ICELL JOURNAL | RESEARCH ARTICLE

Vol.2 No.1 (1)

Annotation and target analysis of human endogenous retroviruses

DOI:https://dx.doi.org/10.71373/PVSC2027
Submitted 1 March 2025
Accepted 1 May 2025
Published 14 May2025

HuilinWei1,BoyingLiang2,ChunlanJiang3,TingweiLyu1,WenLi1,4,QiuxiaMeng5,Tengyue Yan6,XuanyuPan7,YuxiaoHe1,JianpingZhou1,YanlingHu1,

Background:Endogenous retroviruses (ERVs) are important regulatory elements in the human genome. They are involved in the regulation of host gene expression and disease progression through long terminal repeats (LTR) and coding domains (gag, pol, env).Methods:Based on the implicit Markov model, this study integrated LTRharvest and LTRdigest software, combined with 55 ERV-related protein domain databases, systematically annotated ERVs elements in the human genome (GRCh38.p14), and analyzed their potential targets and functions by using STRING, GO and KEGG.Results:A total of 47,666 HERVs candidate sequences (11.05% of the genome) were identified in this study, of which 605 were complete structures, mainly concentrated in chromosomes 1 and 3. It was found that env accounted for the least among the three protein structures. Potential target genes in the upstream and downstream 20kb range of LTR were screened, and core targets such as histone family genes H4C6 and H2BC12 and immune-related genes TLR2 and CCR5 were found to be involved in disease regulation through chromatin remodeling or immune pathways. Enrichment results were significantly associated with nucleosome assembly, innate immune response, and cancer-related pathways (herpes simplex virus infection, systemic lupus erythematosus).Conclusion:This study constructed a comprehensive HERVs annotated database, revealing the potential regulatory ability of LTR, providing a theoretical basis for the application of HERVs in cancer, autoimmune diseases and evolutionary research, and laying a foundation for the development of targeted therapy strategies.

1 Introduction

Endogenous retroviruses(ERVs),also known as LTR transposition elements,belong to a class of retrotransposition elements,which are divided into LTR and non-LTR transposition elements based on whether they have long terminal repeats(LTR) at both ends.ERV was first discovered in the 1960s[1].By infecting somatic cells with exogenous retroviruses,ERV gradually integrates into the vertebrate genome and continuously evolves with the host in the way of Mendelian inheritance.The proportion of endogenous retrovirus sequences in the human genome has reached 8%[2,3].ERV interacts with other factors through transcription and soon,showing a variety of biological functions.As a potential biomarker,ERV can change the course of human diseases to a certain extent,and has gradually become a focus of current research.

The structure of a complete HERV consists of long terminal repeats(LTR) on both sides and an openreadingframe(ORF).In the absence of mutations,HERV contains three functional genes(gag,pol,env).gag gene is responsible for encoding Gag structural protein and promoting the assembly of virus particles.pol gene encodes reverse transcriptase,RNase H and integrase.The env gene encodes the membrane protein Env,which mediates the binding of cell receptors to membranes.But in the case of without the mutation,according to the international classification of virus committee(International Committee on the Taxonomyof VirusesICTV)classification,ERV can be roughly divided into three categories:The sequence of class I was similar to that of gamma retrovirus and Epsilon retrovirus.Class II is similar to α-retroviruses,β-retroviruses,delta retroviruses and lentiviruses.Class III is similar to foam retroviruses and ERV-L[4,5].

Overtime,ERV has become a stable sequence in the genome and is involved in gene regulation by interacting with different signals[6].For example,HERV can affect cancer,immune deficiency diseases and neurodegenerative diseases[7,8].In 2000,it was found that the HERV-Wfamily can produce the envelope protein Syncytin-1,which is a key molecule that drives trophoblast fusion to produce trophoblast,and can regulate embryo implantation and placental trophoblastd evelopment,playing an important role in the process of embryo genesis[9,10].The membrane protein of HERV-K has also been shown to play a role in neuronal degeneration in amyotrophic lateral sclerosis(ASL)[11,12].Neuropsychological diseases such as schizophrenia are close lyrelated to the abnormal expression of HERV-W[13].It has been reported that different transcripts from HERV-H,HERV-K,HERV-R and other families are abnormally expressed in human cancer cells[14].The HERV-K family is more active,with abnormal transcription and transla

AuthorAffiliations:1InstituteofLifeSciences,GuangxiMedicalUniversity,Nanning,Guangxi,China2DepartmentofImmunology,SchoolofBasicMedicalSciences,GuangxiMedicalUniversity,Nanning,Guangxi,China3GuangxiHenbioTechnologyCo.,Ltd,Nanning,Guangxi,China.4DepartmentofBiochemistryandMolecularBiology,SchoolofBasicMedicine,GuangxiMedicalUniversity,Nanning,Guangxi,China.5GuangxiMedicalUniversitySchoolofInformationandManagement,Nanning,Guangxi,China.6CollaborativeInnovationCentreofRegenerativeMedicineandMedicalBioresourceDevelopmentandApplicationCo-constructedbytheProvinceandMinistry,GuangxiMedicalUniversity,Nanning,Guangxi,China.7SchoolofBasicMedicalSciences,GuangxiMedicalUniversity,Nanning,Guangxi,China.
*Correspondingauthor:YanlingHu,Email:ylhupost@163.com

tion detected in various cancers,such as malignant tumors such as melanoma[15],germ cell carcinoma,and ovarian cancer[16].However,in the current study,there is no direct evidence to prove the direct causal relationship between HERV and cancer.However,HERV epigenome modification may be used as a biomarker for the early diagnosis of cancer,so HERV still has strong potential as a biomarker[17].

At present,HERVs have lost the ability to reverse transpose and insert mutations,but can regulate host gene expression through its mRNA and protein products or gene regulatory regions derived from LTR[18,19].Many intact ERVs become targets of transcriptional silencing through modification or mutation of their LTR,playing an important regulatory role in development by regulating the transcripto me.At the same time,LTRs of HERVs can be used as an alternative promoter to drive the expression of oncogenes,thus affecting the occurrence and development of cancer[20].In the current study,the LTR of HERVs is critical for viral replication and integration,typically contains multiple regulatory elements,regulates the expression of nearby genes,participates in host evolution,plays a role in disease,and has potential immunomodulatory effects.LTR in HERVs drives specific gene expression during mammalian oocyte and fertilized egg development by acting as a surrogate promoter and exon[21].Ltr-driven transcription is also affected by various factors, such as epigenetic reprogramming,etc.Such regulatory responses become important expression signals in the evolution of cancer cells,thus causing the occurrence of various diseases[22].

The study of ERV helps to understand the diversity of development and morphological evolution,and it also has a large degree of application in the fields of cancer and autoimmune disease.At present,the integrity of HERV and its components is damaged due to high variation,and the identification and annotation of HERV and its components has been a major difficulty.This study provides relatively complete annotation information of HERV elements,constructs a rich and complete HERV characteristic database, and analyzes the potential target genes upstream and down stream of HERVsLTR,providing important clues for the study of HERVs regulating near by genes and influencing biological traits in the human genome.It provides a strong theoretical basis in pathogens,species evolution,human cancer and other related fields.

2 Materials and methods

2.1 Human whole genome data file acquisition

By GENCODE database (https://www.gencodegenes.org/), the human genome files needed to download, get "gencode_GRCh38p14genome. Fa" the whole genome sequence of the file, the file size is about 3.1 Gb.

Human genome LTR sequence and HERVs annotation

LTRharvest software was used to search for LTR at both ends of the human genome, and the criteria for determining LTR at both ends were as follows: (1) Candidate LTR sequence sum The similarity threshold of the reference sequence is 80%. (2) The LTR length at both ends of the LTR candidate sequence ranges from 1kb to 15kb. (3) Precise search for 4 nucleotides in the motif of LTR initiation and ending sites.

LTRdigest software was used to annotate the features of pairwise LTR and determine the location, direction, distance and sequence composition of the LTR sequence and its internal coding protein domain. With "retro" as the keyword in the Pfam database(http://pfam-legacy.xfam.org/), 55 protein entries associated with ERVs were identified in combination with other published literatures (Table 1). It is used to detect the presence of protein domains encoded by ERV-related genes such as gag, pol, env, and so on. The downloaded protein entries are converted into HMMER2 format, and combined with human tRNA file information, a library is constructed together as the input of LTRdigest software for domain prediction.

