Uncovering the link between human endogenous retroviruses, inflammatory pathways, and gastric cancer development

Zhengtai Li, Hongjiang Li, Kun Fang, Xinglei Lin and Changyuan Yu*
College of Life Science and Technology, Beijing University of Chemical Technology, Beijing, China

Abstract.
BACKGROUND: Endogenous retroviruses, previously deemed "junk" DNA, have gained attention in recent scientific studies. These inherited genomic elements are now recognized for their potential roles in diseases, especially cancer, highlighting their value as potential diagnostic or therapeutic targets.
OBJECTIVE: This research aims to explore the association between human endogenous retroviruses (HERV) and gastric cancer, focusing on discerning HERV expression patterns and understanding their implications in gastric cancer pathology.
METHODS: A quantitative analysis of HERV expression was conducted, employing Support Vector Machine (SVM) and AdaBoost algorithms to identify discriminative HERVs. The co-regulation network between protein-coding genes and HERVs was constructed using the Weighted Gene Co-expression Network Analysis (WGCNA).
RESULTS: Three distinct HERVs (LTR16A|72|451, LTR91|636|874, LTR27D|87|222) were identified as significantly different. Strong correlations were found between HERVs, and gene sets enriched in the inflammatory pathway.
CONCLUSIONS: HERVs appear to influence abnormal inflammatory responses, suggesting a pivotal role in gastric cancer development.

Keywords: Human endogenous retroviruses, machine learning, WGCNA, Gastric cancer

1. Introduction

Human endogenous retroviruses (HERVs) are remnants of ancient retroviral infections and constitute nearly 9% of the human genome [1]. These retroviral elements integrated into the human genome and accumulated mutations, becoming an integral part of the genome. The canonical HERV genome contains three genes (gag, pol, and env) and is flanked by long terminal repeat (LTR) sequences which contain regulatory elements [2]. While most HERVs are usually repressed and considered to be non-functional, they can be triggered by various endogenous and exogenous factors [3].

As a significant health concern, gastric cancer ranked as the sixth most prevalent cancer and the third leading cause of cancer-related fatalities worldwide [4]. In 2020, approximately 1,089,103 new cases were diagnosed, and about 768,793 patients succumbed to the disease [4]. The main recognized causes of gastric cancer include Helicobacter pylori infection [5], various high-risk genetic mutations [6,7], and environmental factors [8].

Recent studies found that aberrant HERV expression may be associated with human cancers [3]. Despite of their potential significance, endogenous viruses have not been extensively studied due to their integration as a part of the genome and epigenetically silenced under routine conditions [9]. Most studies on the relationship

*Corresponding author: Changyuan Yu, College of Life Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China. Tel.: +86 18911295516; E-mail: yucy@mail.buct.edu.cn.

between viruses and tumors have not focused on HERV on human genome. For example, in the study of gastric cancer pathogenesis, much research has focused on the mechanisms of microbial infection, such as Helicobacter pylori or EBV [10], or on the differential expression of protein-coding genes, such as ERBB2 [11] or VEGFR2 [12].

Several studies have investigated the potential association between endogenous retroviruses and cancer development. Yu et al. demonstrated that LTRs of the HERV-W family may cause human uroepithelial cell carcinoma, in which mutant sequences of LTR bind to c-Myb and promoted the expression of Syncytin-1, leading to increased proliferation and viability of human uroepithelial cells [13]. Nakagawa et al. found the expression of ERV3-1 was upregulated in acute myeloid leukemia (AML) patients, suggesting a potential association between HERV and AML [14]. Dolci et al. found that HERV differentially expressed in advanced colon cancer patients and suggested that Env proteins may promote cancer progression [15]. Curty et al. found a significant positive correlation between HERV expression and RAD50 and AC060780.1 in the breast cancer microenvironment, suggesting that HERV may play a role in regulating breast cancer associated genes in HIV carriers [16]. These studies suggest endogenous retroviruses indeed participated in cancer development and worth exploring.

In early studies targeting HERVs as potential biomarkers for tumors, PCR technology was the main method used, focusing on a limited range of HERVs, such as studies on HERV-P, HERV-H, and HERV-K as non-invasive biomarkers for lung cancer [17], and using real-time PCR to study the expression differences of HERV-R, HERV-H, HERV-K, and HERV-P in the blood of breast cancer patients [18]. However, with the gradual reduction in the cost of second-generation sequencing, research on HERVs has entered the realm of large data volumes and multiple types, such as using publicly available datasets of RNA-seq raw data from Alzheimer's disease patients to perform quantitative counting of all HERVs and compare them with healthy controls to find differences in HERVs. It has been found that the expression of HERVs is spread throughout the genome and may affect the normal expression of genes [19].

Our study provides a novel approach to analyze HERV expression in different tumors and explores their potential impact on gene expression. By examining HERV expression in a larger scale and analyzing their correlation with gene expression, we gained a better understanding of the role of HERV in gastric cancer development and progression. The findings of this study may contribute to the development of novel diagnostic and therapeutic strategies.

