Nucleic Acids Research, 2024, **52**, D1315–D1326
https://doi.org/10.1093/nar/gkad904
Advance access publication date: 23 October 2023
Database issue

HervD Atlas: a curated knowledgebase of associations between human endogenous retroviruses and diseases

Cuidan Li[1,†], Qiheng Qian[1,2,3,†], Chenghao Yan[1,3,†], Mingming Lu[1,2,†], Lin Li[4,†], Pan Li[1,3], Zhuojing Fan[1,2], Wenyan Lei[1,3], Kang Shang[1,3], Peihan Wang[1,3], Jie Wang[1,3], Tianyi Lu[1,3], Yuting Huang[5], Hongwei Yang[5], Haobin Wei[1,3], Jingwan Han[4], Jingfa Xiao[1,2,3,*] and Fei Chen[1,3,6,*]

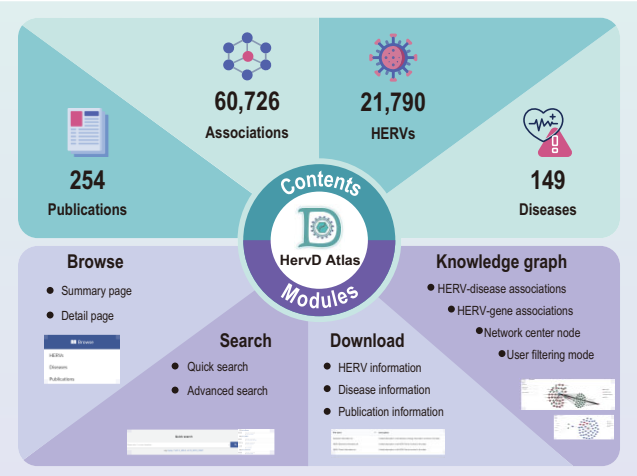
[1]CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China
[2]National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China
[4]State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing 100071, China
[5]State Key Laboratory of Elemento-Organic Chemistry, College of Chemistry, Nankai University, Tianjin 300071, China
[6]Beijing Key Laboratory of Genome and Precision Medicine Technologies, Beijing 100101, China

[*]To whom correspondence should be addressed. Tel: +86 10 84097460; Fax: +86 10 84097720; Email: chenfei@big.ac.cn
Correspondence may also be addressed to Jingfa Xiao. Email: xiaojingfa@big.ac.cn
[†]The authors wish it to be known that, in their opinion, the first five authors should be regarded as Joint First Authors.

Abstract

Human endogenous retroviruses (HERVs), as remnants of ancient exogenous retrovirus infected and integrated into germ cells, comprise ~8% of the human genome. These HERVs have been implicated in numerous diseases, and extensive research has been conducted to uncover their specific roles. Despite these efforts, a comprehensive source of HERV-disease association still needs to be added. To address this gap, we introduce the HervD Atlas (https://ngdc.cncb.ac.cn/hervd/), an integrated knowledgebase of HERV-disease associations manually curated from all related published literature. In the current version, HervD Atlas collects 60 726 HERV-disease associations from 254 publications (out of 4692 screened literature), covering 21 790 HERVs (21 049 HERV-Terms and 741 HERV-Elements) belonging to six types, 149 diseases and 610 related/affected genes. Notably, an interactive knowledge graph that systematically integrates all the HERV-disease associations and corresponding affected genes into a comprehensive network provides a powerful tool to uncover and deduce the complex interplay between HERVs and diseases. The HervD Atlas also features a user-friendly web interface that allows efficient browsing, searching, and downloading of all association information, research metadata, and annotation information. Overall, the HervD Atlas is an essential resource for comprehensive, up-to-date knowledge on HERV-disease research, potentially facilitating the development of novel HERV-associated diagnostic and therapeutic strategies.

Graphical abstract



Received: August 14, 2023. **Revised:** October 3, 2023. **Editorial Decision:** October 4, 2023. **Accepted:** October 6, 2023

Introduction

Human endogenous retroviruses (HERVs), constituting ∼8% of the human genome, are integrated remnants of ancient exogenous retroviruses in germ cells, following Mendelian inheritance (1–5). These ancient retroviral 'roommates' of humans were previously regarded as mere 'junk DNA' (5–7). However, recent research has revealed their significant roles in both normal physiology, such as their regulatory functions in embryonic stem cells (8–11), and in pathological conditions like tumor development (12,13). In healthy individuals, HERVs are typically tightly controlled and remain transcriptionally silent (14); nonetheless, some of them can be reactivated under specific pathological circumstances (13,15). However, it's worth emphasizing that not all HERVs are linked to diseases.

Recent studies have reported the role of aberrant HERV expressions in the progression of various diseases, including cancer such as leukemia (16), clear cell renal carcinoma (17) and hepatocellular carcinoma (18), infectious diseases like COVID-19 (19), HIV (20) and Hepatitis B virus infection (21), age-associated disorders including aging (22) and progeroid syndrome (23), mental disorder like schizophrenia (24) and bipolar disorder (25), autoimmune disorders like systemic lupus erythematosus (26), and neurological diseases like multiple sclerosis (27) and Alzheimer's disease (28). Furthermore, a therapeutic strategy targeting a HERV protein has advanced to phase III trials for treating multiple sclerosis (29), and HERV-E-derived peptide autologous T-cell therapy for clear cell renal cell carcinoma is currently in phase I trial (https://clinicaltrials.gov/ct2/show/NCT03354390). Consequently, HERVs are increasingly believed to play important roles in complex disease progression (14,27), making them as a research hotspot in biomedical research with immense potential for innovative diagnostic and therapeutic applications.

With the development in sequencing and detection technologies these years, considerable effort has been dedicated to uncover the roles of HERVs in various diseases, unraveling numerous important HERV-disease associations (13,30–32). A consolidated resource incorporating these findings will be valuable in the investigation of HERVs and related diseases. Existing databases related to HERVs (e.g. HERVd, gEVE, EnHERV, dbHERV-REs, HESAS, TranspoGene) can be classified into three categories. HERVd (33,34) and gEVE (35) provide general HERV information for understanding their characteristics (e.g. genomic loci, nucleotide and amino acid sequences, functional annotations); EnHERV (36) facilitates the exploration of HERVs' neighboring genes while also enabling analyses of related functional enrichments; dbHERV-Res (37), HESAS (38), TranspoGene (39) focus on HERVs' regulatory function for understanding their impact on gene transcription.

While these databases have provided valuable resources for studying HERVs, they present a few limitations. All the databases mentioned above primarily focus on the basic information of HERVs (including classification, genomic loci, strand, length, nucleotide and amino acid sequences, copy number, nearby genes and corresponding distance, and functional annotations) and their effects on transcription, neglecting their association with diseases. They also need comprehensive integration of HERV information from related publications and external databases. Additionally, the presentation in these databases is merely displayed in the format of listings, rather than intuitive visualizations like knowledge graphs. To tackle these limitations and align with the rapidly advancing HERV research, it is urgent to develop a comprehensive knowledgebase that effectively curates and integrates the existing data resources of HERV-disease associations from all accessible related publications, and offers interactive visualizations encompassing HERVs, related diseases, and implicated genes.

To address these issues, we present HervD Atlas, a curated knowledgebase of HERVs and disease associations. HervD Atlas manually collects tens of thousands of high-quality HERV-disease associations from extensive publications. It employs a unified nomenclature system for curated HERVs and integrates essential information from both publications and existing databases. Additionally, HervD Atlas builds an interactive information-intensive knowledge graph containing all collected HERV-disease associations and related/affected genes. In summary, HervD Atlas offers the latest updated and integrated resources, comprehensive information, and interpretation platforms for HERV-disease association research. Therefore, HervD Atlas will substantially enhance HERV-disease association research and deepen our understanding of the fundamental role of HERVs in human diseases.