Table1 ERVs related protein articles

Accession	name	description
PF00075	RNase_H	Ribonuclease H domain
PF00077	RVP	Retroviral aspartyl protease
PF00078	RVT_1	Reverse transcriptase(RNA-dependent DNA polymerase)
PF00098	zf-CCHC	Zinc knuckle domain
PF00429	TLV_coat	ENV polyprotein (coat polyprotein)
PF00516	GP120	Envelope glycoprotein gp120
PF00517	GP41	Retroviral envelope protein
PF00540	Gag_p17	gag gene protein p17(matrix protein)
PF00552	IN_DBD_C	Integrase DNA binding domain

PF00559	Vif	Retroviral Vif(Viral infectivity)protein
PF00607	Gag_p24	gag protein p24 N-terminal domain
PF00665	rve	Integrase core domain
PF00692	dUTPase	dUTPase domain
PF01021	TYA	Ty transposon capsid protein
PF01140	Gag_MA	Matrix protein (MA),p15
PF01141	Gag_p12	Gag polyprotein, inner coat protein pl2
PF02022	Integrase_Zn	Integrase Zinc binding domain
PF02093	Gag p30	Gag P30 core shell protein
PF02337	Gag p10	Retroviral GAG p10 protein
PF02813	Retro_M	Retroviral matrix protein
PF02994	Transposase 22	L1 transposable element RBD-like domain
PF03078	ATHILA	ATHILA ORF-1 family
PF03276	Gag spuma	Spumavirus gag protein
PF03408	Foamy virus_ENV	Foamy virus envelope protein
PF03708	Avian_gp85	Avian retrovirus envelope protein, gp85
PF03732	Retrotrans gag	Retrotransposon gag protein
PF04160	Borrelia orfX	Orf-X protein
PF04195	Transposase_28	Putative gypsy type transposon
PF05380	Peptidase_A17	Pao retrotransposon peptidase
PF06815	RVT connect	Reverse transcriptase connection domain
PF06817	RVT thumb	Reverse transcriptase thumb domain
PF07253	Gypsy	Gypsy protein,Reverse transcriptase
PF07727	RVT_2	(RNA-dependent DNA polymerase)
PF08284	RVP_2	Retroviral aspartyl protease
PF09590	Env-gp36	Env-gp36 protein(HERV/MMTV type)
PF11988	Dsll_N	Retrograde transport protein Dsl1 N terminal
PF11989	Dsll_C	Retrograde transport protein Dsl1 C terminal
PF12382	Peptidase A22	Retrotransposon peptidase
PF13456	RVT_3	Reverse transcriptase-like
PF13655	RVT_N	N-terminal domain of reverse transcriptase

PF13804	HERV-K_env_2	Retro-transcribing viruses envelope glycoprotein
PF13966	zf-RVT	zinc-binding in reverse transcriptase
PF13975	gag-asp_proteas	gag-polyprotein putative aspartyl protease
PF13976	gag_pre-integrs	GAG-pre-integrase domain
PF14223	Retrotran_gag_2	gag-polypeptide of LTR copia-type
PF14244	Retrotran_gag_3	gag-polypeptide of LTR copia-type
PF14529	Exo_endo_phos_2	Endonuclease-reverse transcriptase
PF17241	Retrotran_gag_4	Ty5 Gag N-terminal region
PF17917	RT_RNaseH	RNase H-like domain found in reverse transcriptase
PF17919	RT_RNaseH_2	RNase H-like domain found in reverse transcriptase
PF17921	Integrase_H2C2	Integrase zinc binding domain
PF17984	TERT_thumb	Telomerase reverse transcriptase thumb DNA binding domain
PF18103	SH3_11	Retroviral integrase C-terminal SH3 domain
PF19259	Ty3_capsid	Ty3 transposon capsid-like protein
PF19317	Gag_p24_C	Gag protein p24 C-terminal domain

2.2 Taxonomic evolutionary analysis of HERVs

According to the classification of retroviruses by the International Committee on Taxonomy of Viruses,we collected 25 reference sequences of endogenous retroviruses with relatively complete RT structure[23,24](Table2).All reference sequences and HERVs with complete structure jointly constructed the evolutionary tree.The comparison tool uses the MEGA7MUSCLEalgorithm[25].trimAL[26]software was used to trim the files,and the threshold of glycine ratio was set to 0.65,and the threshold of sequence similarity was set to 0.001.The Fast tree maximum likelihood method was used to construct the evolutionary tree.Use iTOL[27](Interactive Tree Of Life, https://itol.embl.de/)for beautification.

Table2 Endogenous retrovirus reference sequence

Class	species	Accession	name
ClassI	γ retrovirus	AF053745	Mus dunni endogenous virus,MDEV
ClassI	γ retrovirus	NC_001501	Moloney murine leukemia virus,MMLV
ClassI	γ retrovirus	M77194	Rat leukemia virus,RaLV
ClassI	γ retrovirus	NC_001940	feline leukemia virus,FELV
ClassI	γ retrovirus	NC_001885	gibbon ape leukemia virus,GaLV
ClassI	γ retrovirus	NC_003059	porcine endogenous retrovirus E
ClassI	γ retrovirus	NC_039228	koala retrovirus,KoRV
ClassI	γ retrovirus	U94692	Rauscher murine leukemia virus,RMLV
ClassI	ε retrovirus	NC_001867	Walleye dermal sarcoma virus,WDSV
ClassI	ε retrovirus	AF133051	Walleye cpidermal hyperplasia virus 1,WEHVI

ClassI	ε retrovirus	AF133052	Walleye cpidermal hyperplasia virus 2,WEHV2
ClassII	α retrovirus	NC_015116	Avian leukosis virus,ALV
ClassII	α retrovirus	NC_001407	Rous sarcoma virus,RSV
ClassII	β retrovirus	NC_001550	Mason-Pfizer monkey virus, MPMV
ClassII	β retrovirus	M11841	Simian retrovirus 1 (SRV-1)
ClassII	β retrovirus	NC_001503	mouse mammary tumor virus,MMTV
ClassII	β retrovirus	NC_001494	Jaagsiekte sheep retrovirus,JSRV
ClassII	δ retrovirus	NC_001414	bovine leukemia virus,BLV
ClassII	δ retrovirus	NC_001488	Human T-lymphotropic virus 2,HTLV-2
ClassII	lentivirus	NC_001413	Bovine immunodeficiency virus (BIV)
ClassII	lentivirus	NC_001802	Human immunodeficiency virus 1,HIV-1
ClassII	lentivirus	NC_001722	Human immunodeficiency virus 2,HIV-2
ClassII	lentivirus	NC_001549	Simian immunodeficiency virus (SIV)
ClassII	lentivirus	NC_001482	Feline immunodeficiency virus,FIV
ClassII	lentivirus	NC_001511	Ovine lentivirus,MVV
ClassII	lentivirus	NC_001450	Equine infectious anemia virus,EIAV
ClassIII	spumavirus	NC_039242	feline foamy virus,FeFV
ClassIII	spumavirus	Y07725	human foamy virus,HFV
ClassIII	spumavirus	GU356394	Squirrel monkey virus,SMRV
ClassIII	spumavirus	NC_002201	Equine foamy virus,EFV
ClassIII	spumavirus	NC_075434	Simian foamy virus proviral

2.3 Construction of protein-protein interaction n etwork (PPI) and screening of LTR related core ta rgets

In order to explore the potentially related proteins of HE RVsLTR and visually demonstrate the regulatory role bet ween LTR and human genes, this study extracted the gen es within the range of 20kb upstream of ERVs5 'LTR and 20kb downstream of 3' LTR with complete structure, and constructed the protein interaction network (PPI) using S TRING database. The PPI network was imported into Cyto scape software for the construction of related target net works. CytoHubba and MCODE, Cytoscape plug-ins, were used to screen potential key genes and proteins and dra w the PPI subnetwork map.

2.4 Functional enrichment of upstream and down stream LTR genes

To explore the potential targets and HERVsLTR effect rela tionship with the disease, we to LTR within the scope of upstream and downstream of the gene enrichment analy sis, using DAVID (DatabaseforAnnotation Visualizationan dIntegratedDiscovery, https://davidbioinformatics.nih.go v/) database from cell components (cellularcomponent, C C), molecular function (molecularfunction, MF), biologica l processes (biologicalprocess, BP) for gene ontology (GO) analysis, Kyoto Encyclopedia of Genes and Genomes (KE GG) functional analysis, set $P<0.05$ as the screening condi tion, select the top 10 sequencing pathways, enrichment results using R for data visualization.

3 Results

3.1 Fragmentation analysis of ERVs elements in h uman whole genome

In this study, LTR_harvest and LTR_digest tools were use d to obtain candidate ERVs sequences from the whole hu

man genome. Fragments with LTR sequences at both ends were selected as candidate ERVs, with a total of 47,666 fragments, accounting for 11.05% of the whole human genome, and an average length of 7018bp. The results are shown in Table 3.

Table3 Examples of HERVs annotation results

number	anno_method	stucture type	start	end	E-value	strands	name
###							ID=repeat_region2
seq0	LTRharvest	repeat_region	121265	134398	.	?	Parent=repeat_region2
seq0	LTRharvest	target_site_duplication	121265	121269	.	?	ID=LTR_retrotransposon2;Parent=repeat_region2;ltr_similarity=81.53;seq_number=0
seq0	LTRharvest	LTR_retrotransposon	121270	134393	.	?	Parent=LTR_retrotransposon2
seq0	LTRharvest	long_terminal_repeat	121270	121569	.	?	Parent=LTR_retrotransposon2
seq0	LTRharvest	long_terminal_repeat	134080	134393	.	?	Parent=repeat_region2
seq0	LTRharvest	target_site_duplication	134394	134398	.	?	
###							
seq0	LTRharvest	repeat_region	3801531	3806938	.	-	ID=repeat_region85
seq0	LTRharvest	target_site_duplication	3801531	3801535	.	-	Parent=repeat_region85
seq0	LTRharvest	LTR_retrotransposon	3801536	3806933	.	-	ID=LTR_retrotransposon85;Parent=repeat_region85;ltr_similarity=93.82;seq_number=0
seq0	LTRharvest	long_terminal_repeat	3801536	3801948	.	-	Parent=LTR_retrotransposon85
seq0	LTRdigest	RR_tract	3801957	3801969	.	-	Parent=LTR_retrotransposon85
seq0	LTRdigest	protein_match	3802001	3802541	3.50E-19	-	Parent=LTR_retrotransposon85;reading_frame=0;name=GP41
seq0	LTRdigest	protein_match	3803472	3803568	3.80E-06	-	Parent=LTR_retrotransposon85;reading_frame=2;name=HERV-K_env_2
seq0	LTRdigest	protein_match	3803706	3803793	1.70E-05	-	Parent=LTR_retrotransposon85;reading_frame=2;name=IN_DBD_C
seq0	LTRdigest	protein_match	3804049	3804328	1.20E-18	-	Parent=LTR_retrotransposon85;reading_frame=1;name=rve
seq0	LTRdigest	protein_match	3804552	3804879	1.20E-13	-	Parent=LTR_retrotransposon85;reading_frame=2;name=RNase_H
seq0	LTRharvest	long_terminal_repeat	3806513	3806933	.	-	Parent=LTR_retrotransposon85

seq0	LTRharvest	target_site_duplication	3806934	3806938	.	-	Parent=repeat_region85