2. Materials and methods

2.1. Compilation and analysis of transcriptome datasets

Transcriptome sequencing datasets were compiled from various sources, starting with 56 paired samples from gastric cancer patients, obtained from the NCBI database under the accession number PRJNA435914. To further validate the biomarkers identified, additional data from a separate gastric cancer research project (PRJNA1013242), comprising 61 normal and 48 tumor samples (totaling 109 cases), was also included. For lung cancer, datasets from 20 paired samples were retrieved under the project PRJNA344450 [20], providing a comparative analysis across different cancer types. The liver cancer dataset (PRJNA598835 [21]) included 35 tumor and 30 normal samples, offering insights into the expression patterns in hepatic malignancies. Breast cancer research data (PRJEB50623) contributed to 31 tumor and 10 normal samples, enriching the diversity of the dataset. Pancreatic cancer samples (PRJNA490335 [22]), with an equal distribution of 10 tumor and 10 normal samples, and colorectal cancer data (PRJNA288518 [23]), with 5 tumor and 5 normal samples, were also compiled. These datasets served as an external control to validate the specificity and reliability of the biomarker expression levels identified in our study.

2.2. HERVs annotation and differential expression analysis

The various types of HERVs were identified on the human-reference genome (GRch38) using Repeat-Masker software (V4.0.9_p2) and Dfam gene component database [24]. To ensure that the HERV position is unique, the combination of the HERV name and position was adopted as keywords in querying the annotation to display.

In the transcriptome analysis, the HERVs sequences were retrieved and decompressed to fastq, and the Hisat2 (V2.2.1) [25] was adopted to query the HERVs data against the reference genome(GRch38). The Featurecounts (V2.0.1) [26] was used to quantify the
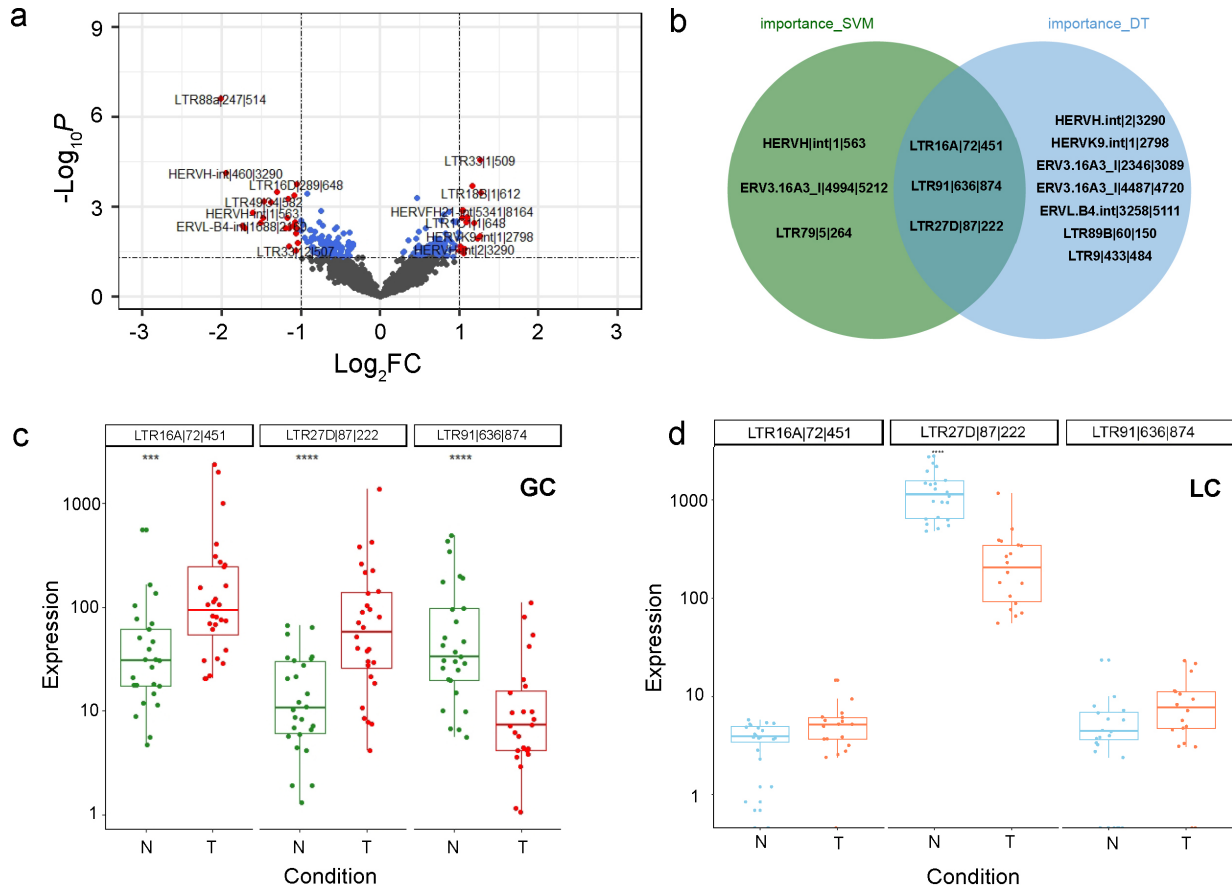
Fig. 1. The differential expression analysis of HEARVs. (a) Differential HERVs in the gastric cancer dataset were screened based on the criteria of $|log2FC| > 1$ and $p < 0.05$. (b) The intersection of the results obtained from the SVM and AdaBoost algorithms. (c) The expression profiles of three potential HERV markers in normal and tumor groups of the gastric cancer dataset. The expression data are presented as log-transformed TPMs. (d) The expression profiles of the three potential HERV markers in normal and tumor groups of lung adenocarcinoma samples. The expression data are presented as log-transformed TPMs.

