-1.5864	0.0000	-0.9617	-1.5481
GSE27011	Asthma white Blood Cells	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GSE31773	Asthma CD4	-1.0094		-0.1771	0.0000	-1.0271	-0.5501	-0.9541	-1.5644
GSE31773	Asthma CD8	-0.9227	-0.2790	-0.8876	-0.0944	-1.1167		-2.6940	-0.6275
GSE43696	Asthma bronchial Epithelial Cells	-0.4328	-1.3312	-0.2292	-0.0705	-0.0511	-1.1715	-0.0518	-1.0961
Cancer									
GSE45829	B cell EBV infection	-2.6365	-1.2296	-2.0438	-2.1371	7.9646	-0.6638	4.3913	13.0290
GSE36474	Myeloma boneMarrow	-0.0598	-0.8678	-1.0675	-1.1189	-2.6319	-0.3585		-2.4148
GSE12453	Hodgkin Lymphoma vs centroblasts	4.6958	-1.0653	14.0910	15.3890		-3.8960	-2.6381	-2.5226
GSE12453	Hodgkin Lymphoma vs centrocytes	5.7359	-0.8208	14.6270	16.3880		-1.7393		
GSE12453	Hodgkin Lymphoma vs memory B cells	4.2863	-0.0315	5.0524	8.7987		-1.9608		
GSE12453	Hodgkin Lymphoma vs naïve B cells	-1.4172	-0.5764				-2.9099		
GSE12453	Hodgkin Lymphoma vs plasma Cells	-0.7689	-0.1444		9.7251		-2.2831		
GSE12453	Diffuse Large B cells Lymphoma vs centroblasts	10.6870		17.0300	18.3890		-2.3661		
GSE12453	Diffuse Large B cells Lymphoma vs centrocytes	11.5906		14.0368	19.2000		-3.1060		
GSE12453	Diffuse Large B cells Lymphoma vs memory B cells	7.0492		7.1040	9.1570		-2.3256		
GSE12453	Diffuse Large B cells Lymphoma vs naïve B cells	7.1427		10.9111	5.1685				
GSE12453	Diffuse Large B cells Lymphoma vs plasma cells	-0.3720	-0.9988	-1.0119			-2.3600		
GSE1299	Breast Cancer Cells	6.0473		13.1814	7.6492	-0.0256	-0.1701	-1.0815	-0.6695
GSE3167	Bladder Carcinoma Situ				-2.0776		-1.1810		
GSE5764	Ductal And Lobular Breast Cancer	-0.7233	-1.0926	-1.3912	-2.2233	-1.2381	-1.3365	-2.8393	-2.7972
GSE5816	Lung Adrenocarcinoma	-0.0502	-2.4285	-0.9688	-0.3790	-3.0735		4.3941	0.0000
GSE6631	Head and neck	-1.2543	-1.3946	-1.5796	-1.9694	-1.1068	-1.0976		
GSE6919	Metastasis Prostate Cancer		-1.5051	4.1664		-1.0184	-1.1170	-1.0618	-1.1169
GSE9750	Cervical Cancer	9.7184	-2.1438	12.2480	5.5965		-3.6560		
GSE13911	Microsatellite Instable Gastric Cancer	-1.1666	-0.0849	-2.5781	-0.2305		-1.4170		

Fig 5. Association heatmap between intragenic HERV and disease conditions. A different intragenic HERV association pattern between cancer and auto-immune disease was identified. Intragenic HERVs were strongly associated with down-regulated genes in cancer. In contrast they are highly associated with up-regulated genes in immune cells under auto immune disease conditions.

of log P-value were calculated to represent the association level of HERVs and certain gene condition. The darker color represents a stronger association level. Red represents an association to genes in down-regulation conditions while green represents genes in up-regulation conditions. Furthermore, with the P-value < 0.001 and odds ratio > 1 cutoff criteria was mainly with the ERV1, ERVL, and ERVL-MaLR superfamilies but not with the ERVK superfamily (S1 Fig). Third, the pattern of association was different between various disease conditions. We found that intragenic HERVs were associated with down-regulated genes in most cancer conditions and in psoriatic skin tissue and associated with up-regulated genes in immune cells from SLE patients, macrophages from RA patients, and Epstein-Barr virus (EBV) infected B cells (Fig 5, S1 Fig and S2 Table).