Data curation and database development

Knowledge curation and integration

To obtain high-quality HERV-disease associations, we implemented a standardized curation process consisting of three main steps: publication search, study curation, and association curation. *Publication search*: First, we conducted a comprehensive literature search using NCBI PubMed. Employing pre-defined keywords 'HERV', 'human endogenous retrovirus', '(HERV) AND (disease)', and '(human endogenous retrovirus) AND (disease)', we identified a total of 4692 non-redundant publications (Supplementary Figure S1). *Study curation*: Next, our curation team conducted a detailed manual review of these articles, only those containing necessary descriptions of significant HERV-disease associations were selected for inclusion in HervD Atlas. Specifically, the criteria were: (i) publications reporting significant upregulation or downregulation of HERVs in disease contexts, substantiated by *P*-values or adjusted *P*-values; (ii) publications investigating the mechanisms underlying disease-associated changes following HERVs stimulation or inhibition through experimental methods. After a conscientious review, 254 publications were carefully screened and incorporated into our knowledgebase. We then meticulously curated the study information from each publication, which involved reported HERVs, associated diseases, methodologies, sample sources, data links and population details. Association Curation: Finally, we retained the data for associations only if they meet one of the following criteria: (i) HERV-disease associations exhibiting significant statistical relevance, with a threshold of P-value <0.05 or adjusted *P*-value <0.05; (ii) HERV-disease associations from mechanism investigation studies that reported exact disease phenotype changes or significant gene regulation in disease contexts. We then documented the information of these associations, including: association at omics level (e.g. copy number, single nucleotide variation, RNA expression, protein expression, DNA methylation, histone modification), HERV trends in diseases, $\log_2$FC, corresponding *P*-values or adjusted *P*-values, effected genes and phenotypic changes. We

also recorded available HERV-gene correlation with a significance of *P*-value <0.05 as reported in the publication.

Moreover, to enhance the comprehensiveness of the HERV information, the expression levels across various human tissues from the Genotype-Tissue Expression (GTEx) project were integrated (40,41), offering a broader perspective of the corresponding HERVs. An overview of the data structure in HervD Atlas is illustrated in Supplementary Figure S2. To standardize disease names and definitions, we utilized disease-related terms and identifiers from multiple ontologies, including Disease Ontology (DO), Experimental Factor Ontology (EFO), Medical Subject Headings (MeSH), Online Mendelian Inheritance in Man (OMIM), and National Cancer Institute (NCI). Utilizing these ontologies, diseases were mapped and categorized into 14 distinct subcategories, including cancer, viral infectious diseases, nervous system diseases, mental disorders, immune system diseases, urogenital diseases, cardiovascular system diseases, digestive system diseases, skin and connective tissue diseases, genetic diseases, and bacterial infectious diseases. This methodology has enabled a unified and comprehensive classification within the HervD Atlas.

HERV classification and annotation

The classification of HERVs was systematically carried out in two main categories: HERV-Term and HERV-Element. HERV-Term refers to individual HERVs with specific genomic locations, whereas HERV-Element encompasses combinations of multiple copies of HERVs present within the genome. To ensure consistency and clarity across the knowledgebase, all HERVs were subjected to a comprehensive process of renaming and encoding. For HERV-Term, information regarding the chromosomal location, specific type (e.g. ERV1, ERV2, ERV3) (42,43) and serial number were considered; while for HERV-Element, only the specific type and serial number were employed.

We also conducted a comprehensive annotation for the HERVs through the following steps: (i) *Manual curation*: Details such as group (3), alias (3,44–46), virus source (42,47,48), earliest shared ancestor (46,47,49–56) and description (42,44) were manually curated from additional publications and databases; (ii) *External database links integration*: Relevant links to HERVs from sources such as PGG.SV (57), Dfam (42) and dbHERV-Res (37) were incorporated; (iii) *Genomic re-annotation*: The genomic locus, region type and nearby genes for each HERV were re-annotated based on the hg38 human reference genome.

Knowledge graph construction

The HervD Atlas features a newly constructed knowledge graph designed to enhance the visualization and interpretation of HERV-disease associations. This graph is organized into two primary panels: *Disease network* (focused on disease as the core node) and *HERV network* (centered around HERV as the core node). The construction of the knowledge graph began with the definition of three main entities: HERV, disease, and gene. Within the HERV entities, further categorization was done into HERV-Term and HERV-Element, aligning with the previous classification process. Relationships were then defined under two main aspects: the connections between disease entities and HERV entities, and the links between HERV entities and gene entities.

To provide deeper insights into the associations, an additional attribute—the number of supporting evidence—was added to these relationships. This attribute serves as an indicator of the strength and reliability of each connection. For an efficient and focused representation, nodes with over 100 links display only the top 100 associated entities supported by the most evidence.

Database implementation

The HervD Atlas was built by the SpringBoot framework (https://spring.io/projects/spring-boot) and Mybatis (https://mybatis.org/mybatis-3) for the back-end system. Several technologies were implemented for the front-end, including AJAX (Asynchronous JavaScript and XML), Bootstrap (https://getbootstrap.com), CSS (Cascading Style Sheets), Semantic UI (https://semantic-ui.com), Select2 (https://select2.org/), JQuery (https://jquery.com), HTML (HyperText Markup Language) and Thymeleaf (https://www.thymeleaf.org). For data rendering and visualization, ECharts was employed, and MySQL was used as the database engine for data storage and querying.

Database contents and usage

HervD Atlas presents a comprehensive integration of high-quality HERV-disease associations (Figure 1), which is achieved through careful manual curation of data from selected scientific literature in a standardized form to ensure the quality and relevance of information. Further, the Atlas creates a visual and interactive knowledge graph, highlighting the complex relationships between HERVs, diseases, and genes. Therefore, HervD Atlas serves as a well-organized repository of HERV-disease associations, provides a comprehensive landscape of HERVs' roles in human diseases, assists in a deeper understanding of disease mechanisms from a new perspective, and further facilitates the development of innovative diagnostic and therapeutic strategies.

Comprehensive association knowledge for diverse HERVs and diseases

In the current version, we have manually collected a total of 60 726 curated high-quality HERV-disease associations for humans from 254 filtered publications, covering 21 790 HERVs belonging to six types (e.g. ERV1, ERV2) (Figure 2A), 149 mapped ontological diseases grouped into 14 categories (e.g. cancer, immune system diseases, nervous system diseases) (Figure 2B) and 610 influenced genes. The curated factors for each HERV type are provided in Supplementary Table S1. A total of 21 049 HERV-Terms and 741 HERV-Elements are collected in our knowledgebase, among which the HERVL and MaLR group is the most prevalent, respectively (Figure 2C, D; Supplementary Figure S3).

Most HERVs are associated with multiple diseases, with a median of three diseases linked to each HERV. Particularly noteworthy are the top 15 HERVs that exhibit the highest number of disease associations (Figure 2E), indicating their multifaceted roles in various diseases. Herein, ERV1_0001 (ERVW-1) and ERV2_0001 (HERV-K-env), as the two hottest HERV-Elements, are found to have 227 and 173 associations with 32 and 28 diseases, respectively. Correspondingly, chr4_ERV1_00001 (ERVMER34-1) and chr6_ERV1_00003 (ERVFRD-1) are the two most prominent HERV-Terms as-