aligned reads of each protein-coding and non-coding genes using EM algorithm. Based on the count matrix, the differential expression analysis was performed by DESeq2 package (V1.32.0) [27], and a combination of $|log2FC| > 1$ and $P < 0.05$ was used as cutoff to select the significantly differentially expressed genes.

2.3. Identification of HERV biomarkers

In the feature selection process, a combination of SVM [28] and AdaBoost [29] algorithm was employed to develop an efficient, interpretable, and robust statistical model. This method created perturbed samples by randomly altering the sample labels and multivariate screening of variance fractions. The importance of each feature in discriminating cancer group from control groups was then determined and the mean value of each feature was subsequently calculated. By several

cycles of eliminating the least significant feature and progressively reducing the features, a combination with the smallest variance was obtained. This approach enhanced model performance during the feature selection process, capitalizing the strong interpretability of linear SVM and generalization performance of AdaBoost.

2.4. HERVs-human protein coding genes co-expression network analysis

According to the selected HERV expression intensity, the samples were divided into high and low groups, and an correlation analysis between HERV and genes was performed using the WGCNA package (V1.71) [30]. The set of genes with a high correlation with the high and low expression of HERV were compiled and the correlation network was constructed. The pathway processes were analyzed using Clusterprofiler
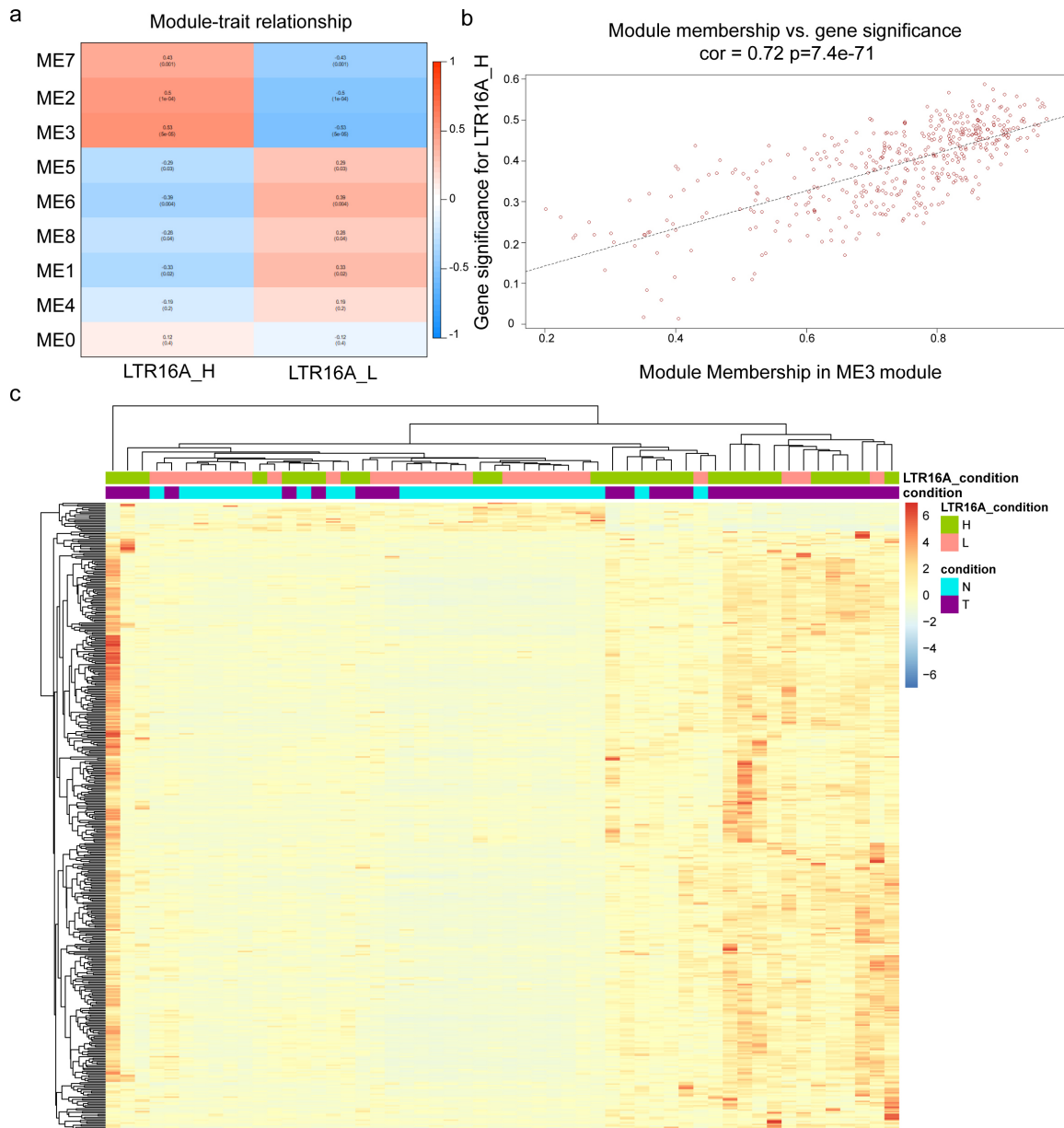
Fig. 2. The WGCNA analysis of protein-coding genes related with HERVs-elevation. (a) the correlation between gene sets identified by WGCNA and LTR16A|72|451 high (LTR16A_H) and low (LTR16A_L) expression status. (b) The correlations between genes and HERVs expression status within ME3 module, the horizontal axis represents correlation coefficients between genes and modules, while the vertical axis represents the correlation coefficient between genes and LTR high expression. (c) The cluster profiles of protein coding genes in the ME3 module in gastric cancer samples.