This finding is interesting because LINE-1s, which are another type of IRS, have also been shown to be associated with down-regulated genes in cancer. It was suggested that LINE-1s, which were found to be globally hypomethylated in cancer tissue, might control neighboring genes by acting as antisense RNAs [32]. Interestingly, our analysis with HERV in this study, another IRS element that was also reported to be hypomethylated in cancer tissues, revealed that they had the same antisense direction. Further proof is required to determine if HERVs can control neighboring genes in the same way as LINE-1s. Another observation is the similar pattern of association in cancer and psoriatic skin tissues. We previously reported LINE-1 hypomethylation in keratinocytes from psoriatic patients and observed that genes with LINE-1 in their vicinity were down-regulated more than genes without LINE-1, similar to what was observed in cancer [33]. We found a similar association with HERVs in the present study. These findings are interesting because keratinocyte proliferation, a characteristic similar to cancer cells, is the main feature in psoriatic skin tissue. Mechanisms of LINE-1- and HERV-mediated gene regulation in cancer cells and keratinocytes remain to be studied further.

Similar to reports in cancer tissues, we have reported global hypomethylation of IRS including LINE-1 and certain HERV types in immune cells of SLE patients [18,34,35]. We hypothesized that these IRS elements could also play roles in controlling neighboring genes. We have reported previously that the up-regulated genes in neutrophils from SLE patients were significantly associated with genes containing LINE-1 [34]. When analyzing genes containing HERVs in the current study, we observed the same association with up-regulated genes in immune cells from SLE. It should be noted that significant associations were observed with only 3 out of 11 microarray data sets, which might be due to disease heterogeneity of the diseases and more data should be further analyzed to validate this observation. However, this observation might suggest a difference in the pathogenesis of how LINE-1 and HERVs affect gene regulation in cancer and SLE. A likely hypothesis is that aberrant LINE-1 or HERV regulation in SLE leads to up-regulation of these retroelement-like transcripts, thereby stimulating immune receptors in the cells and results in the activation of immune genes, including the well-known interferon responsive genes [36–38]. The role of HERV in regulating a gene in SLE was demonstrated by the finding that an alternative transcript of CD5 was regulated by a neighboring HERV-E in SLE B cells [31]. It is possible that these LTRs could function as promoters, enhancers, or cause alternative splicing. Furthermore, TEs also occur in more than two thirds of mature long noncoding RNAs (lncRNA) transcripts and account for a substantial portion of total lncRNA sequence (~30% in human). lncRNAs may be used for various tasks, including post-transcriptional regulation, organization of protein complexes, cell-cell signaling and allosteric regulation of proteins [39–42]. The exonic TEs were proposed to act as RNA domains that are essential for lncRNA function called Repeat Insertion Domains of LncRNAs (RIDLs) [43]. Our analysis also showed that HERVs were associated with up-regulated genes in B cells with EBV infection similar to the HERV association in SLE. This observation is interesting because the EBV has been implicated as a major risk factor for SLE. Interestingly, we

found no significant associations between HERVs and immune cells from other immune-mediated diseases that we analyzed, including asthma, Graves' disease, and rheumatoid arthritis (except for macrophages). Our results indicate that the mechanism in SLE that involved HERVs works mainly in the immune cells and have some specificity with certain HERVs.

The results of individual LTR analysis showed that not all the LTRs from the same superfamily showed significant associations. Our review regarding the role of specific LTR that control certain genes is summarized in S5 Table. Using EnHERV, we would like to give some examples in SLE as the following. The associations were detected between intragenic ERV1, ERVL, and ERVL-MaLR superfamilies with up-regulated genes in SLE T cells and in PBMCs with RNP+ conditions; however, no such associations were found for the ERVK superfamily (S2 Table). This finding correlated with the results of our previous study that hypomethylation of HERV-E, but not HERV-K, was detected in SLE CD4+ T cells [18]. This specific hypomethylation was also associated with up-regulation of HERV-E transcripts in CD4+ T cells [44]. Moreover, our results showed a particularly strong association of SLE with RNP+. This is consistent with the reported sequence homology between HRES-1 and the 70-kDa *gag*-related region of sn-RNP, which supports the suggestion that a possible mechanism in etiopathogenesis of SLE is the induction of a cross-reaction between the two proteins by autoantibodies. EnHERV could help screen for the specific LTR pattern and type that is involved in a disease of interest so that the mechanisms and functions can be further studied.