LTR_digest, combined with ERVS-related protein libraries, was used to detect and annotate the sequence features in HERVs. Among the 55 protein libraries, a total of 23 species were matched across the human genome, including 11 pol, 8 gag and 3 env. The proportion of 25 protein domains in all HERVs is different. As shown in Figure 1, pol occupies the highest proportion in the three protein domains. Sequence features with the highest proportion in the protein domains of pol, gag and env were RVT_1, Transposase_22 and Exo_endo_phos_2, respectively, which accounted for 0.780%, 0.191% and 0.128% of all HERVs, respectively. Among all the protein structures, the average length of the 23 protein structures identified in human genome is 259bp, among which the length of pol structure ranges from 106-496bp with an average length of 247bp, and the length of gag structure ranges from 50-422bp with an average length of 229bp. env structures range in length from 328-428bp, with an average length of 386bp. Among the 47666 HERVs, 12,879 HERVs with protein domains were obtained. The degree of protein domains contained in viral sequences varies greatly in the number of copies in the whole human genome. Among them, 6953 sequences containing only the pol domain account for the majority of all HERVs with protein domains. There were 1709 HERVs containing only gag domains, 1414 HERVs containing both gag and pol domains, and 605 HERVs containing all three domains were the least.

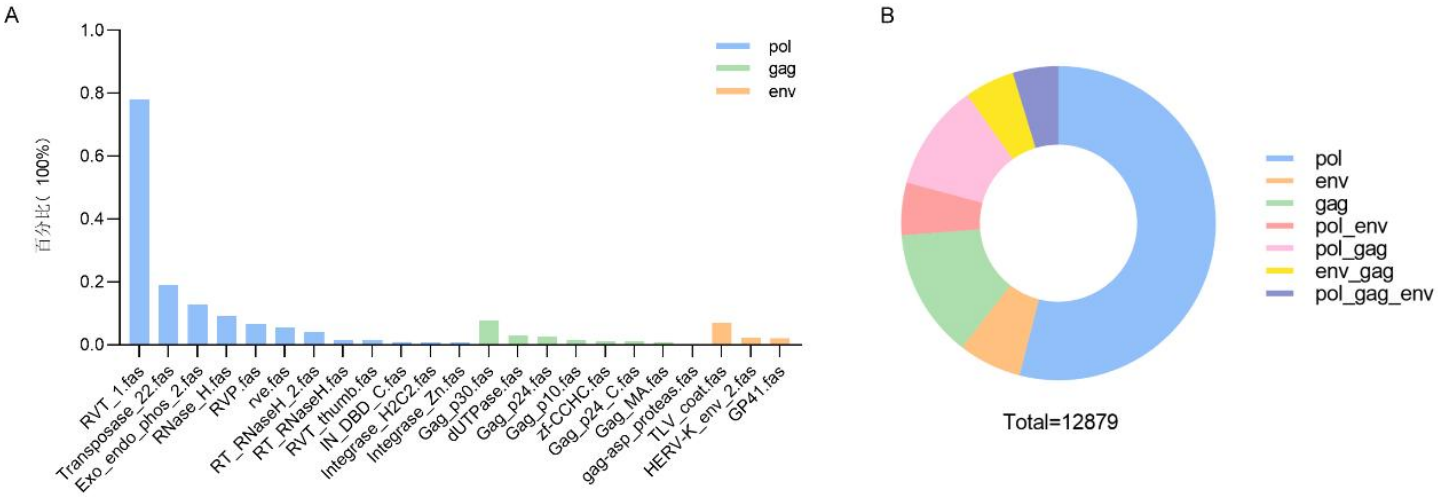

Figure1 The proportion of protein domains in HERVs.A: The proportion of retrovirus protein structures in the human genome. B: The proportion of gag, pol and env structures in all protein-containing domains HERVs.

Table4 The number of sequences containing protein domains in HERVs

gag	pol	env	env gag_pol	gag_env	pol_env	gag_pol_env
1709	6953	838	1414	661	699	605

At the same time, ERVs sequences with complete structure were screened according to the matching results of protein domains. The HERVs with LTR at both ends and containing gag, env and pol protein domains were selected as the standard for complete HERVs, and 605 complete HERVs sequences were screened. They make up only 1.17% of all HERVs and have an average length of 7154 bases. According to the analysis of the distribution of 605 complete HERVs in chromosomes, it can be seen in Figure 2 that the distribution of complete HERVs in the human genome is different, among which the number of complete HERVs in chromosomes 1 and 3 is the largest, with 58, followed by chromosome 8 (41) and chromosome 6 (40), and the number of complete HERVs in chromosomes 21 and 22 is the least. Only three.

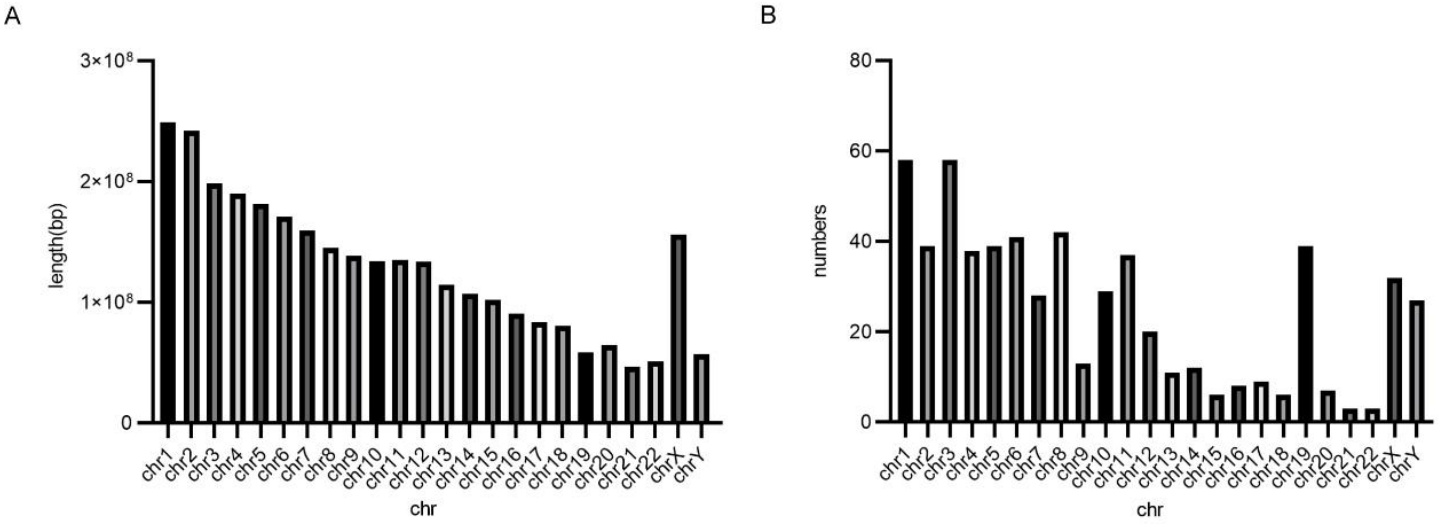

Figure 2 Length of chromosomes and distribution of intact HERVs in chromosomes.A: The length and size of human chromosomes. B: Distribution of structurally intact HERVs in human chromosomes.

3.2Evolutionary analysis of complete HERVs

We applied MEGA7 Neighbor-Joining method and FastTree[28] to construct an evolutionary tree to construct a phylogenetic tree by applying the 25 endogenous retrovirus reference sequences. As shown in Figure 3, the three classes and seven ERVs virus species including joining all had high homology. Red is ClassI, green is ClassII, and blue is ClassIII.