(V4.0.5) [31] to determine the biological function of the gene sets. Finally, we utilized the chooseTopHubInEachModule function from WGCNA to identify the genes with the highest connectivity within the entire gene set. Additionally, we employed gastric cancer data from TCGA to examine the impact of these genes on survival.

3. Results

3.1. Differential expression of HERVs in gastric cancer and screening of biomarkers

We compiled the transcriptome datasets of gastric cancers and processed the raw expression matrix us-
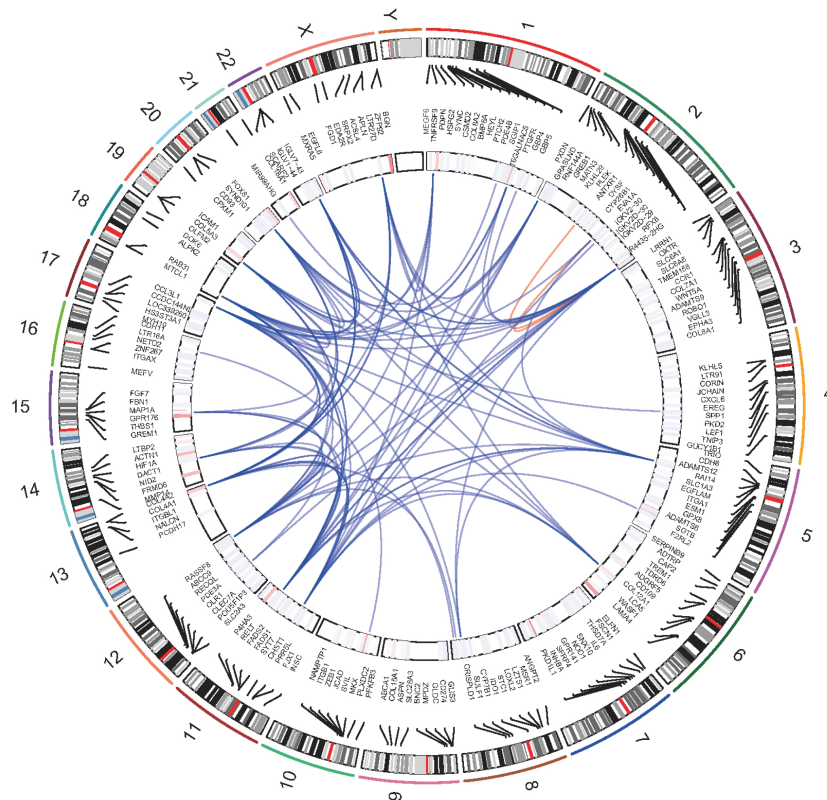
Fig. 3. Genomic positions and the correlation network of protein-coding genes associated with HERVs-elevation, From the outermost ring, the first ring represents chromosomes, the second ring represents the coordinates of genes, the third ring represents the expression status, and the central position connecting lines represent the weight relationships among genes based on WGCNA results. Red lines represent correlations within the same chromosome, blue represents interrelationships between chromosomes.

ing DESeq2. The significantly expressed genes were identified between cancer and normal groups (Fig. 1a), in which 22 HERVs were upregulated and 18 HERVs were downregulated.