Materials and methods

To build EnHERV we used genome data from the UCSC Table Browser [45] as the core information.

The UCSC genome annotation database for the February 2009 assembly of the human genome (hg19, GRCh37 Genome Reference Consortium Human Reference 37 (GCA_000001405.1)) was used in the analysis. Three UCSC tables were used in our analysis. 1) The human repeat annotations, RepeatMasker version open-3.2.7 (rmsk table listed in RepeatMasker track/ Repeats group) containing 5,298,130 repeat records (last updated: 2009-04-24). HERV names in EnHERV were based mainly from the rmsk table. 2) The human gene annotations (knownGene table), and 3) The cross-reference IDs (kgXref table), were used for mapping UCSC gene to gene symbol. The knowGene and kgXref tables contain 82,960 UCSC IDs (last updated: 2013-06-14). Both tables were listed in UCSC Genes track/Genes and Gene Predictions group. Gene information included sequences of the 24 main chromosomes (chromosome 1–22, X, and Y), 59 unplaced contigs, and nine haplotype chromosomes.

Five HERV classes were defined based on the Repbase classification system [46], namely i) class 1 superfamily (ERV1), which contains the HERVs related to gamma retroviruses such as murine leukemia virus (MLV) and baboon endogenous virus (BaEV); ii) class 2 superfamily (ERVK), which contains beta retroviruses including mouse mammary tumor virus (MMTV); iii) class 3 superfamily (ERVL), which is distantly related to spuma retroviruses; iv) the mammalian apparent LTR-retrotransposons (ERVL-MaLR), which is considered as an additional class [47,48]; and v) a group of unclassified fragments, which contain other HERV-like sequences. The full list of these HERVs is shown in S4 Table.

The computational identification and annotation of the HERVs in the reference genome was generally incomplete and the HERV element is usually annotated as a separated fragment. Therefore, the HERV defragmentation process was done by REannotate for joining HERV fragments which belonged to the same HERV element into a single element. The REannotate defragmentation program is based on distance, orientation between fragments, and membership of the same HERV family. The distances in base pairs that we tested were 10, 20, 50, 100,

200, 300, 400, 500, 1 k, 2 k, 5 k, 10 k, 20 k, 40 k, and 50 k. The numbers of defragmented elements, non-defragmented elements, and total elements were determined for each of the distance parameters tested. The equivalent REannotate name list and other information were retrieved from Repbase Update [46]. To manipulate the redundant family names in the REannotate output, the information of each family name was extracted from the comment lines described in the Repbase. Moreover, the information of previous annotated name of LTRs and their internal sequence was also available in Repbase. It was used as reference information in the manipulation process. After defragmentation, the defragmented HERV elements were mapped to every annotated gene isoform by their location in the human genome.

EnHERV was constructed as a web database tool for easy access by users. EnHERV uses PHP to generate dynamic HTML, CSS and Javascript. A MySQL was implemented for recording HERV neighboring gene information. The enrichment function was implement by Python programming language. Two major functions were implemented in EnHERV which are search function and enrichment analysis function. The search function allows users to connect to the pre-built HERV profile database, as described above. Two search options are provided: search by gene(s) and search by HERV characteristics. Seven HERV characteristics can be input: 1) HERV superfamily, 2) HERV family, 3) HERV name, 4) HERV orientation, 5) HERV distance from their neighboring gene, 6) HERV location in gene, and 7) HERV completeness type. The enrichment analysis function implements Fisher's exact test to test the non-random associations between user-defined gene or preset gene lists and genes that contain specific HERV characteristics. Fisher's exact p-value was calculated to examine the significance of the association from contingency table between genes in the giving list with and without specific HERV characteristics in defined conditions and human UCSC genes with and without the same specific HERV characteristics.

HERVs have mainly been reported to be involved in cancer and autoimmune disease, ten different cancer and autoimmune experiments were retrieved from the gene expression omnibus (GEO) [49,50] and built in to EnHERV as sample gene lists. We hypothesized that HERVs could affect the expression of their neighboring genes by either up- or down-regulation, and that the association may be in a specific direction and location. We performed an enrichment analysis in EnHERV to detect the associations between specific HERV properties and various disease conditions. We retrieved 49 GEO accessions from the NCBI GEO database (https://www.ncbi.nlm.nih.gov/geo/) and classified the gene expressions into 49 conditions, including autoimmune and other disease conditions, as shown in S1 Table. Differentially expressed genes were identified using the GEO2R function (http://www.ncbi.nlm.nih.gov/geo/geo2r/). Genes were considered as differentially expressed for P-values with Benjamini & Hochberg adjustment ≤ 0.05 and fold-changes >1-fold.