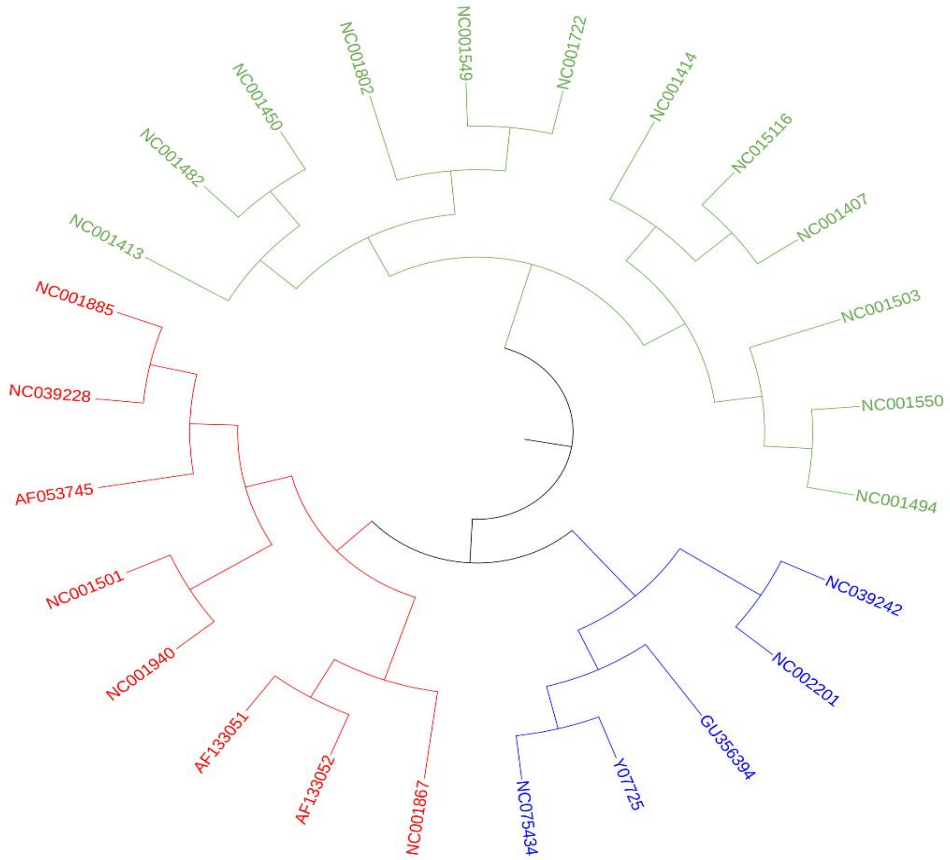

Figure 3. Phylogenetic relationships of 25 reference sequences of endogenous retroviruses.

Endogenous retroviruses are sequences of repeated sequences, and the same endogenous retrovirus fragments may occur simultaneously in different chromosomes. In order to make sure that each sequence identified was independent and specific, we verified the similarity of 47666 HERVs, using blast software to screen out sequences with similarity > 94% and coverage of more than 80% of their own sequences as non-specific sequences. After removing these nonspecific sequences from 605 fully structured HERVs, consensus was reached on 536 independent sequences.

In order to explore the classification of complete HERVs, FastTree was used to construct an evolutionary tree by combining 536 complete HERVs sequences and reference sequences. Among the 536 sequences, only 335 sequences were sufficiently conserved to construct a phylogenetic tree. The results were shown in Figure 4, and the red marks were reference genome sequences. We took Bootstrap value>70% as the basis for classification of HERVs subgroups in human genome with complete structure. Among 335 HERVs with complete structure, 13 belonged to ClassI, accounting for 3.88%, and 152 belonged to ClassI, accounting for 45.37%. 18 belong to Class ClassIII, accounting for 5.37%. However, 45.67% of HERVs are not well classified.

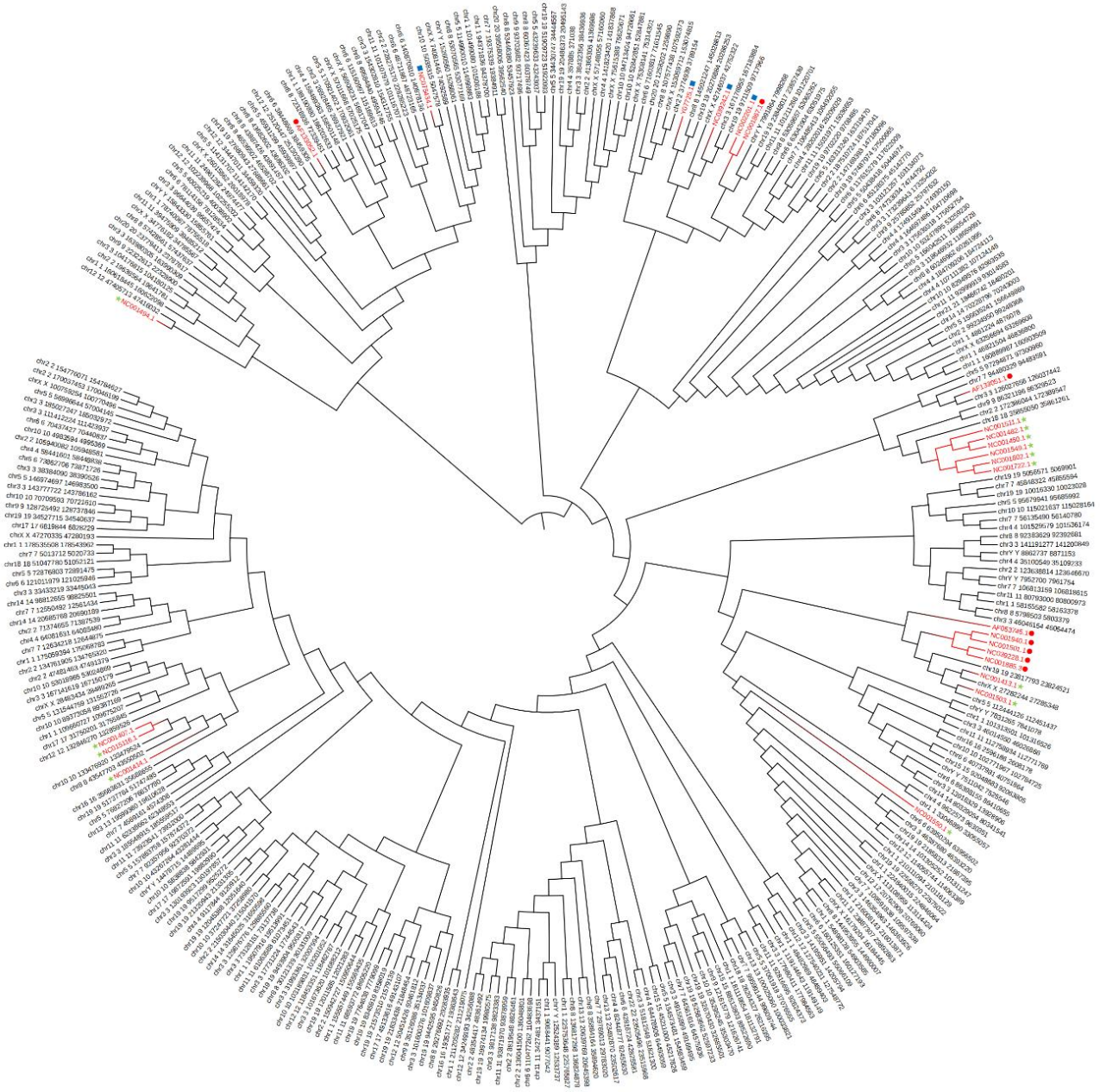

Figure 4 HERVs kinship diagram of the complete structure. Note: The green star is ClassI, the red circle is ClassII, and the blue square is ClassIII.

3.3 Screening of key regulatory targets for complete HERVs LTR

According to existing studies, genes in different ranges of upstream and downstream LTR may affect its function. In order to explore the potential interaction targets of HERVs LTR, We studied the LTR upstream and downstream of 536 complete HERVs sequences within the range of 20 kb (Quantitative and Distribution Characteristics of LTR Retrotransposons in the study) Tetraploid genes will be screened, a total of 632 genes including pseudogenes and long non-coding RNA will be screened. The protein interaction network will be constructed and screened using STRING database and Cytoscape. After appropriate deletion of unrelated genes and nodes, The final result included 141 nodes and 324 intersections (Figure 5). The genes most associated with other nodes were counted. The genes with 15 nodes or more were H4C6, H2BC12, H2BC11, H2BC9, TLR2, and CCR5.