To identify potential HERV markers for gastric cancer, we carried out a multivariate screening process using the SVM model and AdaBoost algorithm. The LTR16A|72|451, LTR27D|87|222, and LTR91|636 |874 emerged as the intersections of the two algorithms (Fig. 1b). The original results and explanations of the two algorithms are provided in Supplementary Material 1. The LTR16A|72|451 and LTR27D|87|222 were significant upregulated in tumor subgroups, while LTR91|636|874 was significantly downregulated (Fig. 1c). Conversely, the expression of LTR16A|72|451 and LTR91|636|874 were not significantly different between lung cancer and normal tissues, and LTR27D|87|222 was significantly downregulated in lung cancer groups. This situation is demonstrated in Fig. 1d. Additionally, we validated the expression levels of three potential biomarkers using raw data collected

from various research projects. In the gastric cancer validation set, the expression patterns were completely consistent with those observed in this study. However, in tumors with higher mortality rates, such as breast cancer, liver cancer, colon cancer, and pancreatic cancer, we did not observe any significant correlation. Due to space limitations in the article, we have presented specific expression matrices and box plots in Supplementary Material 2 and 3.

3.2. Co-expression network analysis of HERV marks and protein-coding genes

To further investigate the relationship between HERV biomarkers and gene expression, we used the median expression levels of these three biomarkers as a standard for high and low grouping. Combined with the previous expression level box plots, we found that the biomarker LTR16A|72|451 exhibits significant differences between normal and tumor groups, with a noticeably higher overall expression level than the other two
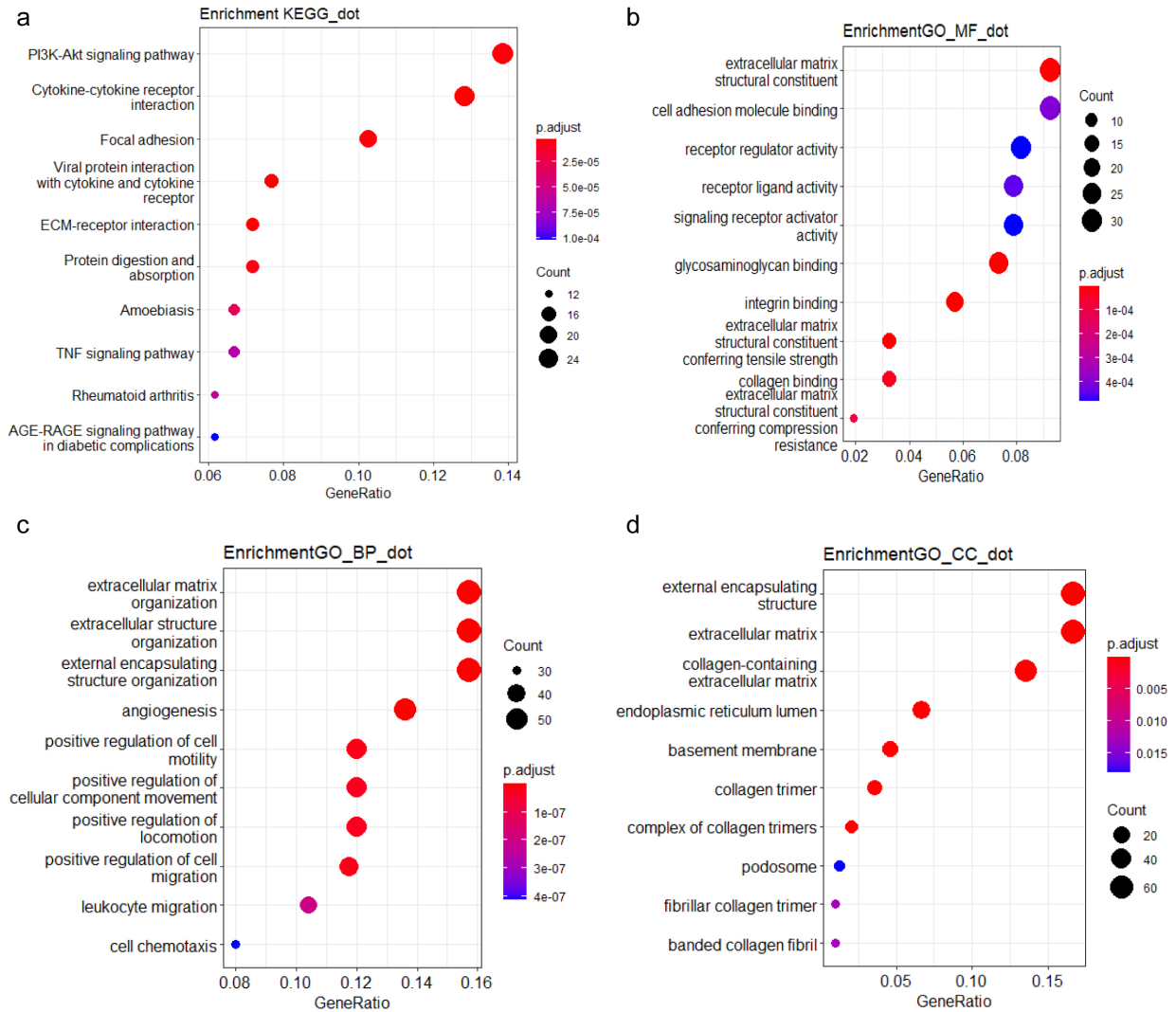
Fig. 4. Functional analysis of HERV-elevation related genes based on (a) KEGG pathway (b) GO molecular function domain (c) GO biological process domain (d) GO cellular component domain.