We tested the association between various disease conditions and four HERV superfamilies. We also used 25 individual HERVs that represent each family for the enrichment analysis (Table 1). Most of the HERVs were found to be expressed HERV elements in previous reports [51–56]. HERVs and disease conditions were considered as associated events when Fisher's exact p-value was <0.001 and the odds ratio was >1.

Conclusions

Many reports have supported the idea that epigenetics plays an import role in disease pathogenesis [57–61]. Previous publications have shown that TEs can alter the expression of their nearby genes as a result of methylation imbalances. Therefore, we investigated the association between HERVs and gene expression under various disease conditions, especially in SLE. First, we developed EnHERV as a HERV database and enrichment tool using repeat and

human genome information from Repbase and the UCSC Table Browser. EnHERV is available at http://sysbio.chula.ac.th/enherv/. EnHERV provides searches by gene names or HERV characteristics and also allows users to perform enrichment analysis between gene lists of user interest and specific selected HERV characteristics. Thousands of enrichment analyses were performed in this study. The results suggested that certain disease conditions were associated with specific LTR types. The EnHERV database and built-in functions will help in further understanding the pathogenesis of not only SLE, but also other diseases where HERVs might be involved in their pathogenesis.

Supporting information

S1 Fig. The highly HERVs diseases signification association. With the P-value < 0.001 and odd ratio > 1 cutoff criteria, the ERV1, ERVL, and ERVL-MaLR superfamilies but not with the ERVK superfamily show the different pattern various disease conditions.
(TIF)

S1 File. Data availability statement.
(DOCX)

S1 Table. Data retrieved from the NCBI GEO database (https://www.ncbi.nlm.nih.gov/geo/) and disease conditions into which the differentially expressed genes were used in demonstration study.
(DOCX)

S2 Table. Association analysis results between various gene expression conditions and the four HERV superfamilies. Significant data with odds ratio >1 and Fisher's p-value <0.001 are indicated in green letters for up-regulated genes and red letters for down-regulated genes.
(XLSX)

S3 Table. Association analysis results between various gene expression conditions and individual HERVs that represent each superfamily. Significant data with odds ratio >1 and Fisher's p-value <0.001 are indicated in in green letters for up-regulated genes and red letters for down-regulated genes.
(XLSX)

S4 Table. Full list of HERVs in the EnHERV database.
(DOCX)

S5 Table. Evidences of LTR involved in gene expression.
(DOCX)

Author Contributions

Conceptualization: SP NH AM.

Formal analysis: PT SP.

Funding acquisition: PT NH.

Investigation: PT NH.

Methodology: PT SP AM.

Project administration: NH AM.

Software: PT WM.

Supervision: SP NH.

Visualization: PT WM.

Writing – original draft: PT NH.

Writing – review & editing: PT SP YA AM NH.

References

1. Griffiths DJ (2001) Endogenous retroviruses in the human genome sequence. Genome Biol 2: REVIEWS1017. PMID: 11423012

2. Feschotte C, Gilbert C (2012) Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet 13: 283–296. https://doi.org/10.1038/nrg3199 PMID: 22421730

3. Chuong EB, Elde NC, Feschotte C (2017) Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet 18: 71–86. https://doi.org/10.1038/nrg.2016.139 PMID: 27867194

4. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860–921. https://doi.org/10.1038/35057062 PMID: 11237011

5. Cohen CJ, Lock WM, Mager DL (2009) Endogenous retroviral LTRs as promoters for human genes: a critical assessment. Gene 448: 105–114. https://doi.org/10.1016/j.gene.2009.06.020 PMID: 19577618

6. Griffiths D (2001) Endogenous retroviruses in the human genome sequence. Genome Biology 2: reviews1017.1011—reviews1017.1015.