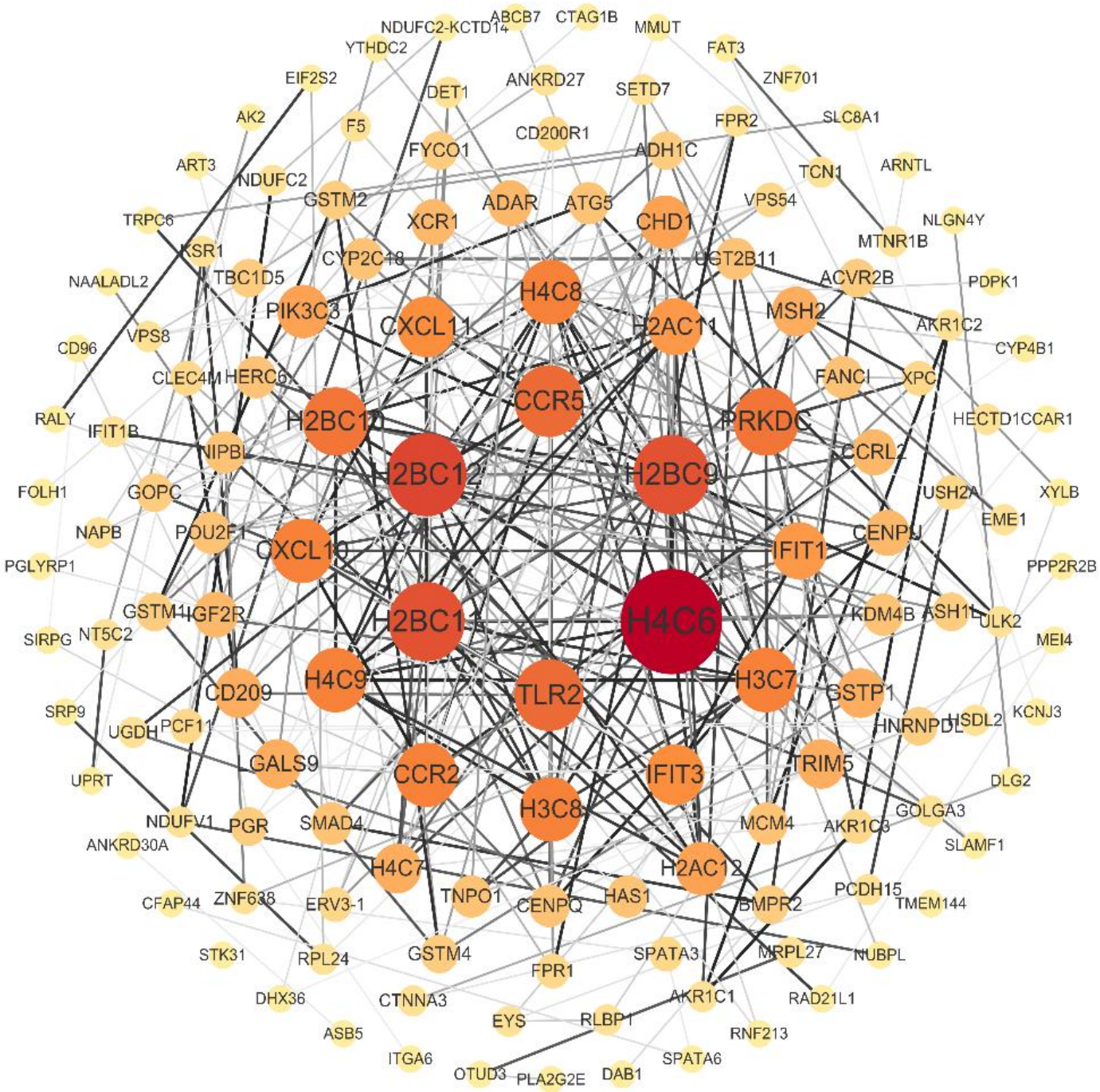

Figure5 Protein interaction network

In order to analyze other key expression protein modules, we constructed a subnetwork using Cytoscape plugin MCODE to screen key expression protein modules. A total of 11 protein modules were obtained, and only the top three protein mod ules were highlighted (Figure 6). A total of 54 interactions we re enriched in the module, with a total of 11 genes, namely H 2AC11, H2BC11, H4C9, H2BC12, H2BC10, H4C8, H3C7, H4C6, H3C8, H2BC9 and H2AC12. In module 2, 26 interactions were enriched, with a total of 10 genes, namely CCRL2, TRIM5, CX CL11, CXCL10, XCR1, HERC6, ADAR, TLR2, CCR5, and CCR2. M odule 3 is enriched to 14 interantagonism, a total of 6 genes, namely GSTM1, GSTM2, GSTP1, CYP2C18, ADH1C, GSTM4.

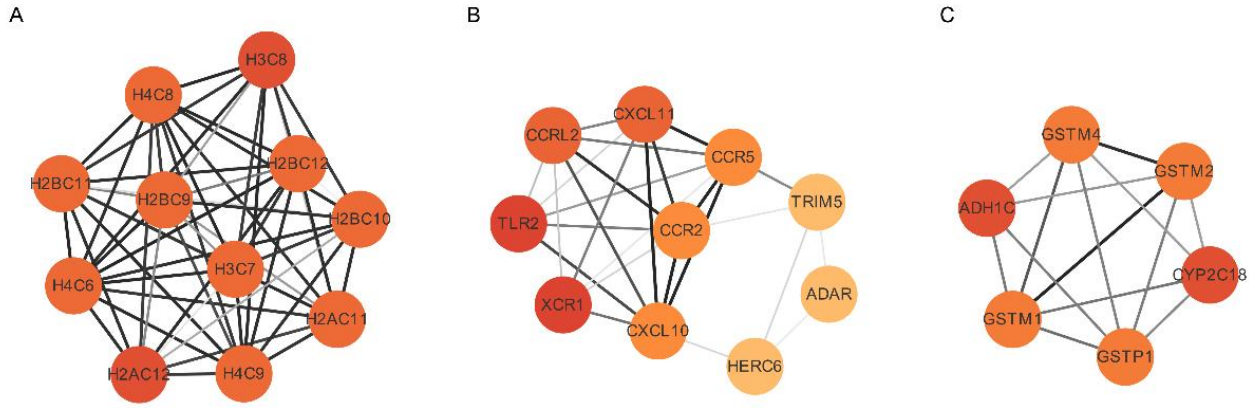

Figure6 Protein interaction subnetwork

3.3Functional enrichment analysis of HERVs LTR a djacent genes

A total of 632 genes in the upstream and downstream 20kb r ange of the structure-complete HERVs LTR sequence were id entified after the removal of duplicates, and gene ontological (GO) enrichment analysis was performed on them, as shown in the figure. Among them, 339 entries were screened for bi ological processes (BP), 352 entries for cell components (CC), and 342 entries for molecular functions (MF) (Figure 7). In te rms of biological processes, it mainly regulates DNA template transcription regulation, prostaglandin metabolism, defense response of gram-positive bacteria, telomere assembly, cell r esponse to jasmonic acid stimulation, and others regulate nit robenzene metabolism, chemotaxis, spermatogenesis, protei n localization chromatin containing CENP-A, nucleosome ass embly and other processes. The cellular components were m ainly enriched in nucleosomes, CENP-A containing nucleoso me, cell membrane, host cell, autophagosome, sperm head-t o-tail coupling device, cytoplasm, nucleus, late endosomes, o uter plasma membrane and so on. The molecular functions w ere mainly concentrated in chromatin structural components, metal ion binding, RNA polymerase II homeopathic regulator y region sequence-specific DNA binding, ketosteroid monoox ygenase activity, protein heterodimerization activity, chemok ine receptor activity, estradiol 17-β-dehydrogenase [NAD(P)+] activity, androsterone dehydrogenase activity, and DNA bind ing parts.

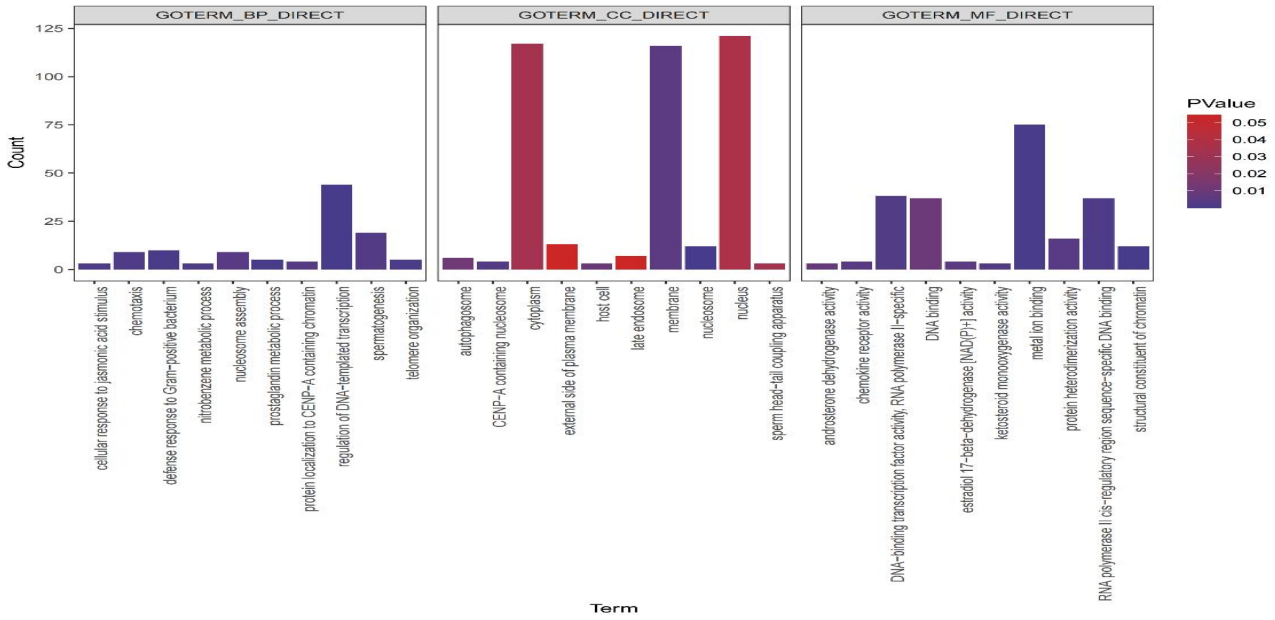

Figure7 GO enrichment analysis

In order to further study the mechanism of signaling pathways in the upstream and downstream 20kb range of the complete HERVs LTR, the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was conducted in this study. A total of 178 genes were concentrated, and only the first 10 pathways were shown (Figure 8). Results The pathways of herpes simplex virus type I infection, neutrophil extracellular trap formation, systemic lupus erythematosus, alcohol, chemical carcinogenicity -DNA admixture, virions - human immunodeficiency virus, cytochrome P450 metabolism to heteroorganisms, drug metabolism - cytochrome P450, chemical carcinogenicity - reactive oxygen species, platinum resistance were enriched.