HERVs, making it more suitable for observing correlations with protein-expressing genes. Therefore, we used this biomarker for high and low grouping to study the correlation between its expression changes and the expression of genes encoding proteins.

Weighted genes co-expression network analysis (WGCNA) analysis was performed to determine the relationship between HERV expression and gene expression in the high and low expression groups, respectively. The Principal Component Analysis (PCA) of the gene expression data within a specific module generated the Module Eigengene (ME) and represented the overall expression characteristics of the genes. In brief, genes with similar expression patterns were grouped

into modules, PCA was then performed on the gene expression data for each module. The first principal component captured the largest portion of variation in the gene expression data and was used as the Module Eigengene. The ME carried the overall expression pattern within the module and was utilized to evaluate the correlation between the module and phenotypic traits, in this way, the gene modules significantly associated with traits could be identified.

We displayed the correlation between the gene sets after clustering and traits in Fig. 2a, where a high correlation between the gene set represented by ME3 and the high expression of LTR16A is clearly visible. Subsequently, the correlation between all genes within the
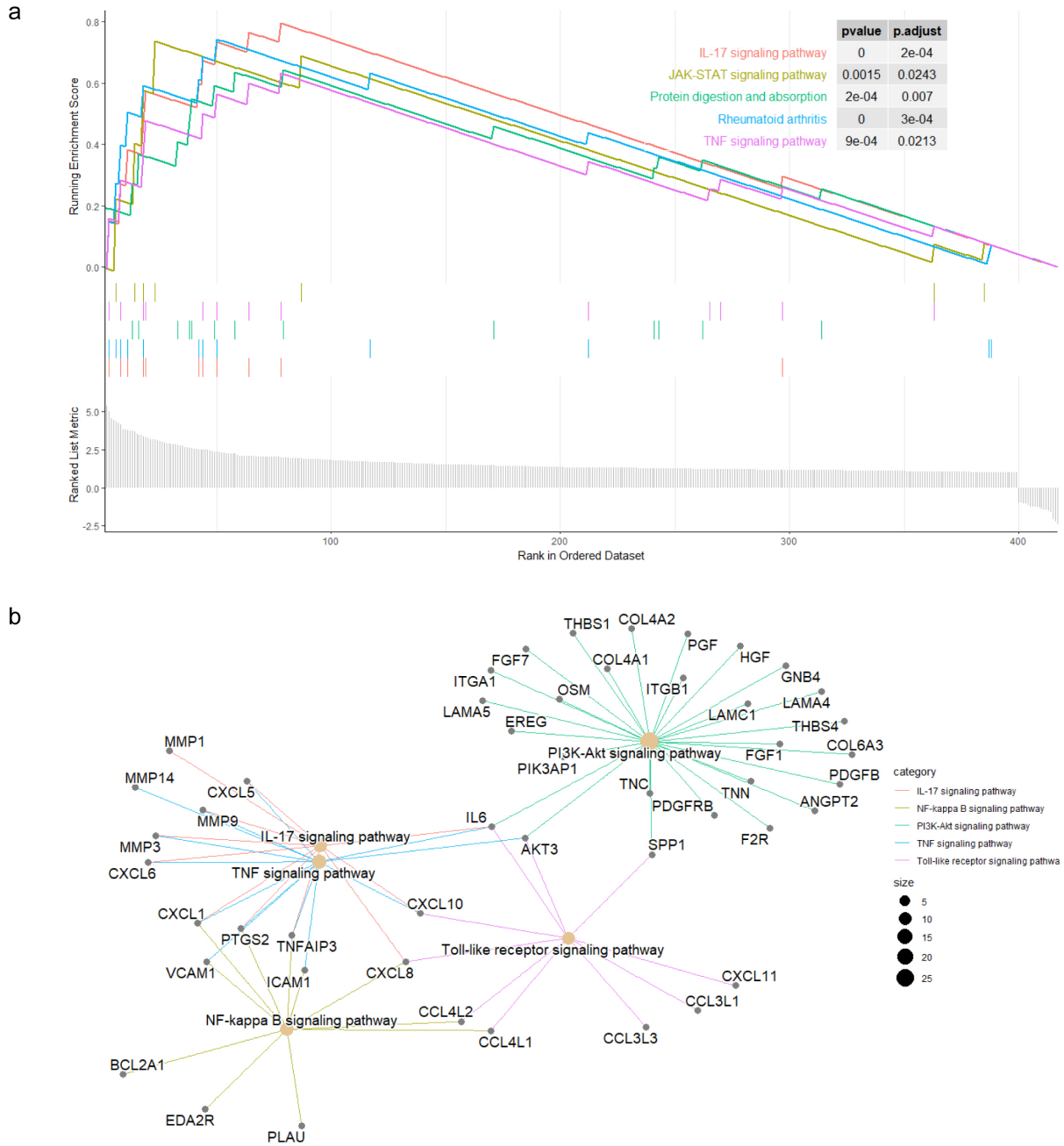
Fig. 5. GSEA of HERV-elevation related genes. (a) Profile of the running ES score and the gene set members on the rank ordered list. (b) The functional relationships among the HERV-elevation related pathways.