7. Haase K, Mosch A, Frishman D (2015) Differential expression analysis of human endogenous retroviruses based on ENCODE RNA-seq data. BMC Med Genomics 8: 71. https://doi.org/10.1186/s12920-015-0146-5 PMID: 26530187

8. Lu X, Sachs F, Ramsay L, Jacques PE, Goke J, et al. (2014) The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat Struct Mol Biol 21: 423–425. https://doi.org/10.1038/nsmb.2799 PMID: 24681886

9. Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, et al. (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet 42: 631–634. https://doi.org/10.1038/ng.600 PMID: 20526341

10. Perron H, Lazarini F, Ruprecht K, Pechoux-Longin C, Seilhean D, et al. (2005) Human endogenous retrovirus (HERV)-W ENV and GAG proteins: physiological expression in human brain and pathophysiological modulation in multiple sclerosis lesions. J Neurovirol 11: 23–33.

11. Grandi N, Cadeddu M, Blomberg J, Tramontano E (2016) Contribution of type W human endogenous retroviruses to the human genome: characterization of HERV-W proviral insertions and processed pseudogenes. Retrovirology 13: 67. https://doi.org/10.1186/s12977-016-0301-x PMID: 27613107

12. Schmitt K, Heyne K, Roemer K, Meese E, Mayer J (2015) HERV-K(HML-2) rec and np9 transcripts not restricted to disease but present in many normal human tissues. Mob DNA 6: 4. https://doi.org/10.1186/s13100-015-0035-7 PMID: 25750667

13. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, et al. (2009) The regulated retrotransposon transcriptome of mammalian cells. Nat Genet 41: 563–571. https://doi.org/10.1038/ng.368 PMID: 19377475

14. Suntsova M, Garazha A, Ivanova A, Kaminsky D, Zhavoronkov A, et al. (2015) Molecular functions of human endogenous retroviruses in health and disease. Cell Mol Life Sci 72: 3653–3675. https://doi.org/10.1007/s00018-015-1947-6 PMID: 26082181

15. Bergman CM, Quesneville H (2007) Discovering and detecting transposable elements in genome sequences. Brief Bioinform 8: 382–392. https://doi.org/10.1093/bib/bbm048 PMID: 17932080

16. Okada M, Ogasawara H, Kaneko H, Hishikawa T, Sekigawa I, et al. (2002) Role of DNA methylation in transcription of human endogenous retrovirus in the pathogenesis of systemic lupus erythematosus. J Rheumatol 29: 1678–1682. PMID: 12180729

17. Piotrowski PC, Duriagin S, Jagodzinski PP (2005) Expression of human endogenous retrovirus clone 4–1 may correlate with blood plasma concentration of anti-U1 RNP and anti-Sm nuclear antibodies. Clin Rheumatol 24: 620–624. https://doi.org/10.1007/s10067-005-1123-8 PMID: 16012778

18. Nakkuntod J, Sukkapan P, Avihingsanon Y, Mutirangura A, Hirankarn N (2013) DNA methylation of human endogenous retrovirus in systemic lupus erythematosus. J Hum Genet 58: 241–249. https://doi.org/10.1038/jhg.2013.6 PMID: 23466822

19. Paces J, Pavlicek A, Paces V (2002) HERVd: database of human endogenous retroviruses. Nucleic Acids Research 30: 205–206. PMID: 11752294

20. Levy A, Sela N, Ast G (2008) TranspoGene and microTranspoGene: transposed elements influence on the transcriptome of seven vertebrates and invertebrates. Nucleic Acids Research 36: D47–D52. https://doi.org/10.1093/nar/gkm949 PMID: 17986453

21. Pereira V (2008) Automated paleontology of repetitive DNA with REANNOTATE. BMC Genomics 9: 614. https://doi.org/10.1186/1471-2164-9-614 PMID: 19094224

22. Giordano J, Ge Y, Gelfand Y, Abrusán G, Benson G, et al. (2007) Evolutionary History of Mammalian Transposons Determined by Genome-Wide Defragmentation. PLoS Comput Biol 3: e137. https://doi.org/10.1371/journal.pcbi.0030137 PMID: 17630829

23. Pereira V (2004) Insertion bias and purifying selection of retrotransposons in the Arabidopsis thaliana genome. Genome Biology 5: R79. https://doi.org/10.1186/gb-2004-5-10-r79 PMID: 15461797

24. Goodier JL (2016) Restricting retrotransposons: a review. Mob DNA 7: 16. https://doi.org/10.1186/s13100-016-0070-z PMID: 27525044

25. Taruscio D, Floridia G, Zoraqi GK, Mantovani A, Falbo V (2002) Organization and integration sites in the human genome of endogenous retroviral sequences belonging to HERV-E family. Mamm Genome 13: 216–222. https://doi.org/10.1007/s00335-001-2118-7 PMID: 11956766