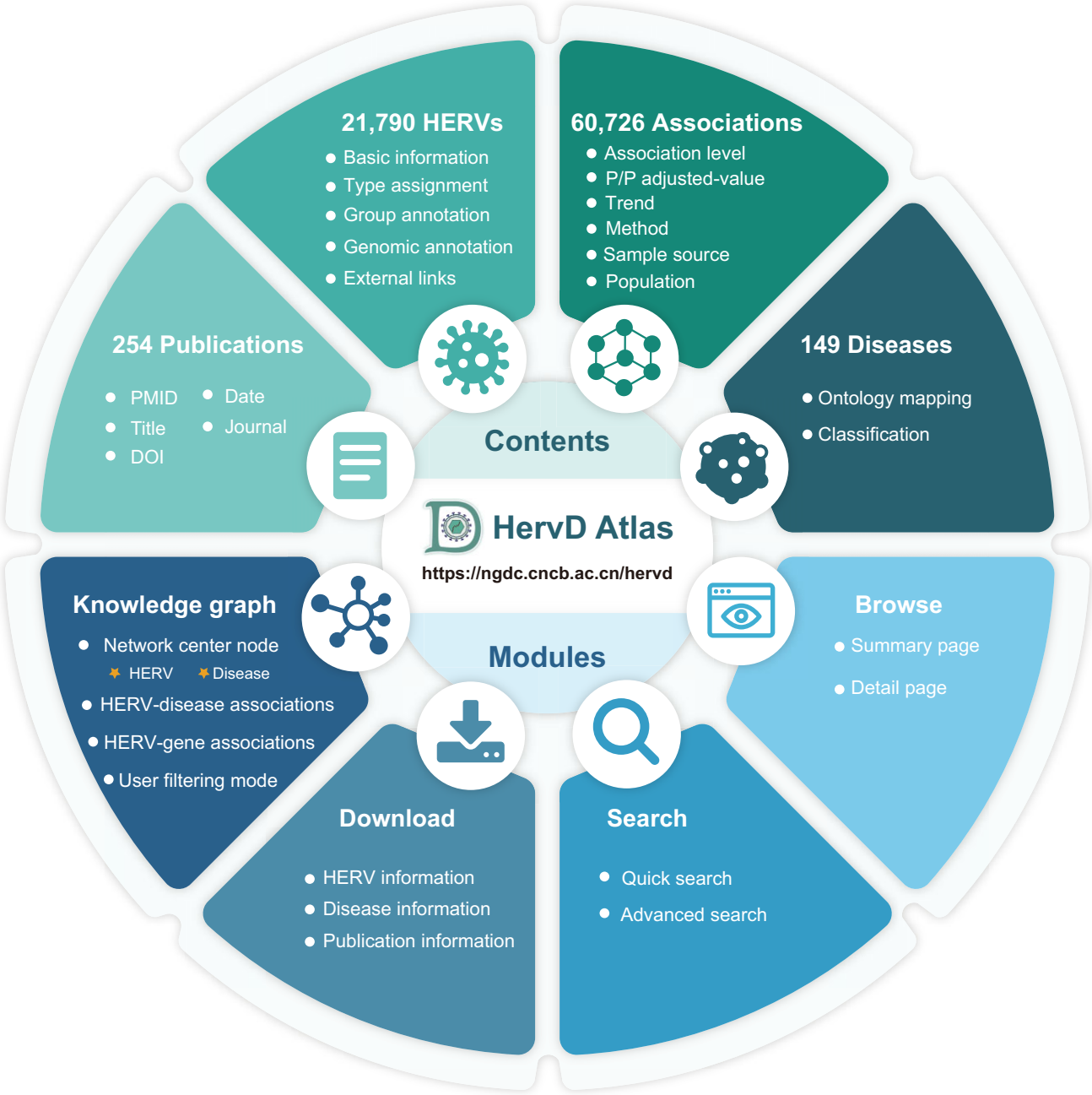


Figure 1. Schematic overview of the contents and modules in HervD Atlas. The data contents and corresponding statistics including associations, HERVs, diseases and publications contained within HervD Atlas (above), and four functional modules, i.e. information browse, multiple search channels, interactive knowledge graph and comprehensive information available for download (below).

sociated with 27 and 22 diseases, respectively. Additionally, the genes influenced by HERVs during disease progression are also collected in our knowledgebase. For instance, ERVW-1 could trigger off PP2A/AKT1/GSK3 pathway, which involves genes like *SYP*, *VAMP1*, *SNAP23*, *SNAP25*, *SNCA* and *GAP43*, leading to abnormal dopaminergic neuron processes associated with schizophrenia (24).

On the other hand, the 149 mapped diseases in HervD Atlas are related to numerous HERVs, with an average of 394 different HERVs for each disease. These associations indicate the importance of HERVs in the study of various diseases. The top 15 diseases with the most references in publications are shown in Figure 2F, indicating their significance in HERV research. For example, multiple sclerosis (MS) stands out as the disease with the most extensive documentation, featuring 165 associations with 59 HERVs from 37 publications. The involvement of HERVs in neurological conditions like MS is well-documented and widely accepted, with support from some scientific literature (58,59). Within the integrated HervD Atlas, strong associations are shown between MS and two specific HERVs belonging to the HERV-W group (ERV1_0001, ERVW-1; ERV1_0003, MSRV-env), comprising 50 and 15 associated items, respectively. These two HERVs are suggested as influential factors in triggering the immuno-pathogenesis

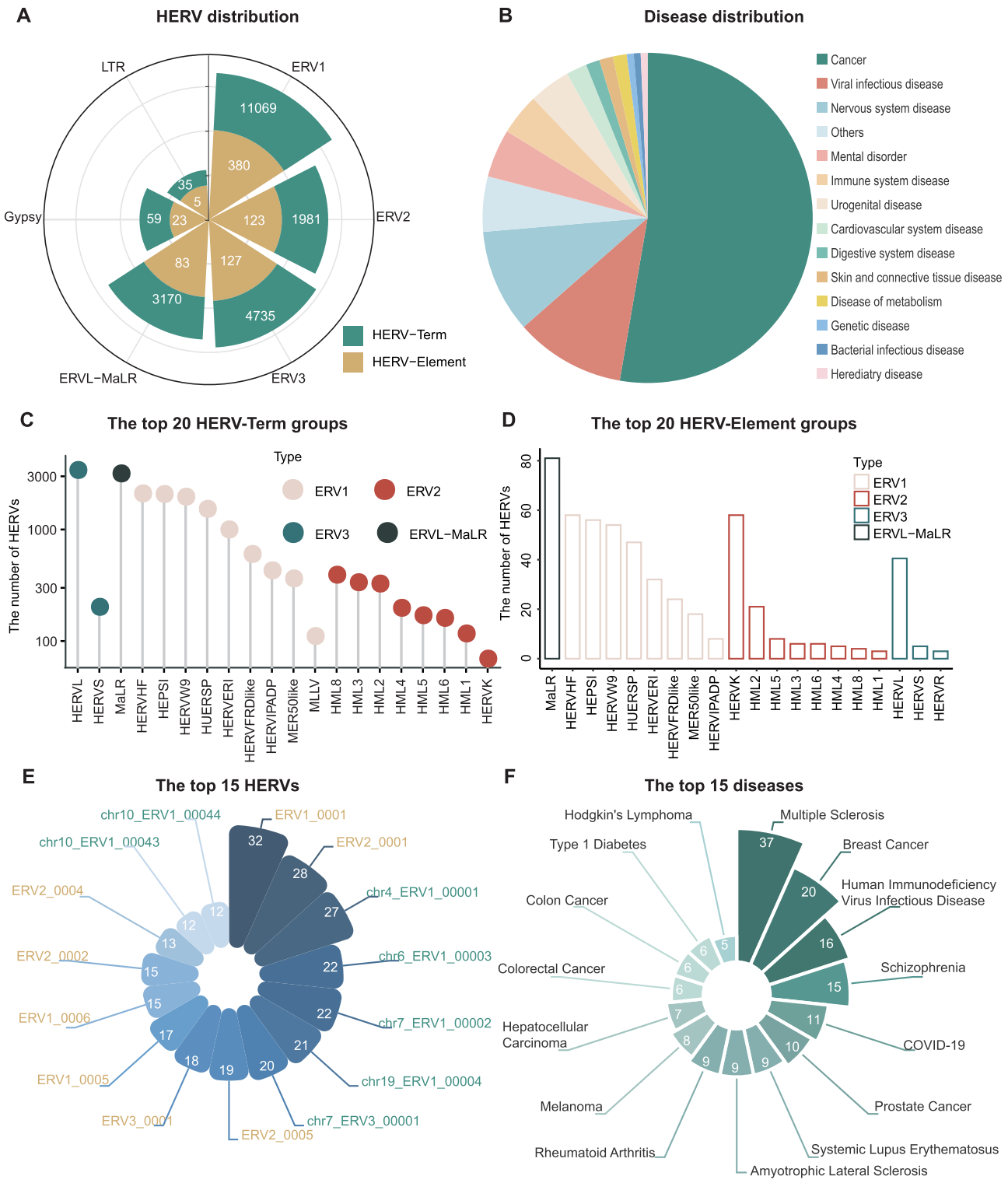


Figure 2. Statistics derived from curated HERV-disease associations. (**A**) The distribution of 21 790 HERVs among the six types. (**B**) The distribution of the diseases grouped into 14 categories. (**C**) The top 20 groups of the HERV-Terms. (**D**) The top 20 groups of the HERV-Elements. (**E**) The top 15 HERVs with the highest number of associated diseases. Fonts in yellow represent HERV-Elements, and those in green represent HERV-Terms. (**F**) The top 15 publication-supported diseases and their corresponding publication number.

of MS, serving as potential prognostic biomarkers for tracking disease progression and therapeutic outcomes (59,60).