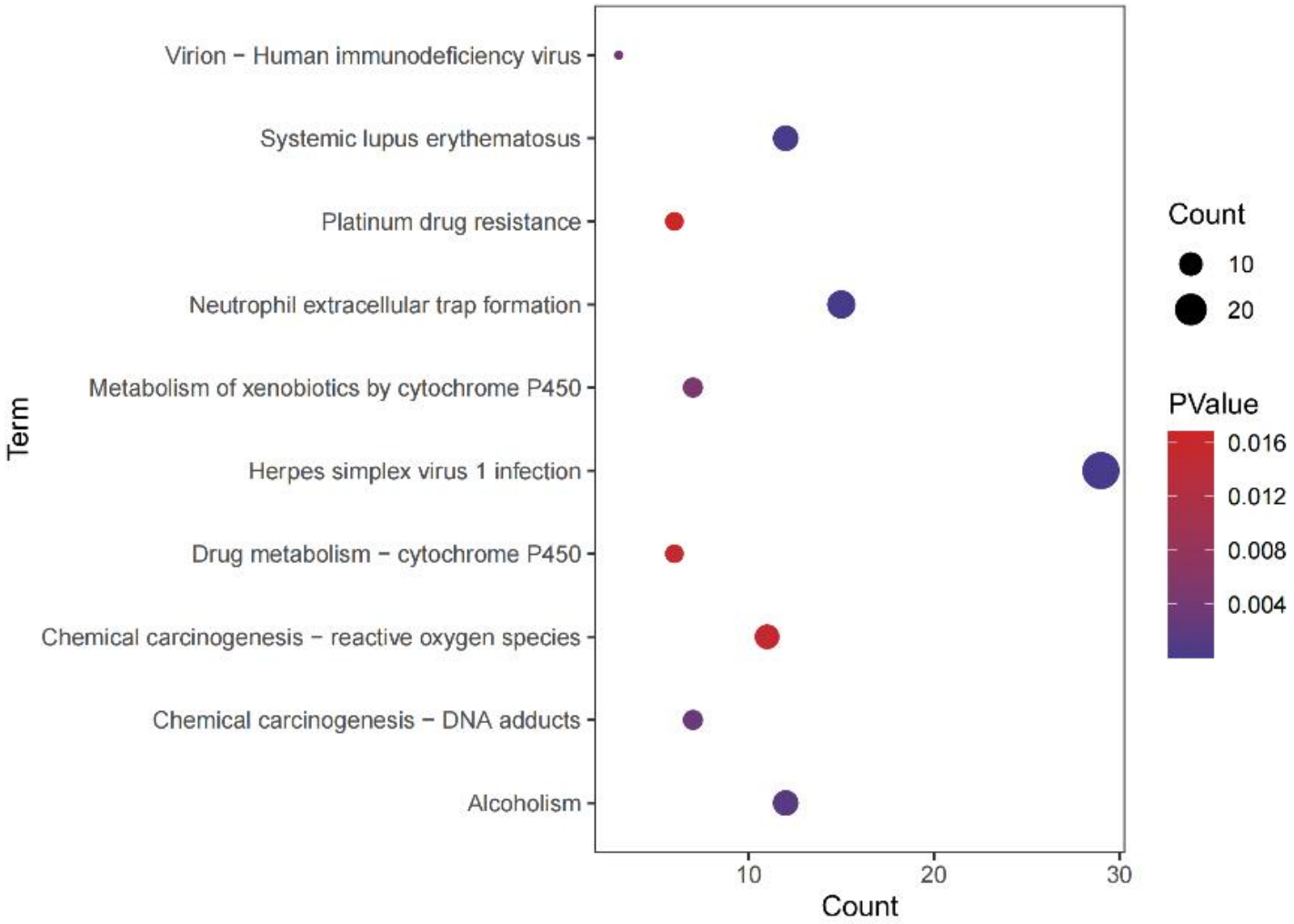

Figure8 Enrichment of KEGG pathway

4 Discussion

The method adopted in this study to identify HERVs in human genome is to search for endogenous retroviruses from the feature sequence LTR at both ends, and then identify each coding sequence and component between LTR based on this. LTRharvest and LTRdigest software are used to identify the LTR sequence from scratch. The ERVs sequences were identified by protein characteristic structure. Due to recombination events, LTR can remain at the integration site during evolution and serve as a marker of the original retrovirus integration site[29]. The traditional method of identifying ERVs sequences, such as using RepeatMasker software, is based on the principle of screening the interspersed repeats and low complexity DNA sequences of the query sequence. It uses the pre-compiled sequence library and a special scoring matrix to detect similar fragments in the query sequence, but its repeat library contains a large number of repeat families from model organisms. Duplicate libraries of non-model organisms can only do limited searches[30]. LTR_par and LTR_STRUC algorithms, while taking the principle of ab initio recognition of repeat sequences, LTRharvest software, using a different combination of features and algorithms than LTR_par and LTR_STRUC, is easier to scale further requirements than LTR_PAR and LTR_STRUC. The advantage of LTRharvest software is that (1) it can quickly read and calculate large data. (2) can incorporate known sequences into prediction parameters, and (3) can accept sequences in multiple FASTA formats[31]. For HERVs that are poorly annotated on the genome, being able to incorporate similar known sequences into predictions can greatly improve HERV annotation. The 5' LTR and 3' LTR at both ends undergo frequent gene recombination and deletion, resulting in the massive formation of solo-LTR elements[32]. LTR contains many transcriptional regulatory sequences of promoter and enhancer binding sites, which are very important and complex for the evolution and development of the host genome[33]. This study only analyzed the gene structure at both ends of LTR and inside the human whole gene, and did not include solo-LTRA total of 47666 HERVs sequences were obtained.

According to the protein domain characteristics, 605 complete HERVs were found statistically, accounting for only 1.27% of HERVs. The rest of the viral sequences were fragmented and silenced to varying degrees. ERVs play an important role in the evolution of vertebrates and produce related physiological effects on the host. The localization of ERVs in the human genome can lead to gene changes, insertional mutagenesis, non-homologous recombination, rearrangement and gene destruction. Due to the accumulation of mutations and deletions in the long process of ERVs evolution, the copy number of endogenous retroviruses of different species has a large difference in the host genome. Mutations such as frameshift and premature termination codon will lead to fragmentation of ERVs fragments, resulting in a large number of viral fragments, only a few of which retain their complete structure and still have regulatory functions. In the process of evolution, ERVs that retain their internal structure usually lose their coding characteristics and infectivity due to accumulated mutations over time, while some proviruses retain part of their coding ability, especially in the env region, but ERVs has not found that they are infectious at present[34]. env gene contains multiple deletions, which can increase intracellular migration and reduce host-to-host infection, thus leading to the termination of replication[35]. The relatively small proportion of env structure in the human genome may be related to rapid evolution under human host stress or functional redundancy. It has been proposed that the proliferation of retroviruses in the host germ line after endogenization is enhanced by the degradation of the env gene, especially in cases where the proviruses that have lost most of the env region appear to be the transmitters of the genome[36]. This may also be the reason why the env region, which has some viral protein coding ability, occupies the least proportion in the three protein domains pol, gag and env.

The results showed that intact HERVs were abundant on chromosomes 1 and 3, but rare on chromosomes 21 and 22. The distribution of fully structured HERVs in human chromosomes is different, which may be related to chromosome size, gene density, or the state of chromatin open. The gene-dense regions may be more likely to accumulate repeats, be affected by transposon insertion, and become part of the host genome in the long-term evolution[37]. The low distribution in chromosomes 21 and 22 May be related to the smaller physical size of these two chromosomes, or they may have stronger selection pressure in evolution[38].In the process of building the evolutionary tree, we obtained the reference retrovirus fragment region with relatively conserved RT, gag and pol were considered to be the most conserved, and env gene was more prone to mutation, so there was a great difference[39]. Among 335 complete structure HERVs, only 182 could be classified, and about half of the sequences were not well classified. Most of the sequences were classified into class ClassII, and the results of ClassI and Class ClassIII were less. In previous studies, Rodriguez et al. 's classification was based on ClassI being similar to gamma retroviruses, ClassII being similar to beta retroviruses, and ClassIII being similar to foam viruses[40]. In previous studies, et al. used differences in RT encoded by pol genes for evolutionary analysis, ClassI was similar to γ and ε retroviruses, ClassII was similar to α, β and δ retroviruses, and ClassIII was similar to foam viruses[41]. (The International Committee on Viruses (ICTV) divides retroviridae into two subfamilies: Ortoretroviridae and Retroviridae, with retrotranscription occurring within viral particles, while foamviridae particles contain double-stranded DNA, and RNA genomes present in ortoretroviruses[42]. This may be one of the reasons why foamvirus genera are not well clustered in evolutionary analysis. At present, there are many mutations, insertions and recombinations of endogenous viruses in the human genome, and the classification basis of HERVs is different. The reliability of the classification method needs to be further verified in the future.

Genes in the upstream and downstream range of LTR may be regulated by LTR promoters or enhancers. In this study, we obtained LTR genes in the upstream and downstream 20kb range of complete HERVs, and screened key targets and target genes to identify the potential regulatory role of LTR in HERVs in humans. The LTR of HERVs as a regulatory element in the genome has multiple potential roles. In HERVs, the gag, env, and pol domains may retain potential transcriptional and transposable activities, and the presence of these domains may enable them to play a regulat

ory role in the host genome[43]. In the results, we also enriched pseudogenes and lncrnas, which may indirectly regulate the activity of HERVs through competitive binding of miRNA[44].