ME3 module and the condition of high expression of LTR16A|72|451 is shown in Fig. 2b, revealing a strong positive correlation between them. According to the clustering heatmap of gene expression, genes with the strongest correlation with high HERV expression displayed a notable distinction between normal and tumor groups (Fig. 2c).

According to the WGCNA results, we retrieved the interaction relationships between genes within the corresponding gene set (Fig. 3). We selected the weight greater than 0.4 as cutoff and constructed the gene correlation network. The network degrees were computed, identifying RAB31 as a central hub. This designation of RAB31 as the hub gene was achieved using the 'choose-

TopHubInEachModule' function. This function operates by selecting genes within distinct modules and assessing the correlation among these genes based on their expression data. These correlations are subsequently converted into measures of connectivity strength, which are used to construct an adjacency matrix. Through this process, RAB31 was pinpointed as the gene exhibiting the highest connectivity within its module, establishing its role as a key hub.

3.3. Functional analysis of protein-coding genes correlated with HERV expression elevation

To further explore the biological functions and pathways related by the genes with high correlation to HERV elevation, we performed KEGG and GO analysis on this gene set. The KEGG analysis revealed that the genes associated with the three potential HERV markers were mainly enriched in the PI3K-Akt pathway, which is critical to tumorigenesis (Fig. 4). We also performed Gene Set Enrichment Analysis (GSEA) to score the identified pathways using the foldchange information. The top five pathways with significant enrichment showed an overall upregulation trend. Among them, the IL-17 signaling pathway showed the most significantly enriched and followed by the JAK-STAT and TNF pathway (Fig. 5a).

Subsequently, we investigated the interrelationships of several major pathways in Fig. 5b. The genes associated with the gene set were mainly enriched in the IL-17 pathway, and several major pathways belong to typical inflammatory response pathways [32]. We can clearly see the process starting from the IL17 pathway, affecting TNF through genes such as MMP3 and CXCL1, and then extending to NF-kappa B through genes like CXCL1 and VCAM1. Meanwhile, IL17 also connects to the PI3K-Akt pathway through IL6 and AKT3, broadly affecting more related genes. In this process, we can observe several hallmark tumor pathways that influence and interact with each other, closely linked.

3.4. Survival analysis of key genes RAB31 which correlated with HERV expression elevation

To investigate the functional impact of HERV-related genes on the survival of gastric cancer patients, we identified RAB31 as the hub gene within our selected gene set, proposing it as a candidate prognostic marker. The gene RAB31 in gastric cancer can also be a potential therapeutic target [33,34]. We obtained a gastric can-
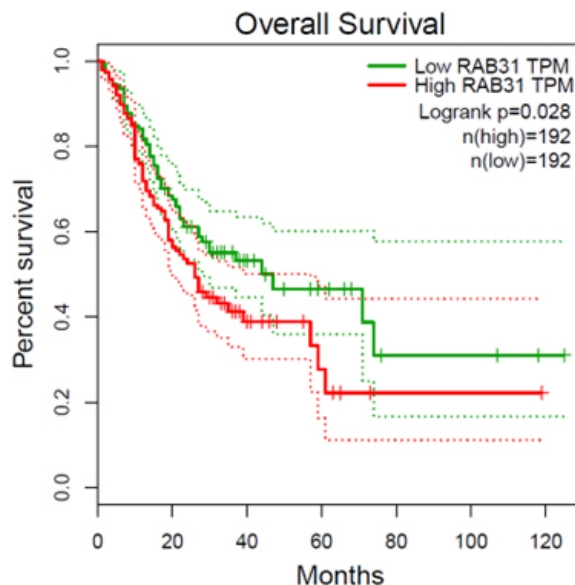

Fig. 6. Survival analysis of the RAB31 gene in TCGA gastric cancer dataset. The patients were classified into RAB31-high and RAB31-low groups according to with median tpm of RAB31 across samples.

cer dataset from TCGA and used the median expression level of RAB31 as the basis for dividing into high and low expression groups. The results of the survival analysis based on these groups were then presented in Fig. 6.

4. Discussion

This study investigated the relationship between differentially expressed HERVs and protein-coding genes in gastric cancer. Our results support that HERVs act as a potential biomarker of gastric cancer.

To investigate the potential role of human endogenous retroviruses (HERV) in gastric cancer, we performed a differential analysis of HERV expression and developed SVM and AdaBoost models for multivariate screening to identify HERVs with major differences between tumor and normal groups. Our screen identified three key markers (LTR16A|72|451, LTR91|636|874, LTR27D|87|222) and demonstrated their significant specificity when compared to lung cancer.