User-friendly modules to access the data of interest

The HervD Atlas provides an intuitive 'Browse' page, designed to facilitate easy access the relevant data. This interface features three interactive and easily navigable tables with three main indexes (HERVs, diseases, and publications), allowing users to explore and access the data of interest.

HERVs, as the core objects within the HervD Atlas, are separated into two categories: HERV-Terms, indicating HERVs with precise chromosomal locations, and HERV-Elements, representing groups of HERVs sharing common genes. The browsing table for HERVs offers basic information such as HERV ID, name, type, and brief summary statistics including the number of associations (Figure 3A, upper panel). For more in-depth information, users can access dedicated pages for each specified HERV, which provide comprehensive records of the basic information (including related hyperlinks to external databases such as Dfam, dbHERV-Res), all associations, correlations with genes and expression across GTEx tissues. An example page for 11q23.3_HML2 is illustrated in Figure 3A, bottom panel.

Disease is another core object in the HervD Atlas. Basic information (e.g. disease name, ontology ID and category) and brief summary statistics (e.g. the number of associations and related publications) are listed in the disease browsing table (Figure 3B, upper panel). For more in-depth exploration, each disease has a corresponding detailed page including basic information (with hyperlinks to external databases like EFO and DOID), all related publications and associated HERVs (Figure 3B, bottom panel with an example of Lung Adenocarcinoma). Overall, the browse module of HervD Atlas is designed to offer researchers smooth and efficient access to the desired data through user-friendly interfaces.

HervD Atlas also offers several search channels to ensure users' effortless and efficient querying: (i) a quick search box on the home page allows real-time queries by specifying HERV ID, HERV name or disease name; (ii) an advanced search function on the 'Search' page that allows users to directly access specific terms within HervD Atlas, including information related to HERVs (like ID, type, name, and genomic location), diseases (like name and ontology ID), and publications (such as ID and PubMed ID) and (iii) an intuitive graph search box on the 'Knowledge Graph' page allows search by HERV ID/name and disease name (Figure 3C, D). HervD Atlas also incorporates an auto-suggestion function for assisting users by providing candidate query terms based on even short inputs, enhancing the overall search experience.

HervD Atlas is committed to promoting global usability and accessibility of curated HERV-disease findings. All data within the platform is publicly available, and query results can be conveniently downloaded from the webpage. This ensures ease of data retrieval and exploration for researchers. Additionally, summarized lists of HERVs, diseases, and publications are also available on the 'Download' page.

Systematically integrated knowledge graph with interactive visualization

HervD Atlas systematically integrates HERV-disease associations and related/influenced genes into a comprehensive relationship network, resulting in an interactive knowledge graph. This graph is a powerful tool to uncover HERV networks that influence various diseases and elucidate the complex interactions between disease development and HERVs.

The graph consists of two panels: Disease Network and HERV Network, and generally defines three types of entities: HERVs (21 790 entries: 21 049 HERV-Terms and 741 HERV-Elements), diseases (149) and genes (610). Both panels feature HERV-disease and HERV-gene associations curated from relevant publications. The Disease Network emphasizes diseases as core nodes, while the HERV Network prioritizes HERVs. Connections are quantitatively represented by the thickness of lines, reflecting the number of associations. To accurately and quickly access the content of interest, users can filter networks based on HERV type, group and disease category. This filtering enables a concentrated view of targeted data. Additionally, the graph's draggable nodes allow for customization in the display, and users can export the high-resolution graph for various purposes, including publication.

The knowledge graph's utility can be illustrated through an examination of HERV-K-env (ERV2_0001), the most complete and biologically active HERV known to have been recently acquired by humans (14). By conducting a search for HERV-K-env-related associations, including 72 links with 28 diseases and 44 affected genes, these connections are cohesively presented in the knowledge graph (Figure 4A). Within the graph, HERV-K-env shows relationships with several disease categories, such as cancer, nervous system disorders, viral infections, and skin and connective tissue diseases (Figure 4B). Specifically, a strong correlation between HERV-K-env and breast cancer is revealed (Figure 4C), substantiated by diverse evidence (RNA and protein) and analytical methods (e.g. RNAi, RT-qPCR, Flow cytometry) from six publications. Breast cancer, being a prevalent global health concern among women (30), necessitates novel diagnostic and therapeutic strategies. The HervD Atlas knowledgebase furnishes comprehensive resources detailing the interaction between HERV-K-env and breast cancer. This wealth of information positions HERV-K-env as a potential key to advancing diagnosis and treatment in the future (61).

Beyond the integration and exploration of known knowledge from publications, the knowledge graph facilitates inference into indirect relationships and novel insights grounded on established connections. For example, within the breast cancer network, we also observe a strong connection with HERV-R besides HERV-K-env. Like HERV-K-env, HERV-R (a member of ERV3) has been reported to be related to various diseases (62,63). At present, HERV-K-env and HERV-R have generally been investigated separately (64), and no report has specifically explored their connection in breast cancer, although several studies have paid attention to their co-expression patterns in diseases (63). Future research into the coupled relationship between the two HERVs in breast cancer may lead to discovery new therapeutic targets or vaccine candidates. On the other hand, the knowledge graph highlights genes influenced by HERV-K-env, including *MAPK14*, *STAT2*, *STAT3*, *HSPD1*, *AKT1*, *CASP3*, *CASP8* and *CASP9* (Figure 4D). Among these, *MAPK14*, as a tyrosine phosphorylated protein detected in activated macrophages, is significant in inducing inflammatory cytokines such as TNFα (65) and is pivotal for IFNG-mediated autophagy defense against *Mycobacterium tuberculosis* (66). Other genes like *STAT3* (67), *HSPD1* (68), *AKT1* (69), *CASP3* (70), *CASP8* (71) and *CASP9* (72) have also been reported to play essential roles in tuberculosis