26. Katoh I, Mirova A, Kurata S, Murakami Y, Horikawa K, et al. (2011) Activation of the long terminal repeat of human endogenous retrovirus K by melanoma-specific transcription factor MITF-M. Neoplasia 13: 1081–1092. PMID: 22131883

27. Pi W, Zhu X, Wu M, Wang Y, Fulzele S, et al. (2010) Long-range function of an intergenic retrotransposon. Proc Natl Acad Sci U S A 107: 12992–12997. https://doi.org/10.1073/pnas.1004139107 PMID: 20615953

28. Absher DM, Li X, Waite LL, Gibson A, Roberts K, et al. (2013) Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to CD4+ T-cell populations. PLoS Genet 9: e1003678. https://doi.org/10.1371/journal.pgen.1003678 PMID: 23950730

29. Feuchter-Murthy AE, Freeman JD, Mager DL (1993) Splicing of a human endogenous retrovirus to a novel phospholipase A2 related gene. Nucleic Acids Res 21: 135–143. PMID: 8382789

30. Lock FE, Rebollo R, Miceli-Royer K, Gagnier L, Kuah S, et al. (2014) Distinct isoform of FABP7 revealed by screening for retroelement-activated genes in diffuse large B-cell lymphoma. Proc Natl Acad Sci U S A 111: E3534–3543. https://doi.org/10.1073/pnas.1405507111 PMID: 25114248

31. Renaudineau Y, Vallet S, Le Dantec C, Hillion S, Saraux A, et al. (2005) Characterization of the human CD5 endogenous retrovirus-E in B lymphocytes. Genes Immun 6: 663–671. https://doi.org/10.1038/sj.gene.6364253 PMID: 16107871

32. Aporntewan C, Phokaew C, Piriyapongsa J, Ngamphiw C, Ittiwut C, et al. (2011) Hypomethylation of intragenic LINE-1 represses transcription in cancer cells through AGO2. PLoS One 6: e17934. https://doi.org/10.1371/journal.pone.0017934 PMID: 21423624

33. Yooyongsatit S, Ruchusatsawat K, Noppakun N, Hirankarn N, Mutirangura A, et al. (2015) Patterns and functional roles of LINE-1 and Alu methylation in the keratinocyte from patients with psoriasis vulgaris. J Hum Genet 60: 349–355. https://doi.org/10.1038/jhg.2015.33 PMID: 25833468

34. Sukapan P, Promnarate P, Avihingsanon Y, Mutirangura A, Hirankarn N (2014) Types of DNA methylation status of the interspersed repetitive sequences for LINE-1, Alu, HERV-E and HERV-K in the neutrophils from systemic lupus erythematosus patients and healthy controls. J Hum Genet 59: 178–188. https://doi.org/10.1038/jhg.2013.140 PMID: 24430577

35. Nakkuntod J, Avihingsanon Y, Mutirangura A, Hirankarn N (2011) Hypomethylation of LINE-1 but not Alu in lymphocyte subsets of systemic lupus erythematosus patients. Clin Chim Acta 412: 1457–1461. https://doi.org/10.1016/j.cca.2011.04.002 PMID: 21496453

36. Nogueira MA, Gavioli CF, Pereira NZ, de Carvalho GC, Domingues R, et al. (2015) Human endogenous retrovirus expression is inversely related with the up-regulation of interferon-inducible genes in the skin of patients with lichen planus. Arch Dermatol Res 307: 259–264. https://doi.org/10.1007/s00403-014-1524-0 PMID: 25384438

37. Mameli G, Cossu D, Cocco E, Frau J, Marrosu MG, et al. (2015) Epitopes of HERV-Wenv induce antigen-specific humoral immunity in multiple sclerosis patients. J Neuroimmunol 280: 66–68. https://doi.org/10.1016/j.jneuroim.2015.03.003 PMID: 25773158

38. Mavragani CP, Sagalovskiy I, Guo Q, Nezos A, Kapsogeorgou EK, et al. (2016) Expression of Long Interspersed Nuclear Element 1 Retroelements and Induction of Type I Interferon in Patients With Systemic Autoimmune Disease. Arthritis Rheumatol 68: 2686–2696. https://doi.org/10.1002/art.39795 PMID: 27338297

39. Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, et al. (2013) Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet 9: e1003470. https://doi.org/10.1371/journal.pgen.1003470 PMID: 23637635