The enrichment of highly connected genes suggests that HERVs may be involved in disease through chromatin remodeling (histone family) or immune pathways (TLR2 and CCR5). The genes enriched in Module 1 are all members of the histone family and contain palindromic terminating elements, which exist in the histone gene cluster of chromosome 6. The abnormal assembly of nucleosomes is closely related to post-translational modification of histones. Studies have found that ubiquitination of H2A can significantly enhance the mechanical stability of nucleosomes, and the loss of modification will lead to nucleosomal depolymerization and DNA exposure[45]. When the deposition of H3A ubiquitination is inhibited, it will lead to loose nucleosome structure, which will affect DNA replication and nucleosome assembly process[46]. TLR2, a member of the Toll-like receptor family, plays an important role in pathogen recognition and innate immune activation. CCR5 gene, a member of the β-chemokine receptor family, is a transmembrane protein, both of which are key molecules in immune signaling pathways. This suggests that HERVs may be involved in inflammation or antiviral response through LTR mediating immune-related genes. For example, HIV infection can activate the HERVs sequence in the host genome, and the env protein of HERVs can promote chronic inflammation through TLR2 and TLR4 pathways, further accelerating the progression of HIV disease[47]. CCR5 inhibitors may improve efficacy in HERVs-related inflammatory pathways by doubly blocking HIV entry[48].

In GO analysis and KEGG pathway enrichment analysis, we obtained a total of 632 LTR genes in the upstream and downstream of HERVs. The biological process results suggest that the upstream and downstream LTR genes of HERVs can play a role in nucleosome assembly and telomere assembly, and the abnormal nucleosome assembly is related to genome stability, and the insertion of HERVs may interfere with the transcriptional silencing mechanism of neighboring genes through LTR promoter activity. In vitro experiments have shown that nucleosome assembly efficiency is highly correlated with histone concentration and DNA sequence[49]. The enrichment of HERVs in Gram-positive bacteria is related to defense response and chemotaxis, suggesting that HERVS may be involved in host innate immune regulation, and may be potentially associated with the activity of chemokine receptor (CCR5), a key target that we have enriched. ERVs influence the defense system through RNA transcripts and affect host regulatory functions, such as RNA interference and innate immune sensing of double-stranded RNA[50]. The propagation of ERVs scatters interferon-induced enhancers, thereby forming an effective innate immunomodulatory network[51]. In another study, HERVs can be used as a proximal regulatory element to promote interferon (IFN) response[52]. In response to interferon stimulation, IFN-stimulating genes such as AIM2, IFIT1, IFIT2, IFIT3, STAT1 and IRF are activated through the JAX-STAT pathway, and STAT1, STAT2 and IRF1 seem to be closely related to HERVs in the regulatory network[53]. The enrichment of items such as prostaglandin metabolism can reflect that HERVs affect the stress response of cells through lipid metabolism. The functional localization of cell components to nucleosomes and CENP-A chromatin sites suggests that HERVs may affect the centromere region through insertion, leading to chromosome separation. At present, it has been found that the insertion of HERV-Ks is correlated with abnormal telomerase activity. Studies have shown that HERVs-interferon signal induced by TERT gene stimulates the expression of chemokines and contributes to the establishment of immunosuppressor tumor microenvironment[54]. The molecular function results showed that HERVs may be involved in cancer or immune diseases by regulating steroid metabolism, such as ketosteroid monooxygenase and estradiol dehydrogenase. There was a strong positive correlation between AR activity and the expression of EERV3-1 and HERV-Ks[55]. The long terminal repeating transposon-like element B (THE1B) of HERVs selectively controls the expression of EPcorticotropin releasing hormone in the placenta, and the 5 'of THE1B interacts with the transcription factor DLX3 expressed in the placenta, thereby influencing the birth time of the fetus[56].

In the KEgg-enriched pathway, HERVs may be associated with viral infection and immune escape. The enriched virion-human immunodeficiency virus pathway suggests that HERVs may interfere with host antiviral immunity by encoding env proteins to mimic viral antigens. Studies have shown that evolutionarily young HERVs can act as enhancers of immune reactivity in COVID-19 patients[57]. The study results of Castro et al. indicated the regulatory role of gene transcription during arbovirus infection induced by HERVs[58]. Enrichment of the NETs pathway suggests that HERVs activation may induce neutrophils to release DNA webs, exacerbating the process of autoimmune disease and directly related to the systemic lupus erythematosus pathway we found. Increased levels of HERVs-related proteins have been found in patients with systemic lupus erythematosus[59]. The expression of ERV-K102 is significantly elevated in the blood of SLE patients and is associated with higher levels of autoantibodies and interferon status, while immunostimulation-specific HERV-K envelope proteins activate neutrophils in SLE IgG immune complexes[60]. Chemical carcinogenicity - The DNA addition pathway suggests that HERVs insertion may induce genomic instability, synergistic with other factors to promote carcinogenic mutations. Studies have shown that high LTR expression level of HERVs predicts high survival rate of patients with small cell lung cancer after chemotherapy[61].

Reference

1.Ra W. On the concept and elucidation of endogenous retroviruses[J/OL]. Philosophical transactions of the Royal Society of London. Series B, Biological Sciences, 2013, 368(1626)

2.A H, M G, P J. Broad-scale phylogenomics provides insights into retrovirus-host evolution[J/OL]. Proceedings of the National Academy of Sciences of the United States of America, 2013, 110(50)

3.X X, H Z, Z G, et al. Endogenous retroviruses of non-avian/mammalian vertebrates illuminate diversity and deep history of retroviruses[J/OL]. PLoS Pathogens, 2018, 14(6)

4.Gifford R, Tristem M. The evolution, distribution and diversity of endogenous retroviruses[J]. Virus Genes, 2003, 26(3): 291–315.

5.Gifford R J, Blomberg J, Coffin J M, et al. Nomenclature for endogenous retrovirus (erv) loci[J]. Retrovirology, 2018, 15(1): 59.

6.Hh K, Jv M. Mobile dna in health and disease[J/OL]. The New England journal of Medicine, 2017, 377(4)