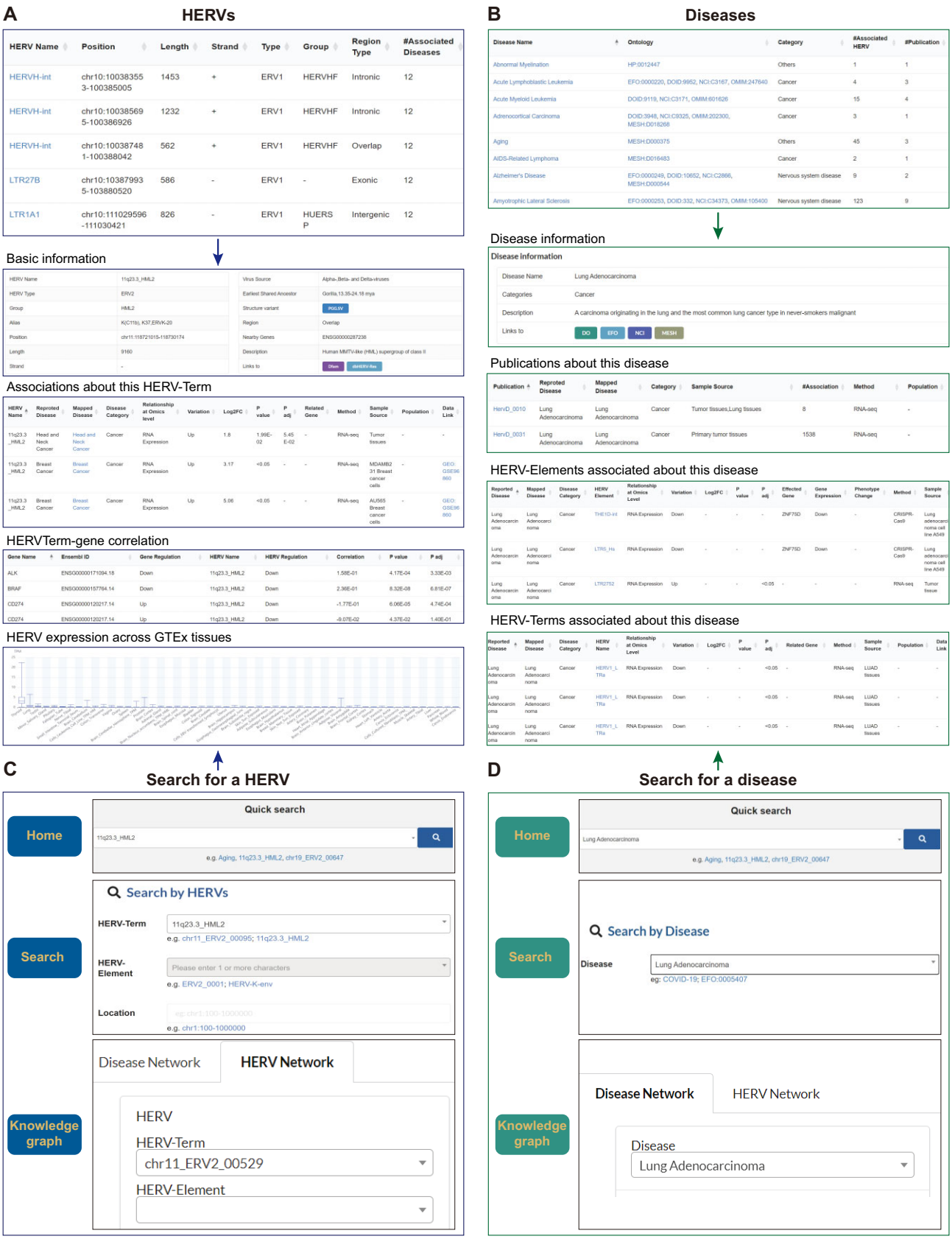


Figure 3. Demonstration of browse and search interfaces in HervD Atlas. (**A**) Screenshot of the browse table of HERVs with fundamental information and summary statistics (above) and detailed HERV information, using 11q23.3_HML2 as an example, including basic information details, relevant associations, related genes, and expression level from GTEx (below). (**B**) Screenshot of the browse table of diseases with basic information and summary statistics (above) and detailed information on 'Lung Adenocarcinoma' used as an example, including disease details, relevant publications and associations (below). (C, D) Three search channels for one specific HERV (**C**) or disease (**D**), respectively.

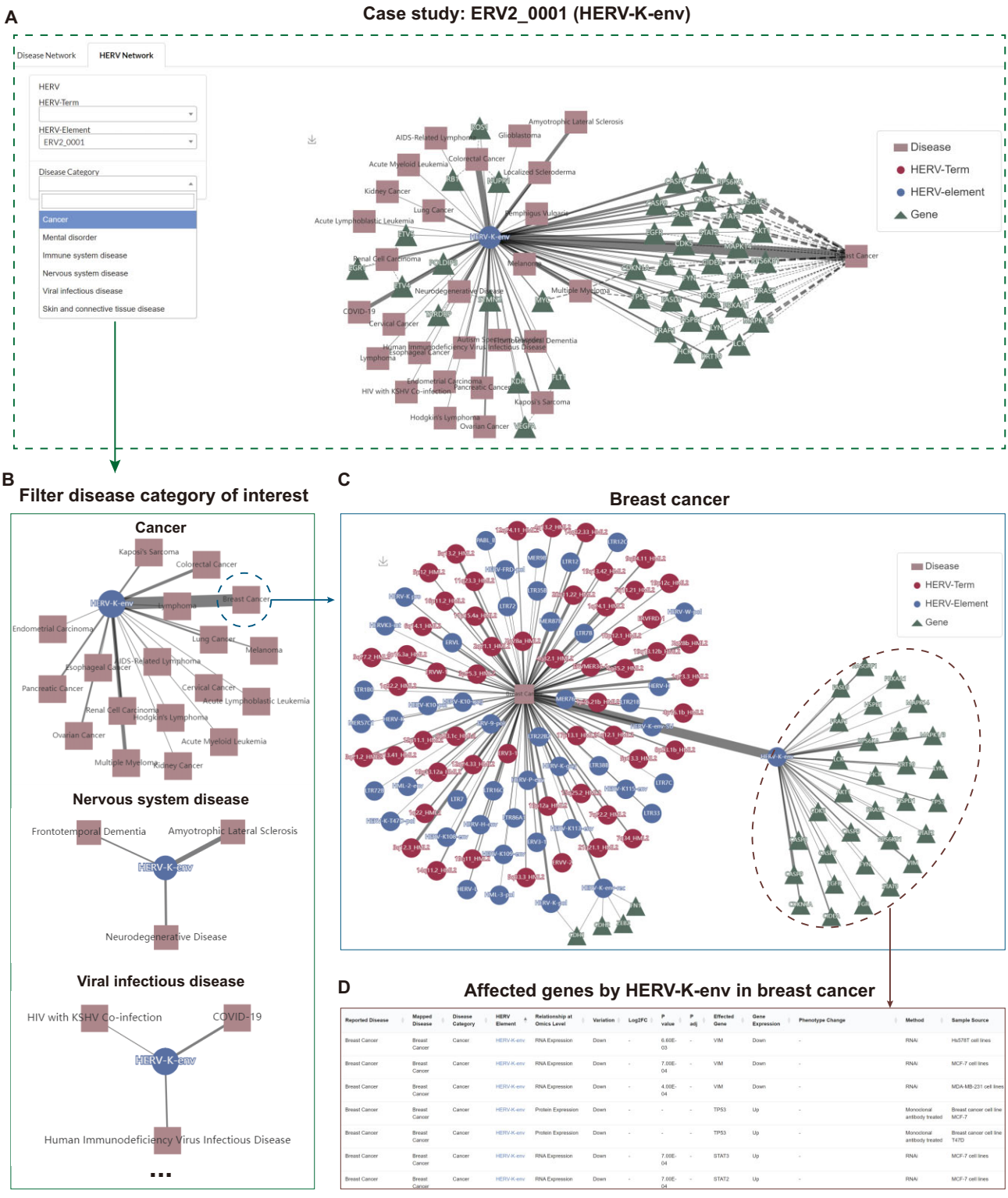


Figure 4. An application case of the knowledge graph in HervD Atlas. (**A**) Knowledge graph centered on the specific HERV, ERV2_0001 (HERV-K-env). (**B**) The knowledge graph centered on the specific HERV, ERV2_0001 (HERV-K-env) with filtered disease categories of cancer, nervous system disease and viral infectious disease. (**C**) The knowledge graph centered on a specific disease, breast cancer. (**D**) Genes affected by ERV2_0001 (HERV-K-env) in breast cancer.

immunity. Although the role of HERV-K in anti-tuberculosis infection remains unexplored, our findings suggest a promising research field for developing host-targeted therapy strategies against tuberculosis.

Overall, the knowledge graph in HervD Atlas offers more than a visually efficient way to access reliable HERV-disease associations; it also provides a valuable resource for researchers to explore and reason from existing knowledge. The knowledge graph fosters the potential for advancements in diagnostics, treatments, and vaccine development by presenting clues and new perspectives on various diseases. This dynamic tool bridges the gap between existing research and future insights, enabling rapid understanding of the field while promoting explorative avenues for novel disease understanding and therapeutic interventions.

Discussion and future developments

Recently, HERVs have been recognized as significant contributors to various pathological conditions in humans, drawing considerable attention for their potential roles in disease. Although numerous important HERV-disease associations have been uncovered, a comprehensive platform to accommodate and integrate these findings still needs to be developed. To bridge this gap, we present the HervD Atlas. To our knowledge, this is the first knowledgebase that systematically collects, curates, and integrates published HERV-disease associations, which is further displayed in an intuitive, visualized, and interactive knowledge graph.