40. Geisler S, Coller J (2013) RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat Rev Mol Cell Biol 14: 699–712. https://doi.org/10.1038/nrm3679 PMID: 24105322

41. Babaian A, Mager DL (2016) Endogenous retroviral promoter exaptation in human cancer. Mob DNA 7: 24. https://doi.org/10.1186/s13100-016-0080-x PMID: 27980689

42. Kelley D, Rinn J (2012) Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol 13: R107. https://doi.org/10.1186/gb-2012-13-11-r107 PMID: 23181609

43. Johnson R, Guigo R (2014) The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. RNA 20: 959–976. https://doi.org/10.1261/rna.044560.114 PMID: 24850885

44. Wu Z, Mei X, Zhao D, Sun Y, Song J, et al. (2015) DNA methylation modulates HERV-E expression in CD4+ T cells from systemic lupus erythematosus patients. J Dermatol Sci 77: 110–116. https://doi.org/10.1016/j.jdermsci.2014.12.004 PMID: 25595738

45. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, et al. (2004) The UCSC Table Browser data retrieval tool. Nucleic Acids Res 32: D493–496. https://doi.org/10.1093/nar/gkh103 PMID: 14681465

46. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, et al. (2005) Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110: 462–467. https://doi.org/10.1159/000084979 PMID: 16093699

47. Smit AF (1993) Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic Acids Res 21: 1863–1872. PMID: 8388099

48. Smit A, Hubley, R & Green, P. (2013–2015) RepeatMasker Open-4.0.

49. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30: 207–210. PMID: 11752295

50. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. (2013) NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res 41: D991–995. https://doi.org/10.1093/nar/gks1193 PMID: 23193258

51. Wang J, Singh M, Sun C, Besser D, Prigione A, et al. (2016) Isolation and cultivation of naive-like human pluripotent stem cells based on HERVH expression. Nat Protoc 11: 327–346. https://doi.org/10.1038/nprot.2016.016 PMID: 26797457

52. Lamprecht B, Walter K, Kreher S, Kumar R, Hummel M, et al. (2010) Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nat Med 16: 571–579, 571p following 579. https://doi.org/10.1038/nm.2129 PMID: 20436485

53. Kronung SK, Beyer U, Chiaramonte ML, Dolfini D, Mantovani R, et al. (2016) LTR12 promoter activation in a broad range of human tumor cells by HDAC inhibition. Oncotarget 7: 33484–33497. https://doi.org/10.18632/oncotarget.9255 PMID: 27172897

54. Le Dantec C, Vallet S, Brooks WH, Renaudineau Y (2015) Human endogenous retrovirus group E and its involvement in diseases. Viruses 7: 1238–1257. https://doi.org/10.3390/v7031238 PMID: 25785516

55. Watanabe Y, Tenzen T, Nagasaka Y, Inoko H, Ikemura T (2000) Replication timing of the human X-inactivation center (XIC) region: correlation with chromosome bands. Gene 252: 163–172. PMID: 10903448

56. Glinsky GV (2015) Transposable Elements and DNA Methylation Create in Embryonic Stem Cells Human-Specific Regulatory Sequences Associated with Distal Enhancers and Noncoding RNAs. Genome Biol Evol 7: 1432–1454. https://doi.org/10.1093/gbe/evv081 PMID: 25956794

57. Miao CG, Yang JT, Yang YY, Du CL, Huang C, et al. (2014) Critical role of DNA methylation in the pathogenesis of systemic lupus erythematosus: new advances and future challenges. Lupus 23: 730–742. https://doi.org/10.1177/0961203314527365 PMID: 24644011

58. Long H, Yin H, Wang L, Gershwin ME, Lu Q (2016) The critical role of epigenetics in systemic lupus erythematosus and autoimmunity. J Autoimmun.

59. Chen SH, Lv QL, Hu L, Peng MJ, Wang GH, et al. (2016) DNA methylation alterations in the pathogenesis of lupus. Clin Exp Immunol.

60. Sharma S, Kelly TK, Jones PA (2010) Epigenetics in cancer. Carcinogenesis 31: 27–36. https://doi.org/10.1093/carcin/bgp220 PMID: 19752007

61. Tsai HC, Baylin SB (2011) Cancer epigenetics: linking basic biology to clinical medicine. Cell Res 21: 502–517. https://doi.org/10.1038/cr.2011.24 PMID: 21321605