7.C V, Ra W, Dj G. Human rna “rumor” viruses: the search for novel huma n retroviruses in chronic disease[J/OL]. Microbiology and molecular biolog y reviews : MMBR, 2008, 72(1)
8.Jp S. Studies of endogenous retroviruses reveal a continuing evolutionar y saga[J/OL]. Nature reviews. Microbiology, 2012, 10(6)
9.Huang Q, Chen H, Li J, et al. Epigenetic and non-epigenetic regulation of syncytin-1 expression in human placenta and cancer tissues[J]. Cellular Sig nalling, 2014, 26(3): 648–656.
10.Mi S, Lee X, Li X, et al. Syncytin is a captive retroviral envelope protein i nvolved in human placental morphogenesis[J]. Nature, 2000, 403(6771): 785–789.
11.Douville R, Liu J, Rothstein J, et al. Identification of active loci of a hum an endogenous retrovirus in neurons of patients with amyotrophic lateral sclerosis[J]. Annals of Neurology, 2011, 69(1): 141–151.
12.Eirew P, Steif A, Khattra J, et al. Dynamics of genomic clones in breast c ancer patient xenografts at single-cell resolution[J]. Nature, 2015, 518(75 39): 422–426.
13.Aftab A, Shah A A, Hashmi A M. Pathophysiological role of herv-w in sc hizophrenia[J]. The Journal of Neuropsychiatry and Clinical Neurosciences, 2016, 28(1): 17–25.
14.Küry P, Nath A, Créange A, et al. Human endogenous retroviruses in ne urological diseases[J]. Trends in Molecular Medicine, 2018, 24(4): 379–39 4.
15.Katsura Y, Asai S. Evolutionary medicine of retroviruses in the human g enome[J]. The American Journal of the Medical Sciences, 2019, 358(6): 38 4–388.
16.Ruprecht K, Mayer J, Sauter M, et al. Endogenous retroviruses and can cer[J]. Cellular and Molecular Life Sciences: CMLS, 2008, 65(21): 3366–338 2.
17.Hu T, Zhu X, Pi W, et al. Hypermethylated ltr retrotransposon exhibits enhancer activity[J]. Epigenetics, 2017, 12(3): 226–237.
18.Geis F K, Goff S P. Silencing and transcriptional regulation of endogeno us retroviruses: an overview[J]. Viruses, 2020, 12(8): 884.
19.Chen J, Foroozesh M, Qin Z. Transactivation of human endogenous ret roviruses by tumor viruses and their functions in virus-associated maligna ncies[J]. Oncogenesis, 2019, 8(1): 6.
20.Em J, D B, R T, et al. Human endogenous retroviral protein triggers defi cit in glutamate synapse maturation and behaviors associated with psych osis[J/OL]. Science Advances, 2020, 6(29)
21.Franke V, Ganesh S, Karlic R, et al. Long terminal repeats power evoluti on of genes and gene expression programs in mammalian oocytes and zy gotes[J]. Genome Research, 2017, 27(8): 1384–1394.
22.Lee J-Y, Kong G. Roles and epigenetic regulation of epithelial-mesench ymal transition and its transcription factors in cancer initiation and progre ssion[J]. Cellular and Molecular Life Sciences: CMLS, 2016, 73(24): 4643–4 660.
23.王潇. 川金丝猴、滇金丝猴和黑叶猴内源性逆转录病毒的研究[D/OL]. 北京林业大学, 2020
24.唐洲. 内源性反转录病毒在主要农业动物基因组上分布及其特征的研究[D/OL]. 华中农业大学, 2016
25.Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genet ics analysis version 11[J]. Molecular Biology and Evolution, 2021, 38(7): 30 22–3027.
26.Capella-Gutiérrez S, Silla-Martínez J M, Gabaldón T. TrimAl: a tool for a utomated alignment trimming in large-scale phylogenetic analyses[J]. Bioi nformatics (Oxford, England), 2009, 25(15): 1972–1973.
27.Letunic I, Bork P. Interactive tree of life (itol) v5: an online tool for phyl ogenetic tree display and annotation[J]. Nucleic Acids Research, 2021, 49(W1): W293–W296.
28.Price M N, Dehal P S, Arkin A P. FastTree: computing large minimum ev olution trees with profiles instead of a distance matrix[J]. Molecular Biolo gy and Evolution, 2009, 26(7): 1641–1650.
29.Yang L, Malhotra R, Chikhi R, et al. Recombination marks the evolution ary dynamics of a recently endogenized retrovirus[J]. Molecular Biology a nd Evolution, 2021, 38(12): 5423. DOI:10.1093/molbev/msab252.
30.Jurka J. Repbase update: a database and an electronic journal of repeti tive elements[J]. Trends in Genetics: TIG, 2000, 16(9): 418–420.
31.Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible s oftware for de novo detection of ltr retrotransposons[J]. BMC Bioinforma tics, 2008, 9: 18.
32.Stoye J P. Endogenous retroviruses: still active after all these years?[J]. Current Biology: CB, 2001, 11(22): R914-916.
33.Song Y, Wen H, Zhai X, et al. Functional bidirectionality of erv-derived l ong non-coding rnas in humans[J]. International Journal of Molecular Scie nces, 2024, 25(19): 10481.
34.Russ E, Iordanskiy S. Endogenous retroviruses as modulators of innate immunity[J]. Pathogens (Basel, Switzerland), 2023, 12(2): 162.
35.Chabukswar S, Grandi N, Lin L-T, et al. Envelope recombination: a majo r driver in shaping retroviral diversification and evolution within the host genome[J]. Viruses, 2023, 15(9): 1856.
36.Magiorkinis G, Gifford R J, Katzourakis A, et al. Env-less endogenous ret roviruses are genomic superspreaders[J]. Proceedings of the National Aca demy of Sciences of the United States of America, 2012, 109(19): 7385–7 390.
37.J S, M D, Z X, et al. The landscape of hervrnas transcribed from human endogenous retroviruses across human body sites[J/OL]. Genome Biology, 2022, 23(1)[2024–04–01]. https://pubmed.ncbi.nlm.nih.gov/36329469/. DOI:10.1186/s13059-022-02804-w.
38.Li C, Qian Q, Yan C, et al. HervD atlas: a curated knowledgebase of asso ciations between human endogenous retroviruses and diseases[J]. Nuclei c Acids Research, 2024, 52(D1): D1315–D1326.
39.Zhang Y, Wang G, Zhu Y, et al. Exploring the role of endogenous retrovi ruses in seasonal reproductive cycles: a case study of the erv-v envelope g ene in mink[J]. Frontiers in Cellular and Infection Microbiology, 2024, 14: 1404431.
40.Vargiu L, Rodriguez-Tomé P, Sperber G O, et al. Classification and chara cterization of human endogenous retroviruses; mosaic forms are commo n[J]. Retrovirology, 2016, 13: 7.
41.Yedavalli V R K, Patil A, Parrish J, et al. A novel class iii endogenous retr ovirus with a class i envelope gene in african frogs with an intact genome and developmentally regulated transcripts in xenopus tropicalis[J]. Retrov irology, 2021, 18(1): 20.
42.Greenwood A D, Ishida Y, O’Brien S P, et al. Transmission, evolution, an d endogenization: lessons learned from recent retroviral invasions[J]. Mic robiology and Molecular Biology Reviews : MMBR, 2017, 82(1): e00044-1 7.
43.Yi J M. Epigenetic regulation of hervs: implications for cancer immunot herapy[J]. Genes & Genomics, 2024, 46(11): 1303–1312.
44.Zhang T, Zheng R, Li M, et al. Active endogenous retroviral elements in human pluripotent stem cells play a role in regulating host gene expressio n[J]. Nucleic Acids Research, 2022, 50(9): 4959–4973.
45.Xiao X, Liu C, Pei Y, et al. Histone h2a ubiquitination reinforces mechan ical stability and asymmetry at the single-nucleosome level[J]. Journal of t he American Chemical Society, 2020, 142(7): 3340–3345. DOI:10.1021/jac s.9b12448.
46.Zhang W, Feng J, Li Q. The replisome guides nucleosome assembly duri ng dna replication[J]. Cell & Bioscience, 2020, 10: 37.
47.方浩, 杨金轩, 罗荣华, et al. CCR5单核苷酸多态性影响HIV感染及艾滋病疾病进程的研究进展[J]. 病毒学报, 2022, 38(4): 991–1000.

48.Karuppusamy K V, Babu P, Thangavel S. The strategies and challenges of ccr5 gene editing in hematopoietic stem and progenitor cells for the tre atment of hiv[J]. Stem Cell Reviews and Reports, 2021, 17(5): 1607–1618. DOI:10.1007/s12015-021-10145-7
49.Anonymous. 核小体体外组装的理论模型建立和实验研究[EB/OL](2020–09–01)[2025–02–09]. https://d.wanfangdata.com.cn/Thesis/D01972189.
50.Meyer T J, Rosenkrantz J L, Carbone L, et al. Endogenous retroviruses: with us and against us[J]. Frontiers in Chemistry, 2017, 5: 23.
51.Chuong E B, Elde N C, Feschotte C. Regulatory evolution of innate imm unity through co-option of endogenous retroviruses[J]. Science (New York, N.Y.), 2016, 351(6277): 1083–1087.
52.Studstill C, Huang N, Sundstrom S, et al. Apoptotic caspases suppress e xpression of endogenous retroviruses in hpv31+ cells that are associated with activation of an innate immune response[J]. Viruses, 2024, 16(11): 1695.
53.Wang Y, Liu M, Guo X, et al. Endogenous retrovirus elements are co-ex pressed with ifn stimulation genes in the jak-stat pathway[J]. Viruses, 2022, 15(1): 60.
54.Mao J, Zhang Q, Wang Y, et al. TERT activates endogenous retroviruses to promote an immunosuppressive tumour microenvironment[J]. EMBO Reports, 2022, 23(4): e52984.
55.Alizadeh-Ghodsi M, Owen K L, Townley S L, et al. Potent stimulation of the androgen receptor instigates a viral mimicry response in prostate can cer[J]. Cancer Research Communications, 2022, 2(7): 706–724.
56.Dunn-Fletcher C E, Muglia L M, Pavlicev M, et al. Anthropoid primate–s pecific retroviral element the1b controls expression of crh in placenta and alters gestation length[J]. PLoS Biology, 2018, 16(9): e2006337.
57.Yoshida R, Ohtani H. Activation of evolutionarily young endogenous re troviruses is implicated in covid-19 immunopathology[J]. Genes to Cells, 2025, 30(1): e13194.
58.De Castro F L, Brustolini O J B, Geddes V E V, et al. Modulation of herv expression by four different encephalitic arboviruses during infection of h uman primary astrocytes[J]. Viruses, 2022, 14(11): 2505.
59.Greenig M. HERVs, immunity, and autoimmunity: understanding the c onnection[J]. PeerJ, 2019, 7: e6711.
60.Tokuyama M, Gunn B M, Venkataraman A, et al. Antibodies against hu man endogenous retrovirus k102 envelope activate neutrophils in system ic lupus erythematosus[J]. The Journal of Experimental Medicine, 2021, 218(7): e20191766.
61.Russo M, Morelli S, Capranico G. Expression of down-regulated erv ltr elements associates with immune activation in human small-cell lung can cers[J]. Mobile DNA, 2023, 14: 2.

Summary: This study provides a relatively comprehensive human genome HERVs annotation database, which is of great value for understanding the regulation of human HERVs on human genome. Importantly, the identified upstream and downstream regulatory genes of LTR also provide further reference value for the evolution and regulatory elements of HERVs, suggesting that HERVs may participate in the disease mechanism by regulating chromatin structure, immune response and metabolic pathways. LTR or interacting genes that target HERVs may become novel strategies for cancer or autoimmune diseases, providing a theoretical basis for drug target screening and therapeutic potential. Further studies and experimental validation are needed to evaluate the potential impact of HERVs expression on host health and its role in specific disease signaling pathways. Meanwhile, the regulatory role of HERVs in specific pathways can be further analyzed in combination with epigenome and single-cell transcriptome.**Competing interests:**The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.**Acknowledgements:**The authors would like to express their gratitude to all participants who participated in this study.**Author Contributions:**HLW: Conceptualization, Formal analysis, Writing - original draft. BYL: Conceptualization, review & editing. CLJ: Analysis, writing - methods. TWL: Review & editing.WL: review & editing. QXM, TYY: Platform software environment construction.XYP, YXH: Visualization of results. JPZ: Picture beautification, review. YLH*: Conceptualization, review & editing. All authors have read and agreed to the published version of the manuscript.**Funding:**This work is supported by the Key Research and Development Program of Guangxi(GuikeAB22035027), the National Key Research and Development Program of China(2023YFC2605400), and the Natural Science Foundation (2025KY0127).**Data Availability:**The software used is derived from the GenomeTools package (Linux platform). All R packages are available online.**Ethics approval and consent to participate:**Not applicable.**Consent for publication:**Not applicable.