Compared with the existing HERV-associated databases, HervD Atlas mainly features: (i) *Manual collection and curation*: For the first time, HervD Atlas manually collects high-quality data concerning HERV-disease associations from various scientific publications. (ii) *Unified nomenclature and integration*: A unified naming system for these curated HERVs has been implemented, including details like genomic locations, HERV types, and serial numbers. Related information from publications and other external databases, such as earliest shared ancestors, aliases, and structural variants, has also been integrated. This organization aids users in quickly overviewing the HERVs of interest and associated diseases. Additionally, we have adopted an ontology mapping and classification system for diseases. (iii) *Knowledge graph construction*: The knowledge graph is constructed based on the curated associations for the collected HERVs, diseases, and related/affected genes, encompassing both a HERV-network and a disease-network. This integration and visualization enable users to browse, summarize, download, and reuse the associations, enhancing the readability and applicability of the HERV-disease data.

HervD Atlas presents an information-intensive and highly interconnected knowledge graph encompassing various HERVs and their associations with diverse diseases, representing the comprehensive integration of worldwide HERV-related findings. As one vital resource within the National Genomics Data Center (NGDC, https://ngdc.cncb.ac.cn), HervD Atlas will be periodically updated to incorporate the latest HERV-disease association discoveries. Looking ahead, we anticipate the inclusion of numerous HERV-disease associations at the single-cell level, driven by the advent of single-cell studies and corresponding analysis tools (73–75). Such information will be integrated into future versions of the knowledgebase. We will also explore the interaction of HERVs with endogenous microbes in the human body, such as gut microbes, in light of recent findings illustrating how HERVs can influence the intestinal microenvironment and affect gut health (76,77). We also plan to incorporate information on the artificial modification of HERVs through synthetic biology and other methods, such as the packaging of HERVs into viral particles and resurrecting them (23,78–81). Additionally, it's important to highlight that HERVs have been shown to have significant functions as gene regulatory elements. They can act as enhancers, promoters, and insulators, affecting the activity of nearby genes (13,37–39). This valuable insight will be integrated into future versions of the knowledgebase.

In conclusion, HervD Atlas will contribute significantly to advancing our understanding of HERVs and their implications for human health and disease. Its scope, accessibility, and methodologies make it a promising tool for researchers and clinicians, fostering new insights and applications in this emerging field.

Data availability

HervD Atlas is a curated knowledge database of HERV-disease association studies at https://ngdc.cncb.ac.cn/hervd/.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank a number of users for reporting bugs and providing suggestions.

Funding

National Key Research Program of China [2022YFC2602302]; Strategic Priority Research Program of the Chinese Academy of Sciences [XDB38030400]; National Natural Science Foundation of China [32170669]. Funding for open access charge: Funds for International Cooperation and Exchange of the National Natural Science Foundation of China [32061143024].

Conflict of interest statement

None declared.

References

1. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
2. Bannert,N. and Kurth,R. (2006) The evolutionary dynamics of human endogenous retroviral families. *Annu. Rev. Genomics Hum. Genet.*, **7**, 149–173.
3. Vargiu,L., Rodriguez-Tome,P., Sperber,G.O., Cadeddu,M., Grandi,N., Blikstad,V., Tramontano,E. and Blomberg,J. (2016) Classification and characterization of human endogenous retroviruses; mosaic forms are common. *Retrovirology*, **13**, 7.
4. Grandi,N. and Tramontano,E. (2018) Human endogenous retroviruses are ancient acquired elements still shaping innate immune responses. *Front. Immunol.*, **9**, 2039.

5. Nelson,P.N., Carnegie,P.R., Martin,J., Davari Ejtehadi,H., Hooley,P., Roden,D., Rowland-Jones,S., Warren,P., Astley,J. and Murray,P.G. (2003) Demystified. Human endogenous retroviruses. *Mol. Pathol.*, **56**, 11–18.
6. Gorbunova,V., Seluanov,A., Mita,P., McKerrow,W., Fenyo,D., Boeke,J.D., Linker,S.B., Gage,F.H., Kreiling,J.A., Petrashen,A.P., *et al.* (2021) The role of retrotransposable elements in ageing and age-associated diseases. *Nature*, **596**, 43–53.
7. Kim,H.S. (2012) Genomic impact, chromosomal distribution and transcriptional regulation of HERV elements. *Mol. Cells*, **33**, 539–544.
8. Bakoulis,S., Krautz,R., Alcaraz,N., Salvatore,M. and Andersson,R. (2022) Endogenous retroviruses co-opted as divergently transcribed regulatory elements shape the regulatory landscape of embryonic stem cells. *Nucleic Acids Res.*, **50**, 2111–2127.
9. Ku,Y., Park,J.H., Cho,R., Lee,Y., Park,H.M., Kim,M., Hur,K., Byun,S.Y., Liu,J., Lee,Y.S., *et al.* (2021) Noncanonical immune response to the inhibition of DNA methylation by Staufen1 via stabilization of endogenous retrovirus RNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2016289118.
10. Lu,X., Sachs,F., Ramsay,L., Jacques,P.E., Goke,J., Bourque,G. and Ng,H.H. (2014) The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat. Struct. Mol. Biol.*, **21**, 423–425.
11. Kunarso,G., Chia,N.Y., Jeyakani,J., Hwang,C., Lu,X., Chan,Y.S., Ng,H.H. and Bourque,G. (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.*, **42**, 631–634.
12. Bannert,N., Hofmann,H., Block,A. and Hohn,O. (2018) HERVs new role in cancer: from accused perpetrators to cheerful protectors. *Front. Microbiol.*, **9**, 178.
13. Suntsova,M., Garazha,A., Ivanova,A., Kaminsky,D., Zhavoronkov,A. and Buzdin,A. (2015) Molecular functions of human endogenous retroviruses in health and disease. *Cell. Mol. Life Sci.*, **72**, 3653–3675.
14. Golkaram,M., Salmans,M.L., Kaplan,S., Vijayaraghavan,R., Martins,M., Khan,N., Garbutt,C., Wise,A., Yao,J., Casimiro,S., *et al.* (2021) HERVs establish a distinct molecular subtype in stage II/III colorectal cancer with poor outcome. *NPJ Genom Med*, **6**, 13.
15. Hosseiniporgham,S. and Sechi,L.A. (2023) Anti-HERV-K Drugs and Vaccines, Possible Therapies against Tumors. *Vaccines (Basel)*, **11**, 751.
16. Deniz,O., Ahmed,M., Todd,C.D., Rio-Machin,A., Dawson,M.A. and Branco,M.R. (2020) Endogenous retroviruses are a source of enhancers with oncogenic potential in acute myeloid leukaemia. *Nat. Commun.*, **11**, 3506.
17. Yi,J.M., Kim,H.M. and Kim,H.S. (2004) Expression of the human endogenous retrovirus HERV-W family in various human tissues and cancer cells. *J. Gen. Virol.*, **85**, 1203–1210.
18. Zhou,Y., Liu,L., Liu,Y., Zhou,P., Yan,Q., Yu,H., Chen,X. and Zhu,F. (2021) Implication of human endogenous retrovirus W family envelope in hepatocellular carcinoma promotes MEK/ERK-mediated metastatic invasiveness and doxorubicin resistance. *Cell Death Discov*, **7**, 177.
19. Grandi,N., Erbi,M.C., Scognamiglio,S. and Tramontano,E. (2023) Human endogenous retrovirus (HERV) transcriptome is dynamically modulated during SARS-CoV-2 infection and allows discrimination of COVID-19 clinical stages. *Microbiol. Spectr.*, **11**, e0251622.
20. van der Kuyl,A.C. (2012) HIV infection and HERV expression: a review. *Retrovirology*, **9**, 6.
21. Liu,C., Liu,L., Wang,X., Liu,Y., Wang,M. and Zhu,F. (2017) HBV X Protein induces overexpression of HERV-W env through NF-kappaB in HepG2 cells. *Virus Genes*, **53**, 797–806.
22. Arancio,W. (2019) Progerin expression induces a significant downregulation of transcription from human repetitive sequences in iPSC-derived dopaminergic neurons. *Geroscience*, **41**, 39–49.
23. Liu,X., Liu,Z., Wu,Z., Ren,J., Fan,Y., Sun,L., Cao,G., Niu,Y., Zhang,B., Ji,Q., *et al.* (2023) Resurrection of endogenous retroviruses during aging reinforces senescence. *Cell*, **186**, 287–304.
24. Yan,Q., Wu,X., Zhou,P., Zhou,Y., Li,X., Liu,Z., Tan,H., Yao,W., Xia,Y. and Zhu,F. (2022) HERV-W envelope triggers abnormal dopaminergic neuron process through DRD2/PP2A/AKT1/GSK3 for schizophrenia risk. *Viruses*, **14**, 145.
25. Frank,O., Giehl,M., Zheng,C., Hehlmann,R., Leib-Mosch,C. and Seifarth,W. (2005) Human endogenous retrovirus expression profiles in samples from brains of patients with schizophrenia and bipolar disorders. *J. Virol.*, **79**, 10890–10901.
26. Krzysztalowska-Wawrzyniak,M., Ostanek,M., Clark,J., Binczak-Kuleta,A., Ostanek,L., Kaczmarczyk,M., Loniewska,B., Wyrwicz,L.S., Brzosko,M. and Ciechanowicz,A. (2011) The distribution of human endogenous retrovirus K-113 in health and autoimmune diseases in Poland. *Rheumatology (Oxford)*, **50**, 1310–1314.
27. Balada,E., Ordi-Ros,J. and Vilardell-Tarres,M. (2009) Molecular mechanisms mediated by human endogenous retroviruses (HERVs) in autoimmunity. *Rev. Med. Virol.*, **19**, 273–286.
28. Dembny,P., Newman,A.G., Singh,M., Hinz,M., Szczepek,M., Kruger,C., Adalbert,R., Dzaye,O., Trimbuch,T., Wallach,T., *et al.* (2020) Human endogenous retrovirus HERV-K(HML-2) RNA causes neurodegeneration through Toll-like receptors. *JCI Insight*, **5**, e131093.
29. Hartung,H.P., Derfuss,T., Cree,B.A., Sormani,M.P., Selmaj,K., Stutters,J., Prados,F., MacManus,D., Schneble,H.M., Lambert,E., *et al.* (2022) Efficacy and safety of temelimab in multiple sclerosis: results of a randomized phase 2b and extension study. *Mult. Scler.*, **28**, 429–440.
30. Stricker,E., Peckham-Gregory,E.C. and Scheurer,M.E. (2023) HERVs and cancer-a comprehensive review of the relationship of human endogenous retroviruses and human cancers. *Biomedicines*, **11**, 936.
31. Grabski,D.F., Hu,Y., Sharma,M. and Rasmussen,S.K. (2019) Close to the bedside: a systematic review of endogenous retroviruses and their impact in oncology. *J. Surg. Res.*, **240**, 145–155.
32. Pisano,M.P., Grandi,N. and Tramontano,E. (2020) High-throughput sequencing is a crucial tool to investigate the contribution of human endogenous Retroviruses (HERVs) to human biology and development. *Viruses*, **12**, 633.
33. Paces,J., Pavlicek,A. and Paces,V. (2002) HERVd: database of human endogenous retroviruses. *Nucleic Acids Res.*, **30**, 205–206.
34. Paces,J., Pavlicek,A., Zika,R., Kapitonov,V.V., Jurka,J. and Paces,V. (2004) HERVd: the Human Endogenous RetroViruses Database: update. *Nucleic Acids Res.*, **32**, D50.
35. Nakagawa,S. and Takahashi,M.U. (2016) gEVE: a genome-based endogenous viral element database provides comprehensive viral protein-coding sequences in mammalian genomes. *Database (Oxford)*, **2016**, baw087.
36. Tongyoo,P., Avihingsanon,Y., Prom-On,S., Mutirangura,A., Mhuantong,W. and Hirankarn,N. (2017) EnHERV: enrichment analysis of specific human endogenous retrovirus patterns and their neighboring genes. *PLoS One*, **12**, e0177119.
37. Ito,J., Sugimoto,R., Nakaoka,H., Yamada,S., Kimura,T., Hayano,T. and Inoue,I. (2017) Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. *PLoS Genet.*, **13**, e1006883.
38. Kim,T.H., Jeon,Y.J., Kim,W.Y. and Kim,H.S. (2005) HESAS: hERVs expression and structure analysis system. *Bioinformatics*, **21**, 1699–1700.
39. Levy,A., Sela,N. and Ast,G. (2008) TranspoGene and microTranspoGene: transposed elements influence on the transcriptome of seven vertebrates and invertebrates. *Nucleic Acids Res.*, **36**, D47–D52.
40. GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318–1330.

41. Burn,A., Roy,F., Freeman,M. and Coffin,J.M. (2022) Widespread expression of the ancient HERV-K (HML-2) provirus group in normal human tissues. *PLoS Biol.*, **20**, e3001826.
42. Storer,J., Hubley,R., Rosen,J., Wheeler,T.J. and Smit,A.F. (2021) The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob DNA*, **12**, 2.
43. Bao,W., Kojima,K.K. and Kohany,O. (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*, **6**, 11.
44. Benson,D.A., Cavanaugh,M., Clark,K., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Sayers,E.W. (2012) GenBank. *Nucleic Acids Res.*, **41**, D36–D42.
45. Tokuyama,M., Kong,Y., Song,E., Jayewickreme,T., Kang,I. and Iwasaki,A. (2018) ERVmap analysis reveals genome-wide transcription of human endogenous retroviruses. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 12565–12572.
46. Subramanian,R.P., Wildschutte,J.H., Russo,C. and Coffin,J.M. (2011) Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology*, **8**, 90.
47. Glinsky,G.V. (2022) Molecular diversity and phenotypic pleiotropy of ancient genomic regulatory loci derived from human endogenous retrovirus type H (HERVH) promoter LTR7 and HERVK promoter LTR5_Hs and their contemporary impacts on pathophysiology of Modern Humans. *Mol. Genet. Genomics*, **297**, 1711–1740.
48. Phan,K., He,Y., Fu,Y., Dzamko,N., Bhatia,S., Gold,J., Rowe,D., Ke,Y.D., Ittner,L.M., Hodges,J.R., *et al.* (2021) Pathological manifestation of human endogenous retrovirus K in frontotemporal dementia. *Commun Med (Lond)*, **1**, 60.
49. Grandi,N., Cadeddu,M., Blomberg,J. and 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.
50. Grandi,N., Pisano,M.P., Pessiu,E., Scognamiglio,S. and Tramontano,E. (2021) HERV-K (HML7) integrations in the human genome: comprehensive characterization and comparative analysis in non-human primates. *Biology (Basel)*, **10**, 439.
51. Scognamiglio,S., Grandi,N., Pessiu,E. and Tramontano,E. (2022) Identification, comprehensive characterization, and comparative genomics of the HERV-K (HML8) integrations in the human genome. *Virus Res.*, **323**, 198976.
52. Zahn,J., Kaplan,M.H., Fischer,S., Dai,M., Meng,F., Saha,A.K., Cervantes,P., Chan,S.M., Dube,D., Omenn,G.S., *et al.* (2015) Expansion of a novel endogenous retrovirus throughout the pericentromeres of modern humans. *Genome Biol.*, **16**, 74.
53. Mager,D.L. and Freeman,J.D. (1995) HERV-H endogenous retroviruses: presence in the New World branch but amplification in the Old World primate lineage. *Virology*, **213**, 395–404.
54. Huh,J.W., Kim,T.H., Yi,J.M., Park,E.S., Kim,W.Y., Sin,H.S., Kim,D.S., Min,D.S., Kim,S.S., Kim,C.B., *et al.* (2006) Molecular evolution of the periphilin gene in relation to human endogenous retrovirus m element. *J. Mol. Evol.*, **62**, 730–737.
55. Lee,J.W. and Kim,H.S. (2006) Endogenous retrovirus HERV-I LTR family in primates: sequences, phylogeny, and evolution. *Arch. Virol*, **151**, 1651–1658.
56. Flockerzi,A., Burkhardt,S., Schempp,W., Meese,E. and Mayer,J. (2005) Human endogenous retrovirus HERV-K14 families: status, variants, evolution, and mobilization of other cellular sequences. *J. Virol.*, **79**, 2941–2949.
57. Zhang,C., Gao,Y., Ning,Z., Lu,Y., Zhang,X., Liu,J., Xie,B., Xue,Z., Wang,X., Yuan,K., *et al.* (2019) PGG.SNV: understanding the evolutionary and medical implications of human single nucleotide variations in diverse populations. *Genome Biol.*, **20**, 215.
58. Gruchot,J., Kremer,D. and Kury,P. (2019) Neural cell responses upon exposure to human endogenous retroviruses. *Front. Genet.*, **10**, 655.
59. Dolei,A., Ibba,G., Piu,C. and Serra,C. (2019) Expression of HERV genes as possible biomarker and target in neurodegenerative diseases. *Int. J. Mol. Sci.*, **20**, 3706.
60. Dolei,A. (2018) The aliens inside us: HERV-W endogenous retroviruses and multiple sclerosis. *Mult. Scler.*, **24**, 42–47.
61. Johanning,G.L., Malouf,G.G., Zheng,X., Esteva,F.J., Weinstein,J.N., Wang-Johanning,F. and Su,X. (2017) Expression of human endogenous retrovirus-K is strongly associated with the basal-like breast cancer phenotype. *Sci. Rep.*, **7**, 41960.
62. Rhyu,D.W., Kang,Y.J., Ock,M.S., Eo,J.W., Choi,Y.H., Kim,W.J., Leem,S.H., Yi,J.M., Kim,H.S. and Cha,H.J. (2014) Expression of human endogenous retrovirus env genes in the blood of breast cancer patients. *Int. J. Mol. Sci.*, **15**, 9173–9183.
63. Ko,E.J., Song,K.S., Ock,M.S., Choi,Y.H., Kim,S., Kim,H.S. and Cha,H.J. (2021) Expression profiles of human endogenous retrovirus (HERV)-K and HERV-R Env proteins in various cancers. *BMB Rep*, **54**, 368–373.
64. Andersson,A.C., Venables,P.J., Tonjes,R.R., Scherer,J., Eriksson,L. and Larsson,E. (2002) Developmental expression of HERV-R (ERV3) and HERV-K in human tissue. *Virology*, **297**, 220–225.
65. Han,J., Lee,J.D., Bibbs,L. and Ulevitch,R.J. (1994) A MAP kinase targeted by endotoxin and hyperosmolarity in mammalian cells. *Science*, **265**, 808–811.
66. Tateosian,N.L., Pellegrini,J.M., Amiano,N.O., Rolandelli,A., Casco,N., Palmero,D.J., Colombo,M.I. and Garcia,V.E. (2017) IL17A augments autophagy in *Mycobacterium tuberculosis*-infected monocytes from patients with active tuberculosis in association with the severity of the disease. *Autophagy*, **13**, 1191–1204.
67. Queval,C.J., Song,O.R., Deboosere,N., Delorme,V., Debrie,A.S., Iantomasi,R., Veyron-Churlet,R., Jouny,S., Redhage,K., Deloison,G., *et al.* (2016) STAT3 represses nitric oxide synthesis in human macrophages upon *Mycobacterium tuberculosis* infection. *Sci. Rep.*, **6**, 29297.
68. Hadifar,S., Mostafaei,S., Behrouzi,A., Fateh,A., Riahi,P., Siadat,S.D. and Vaziri,F. (2021) Strain-specific behavior of *Mycobacterium tuberculosis* in A549 lung cancer cell line. *BMC Bioinf.*, **22**, 154.
69. Wang,X., Cao,Z., Jiang,J., Zhu,Y., Dong,M., Tong,A. and Cheng,X. (2010) AKT1 polymorphisms are associated with tuberculosis in the Chinese population. *Int. J. Immunogenet.*, **37**, 97–101.
70. Adekambi,T., Ibegbu,C.C., Cagle,S., Ray,S.M. and Rengarajan,J. (2018) High frequencies of caspase-3 expressing *Mycobacterium tuberculosis*-specific CD4(+) T cells are associated with active tuberculosis. *Front. Immunol.*, **9**, 1481.
71. Stutz,M.D., Allison,C.C., Ojaimi,S., Preston,S.P., Doerflinger,M., Arandjelovic,P., Whitehead,L., Bader,S.M., Batey,D., Asselin-Labat,M.L., *et al.* (2021) Macrophage and neutrophil death programs differentially confer resistance to tuberculosis. *Immunity*, **54**, 1758–1771.
72. Ramon-Luing,L.A., Olvera,Y., Flores-Gonzalez,J., Palacios,Y., Carranza,C., Aguilar-Duran,Y., Vargas,M.A., Gutierrez,N., Medina-Quero,K. and Chavez-Galan,L. (2022) Diverse cell death mechanisms are simultaneously activated in macrophages infected by virulent *Mycobacterium tuberculosis*. *Pathogens*, **11**, 492.
73. Wang,J., Ren,M., Yu,J., Hu,M., Wang,X., Ma,W., Jiang,X. and Cui,J. (2022) Single-cell RNA sequencing highlights the functional role of human endogenous retroviruses in gallbladder cancer. *EBioMedicine*, **85**, 104319.
74. Li,C., Zhang,Y., Leng,L., Pan,X., Zhao,D., Li,X., Huang,J., Bolund,L., Lin,G., Luo,Y., *et al.* (2022) The single-cell expression profile of transposable elements and transcription factors in human early biparental and uniparental embryonic development. *Front. Cell Dev. Biol.*, **10**, 1020490.
75. Shao,W. and Wang,T. (2021) Transcript assembly improves expression quantification of transposable elements in single-cell RNA-seq data. *Genome Res.*, **31**, 88–100.

76. Lima-Junior,D.S., Krishnamurthy,S.R., Bouladoux,N., Collins,N., Han,S.J., Chen,E.Y., Constantinides,M.G., Link,V.M., Lim,A.I., Enamorado,M., *et al.* (2021) Endogenous retroviruses promote homeostatic and inflammatory responses to the microbiota. *Cell*, **184**, 3794–3811.
77. Bergallo,M., Galliano,I., Montanari,P., Zaniol,E., Graziano,E., Calvi,C., Alliaudi,C., Dapra,V. and Savino,F. (2020) Modulation of human endogenous retroviruses -H, -W and -K transcription by microbes. *Microbes Infect.*, **22**, 366–370.
78. Contreras-Galindo,R., Kaplan,M.H., Dube,D., Gonzalez-Hernandez,M.J., Chan,S., Meng,F., Dai,M., Omenn,G.S., Gitlin,S.D. and Markovitz,D.M. (2015) Human Endogenous Retrovirus Type K (HERV-K) Particles Package and Transmit HERV-K-Related Sequences. *J. Virol.*, **89**, 7187–7201.
79. Duroy,P.O., Bosshard,S., Schmid-Siegert,E., Neuenschwander,S., Arib,G., Lemercier,P., Masternak,J., Roesch,L., Buron,F., Girod,P.A., *et al.* (2020) Characterization and mutagenesis of Chinese hamster ovary cells endogenous retroviruses to inactivate viral particle release. *Biotechnol. Bioeng.*, **117**, 466–485.
80. Lee,Y.N. and Bieniasz,P.D. (2007) Reconstitution of an infectious human endogenous retrovirus. *PLoS Pathog.*, **3**, e10.
81. Boller,K., Schonfeld,K., Lischer,S., Fischer,N., Hoffmann,A., Kurth,R. and Tonjes,R.R. (2008) Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles. *J. Gen. Virol.*, **89**, 567–572.

Received: August 14, 2023. **Revised:** October 3, 2023. **Editorial Decision:** October 4, 2023. **Accepted:** October 6, 2023