To explore the relationship between HERVs and protein-coding genes, we divided the samples into HERV-high and HERV-low groups according to the median expression tpm of LTR16A|72|451. Using the weighted gene co-expression network analysis (WGCNA), we identified a set of highly expressed

genes that were strongly associated with HERV expression. Furthermore, we conducted pathway enrichment analysis using KEGG and GO to gain insights into the biological functions of these genes and their potential role of HERV in tumor development. Our findings suggest that HERV expression may play a crucial role in tumor development through its association with the expression of specific genes.

We observed that the well-known IL-17 pathway was the most significantly correlated functional modules, with a broad role in activating downstream signaling pathways involved in inflammatory responses, autoimmune diseases, and tumor development [35]. Additionally, the PI3K-Akt pathway showed consistent results with the pathway enrichment analysis [36]. Our analysis of HERV markers showed a significant correlation between inflammatory pathways and their expression changes. We propose that HERV may be activated by the inflammatory response, resulting in a significantly altered expression pattern during the inflammation-to-tumor process. The overexpression of HERV may affect the expression of downstream genes, leading to an overall acceleration of tumor development.

In this study, we utilized the WGCNA approach to analyze the correlation between HERV markers and protein-coding genes and performed a detailed functional analysis of HERV in tumorigenesis. We can predict the future survival status of patients through the changes in the levels of HERV markers in transcriptome sequencing results, and by further identifying the conditions of patients through these highly related genes and pathways, this provides a multi-dimensional assessment basis for the early screening of gastric cancer. The HERV markers and correlated genes may pave the way for experimental verification of their potential roles in tumor development.

5. Conclusion

Our study revealed a significant relationship between differentially expressed HERV and protein-coding genes in gastric cancer. The functional analysis supported that HERVs may be potential biomarkers of gastric cancer. Based on the differential analysis of HERV expression, we developed SVM and AdaBoost models, and identified three key HERV markers (LTR16A|72|451, LTR91|636|874, LTR27D|87|222) with high specificity in gastric cancer. Our analysis further demonstrated a strong association between HERV expression and the expression of Hallmark pathways,

suggesting its underexplored role in tumor development. Pathway analysis highlighted the involvement of IL-17 and PI3K-Akt pathways, indicating a correlation between HERV expression changes and inflammatory pathways in gastric cancer. We also identified RAB31 as a hub gene and revealed its impact on patient survival as a candidate of therapeutic target.

Author contributions

Conception: Zhengtai Li and Changyuan Yu.
Interpretation or analysis of data: Zhengtai Li, Hongjiang Li.
Preparation of the manuscript: Zhengtai Li, Kun Fang and Xinglei Lin.
Revision for important intellectual content: Changyuan Yu.
Supervision: Changyuan Yu.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability

The annotation file used for quantification and the quantification results can be found at the following URL: https://github.com/LiZhengTai2022/HERV_RES.

Supplementary data

The supplementary files are available to download from http://dx.doi.org/10.3233/CBM-230417.

References

[1] E. Alfaro, M. Gamez and N. García, adabag: An R Package for Classification with Boosting and Bagging, *Journal of Sta-*

tistical Software **54** (2013), 1–35.

[2] N. Amatya, A.V. Garg and S.L. Gaffen, IL-17 Signaling: The Yin and the Yang, *Trends Immunol* **38** (2017), 310–322.

[3] M. Bhasin and G.P.S.J.V. Raghava, Prediction of CTL epitopes using QM, *SVM and ANN Techniques* **22** (2004), 0–3204.

[4] V. Blair, I. Martin, D. Shaw, I. Winship, D. Kerr, J. Arnold, P. Harawira, M. McLeod, S. Parry, A. Charlton, M. Findlay, B. Cox, B. Humar, H. More and P. Guilford, Hereditary diffuse gastric cancer: diagnosis and management, *Clin Gastroenterol Hepatol* **4** (2006), 262–275.

[5] J. Bornschein, M. Selgrad, M. Warnecke, D. Kuester, T. Wex and P. Malfertheiner, H. pylori infection is a key risk factor for proximal gastric cancer, *Dig Dis Sci* **55** (2010), 3124–3131.

[6] K. Chen, J. Xu, Y.L. Tong, J.F. Yan, Y. Pan, W.J. Wang, L. Zheng, X.X. Zheng, C. Hu, X. Hu, X. Shen and W. Chen, Rab31 promotes metastasis and cisplatin resistance in stomach adenocarcinoma through Twist1-mediated EMT, *Cell Death Dis* **14** (2023), 115.

[7] G. Curty, G.A. Beckerle, L.P. Iniguez, R.L. Furler, P.S. de Carvalho, J.L. Marston, S. Champiat, J.J. Heymann, C.E. Ormsby, G. Reyes-Teran, M.A. Soares, D.F. Nixon, M.L. Bendall, F.E. Leal and M. de Mulder Rougvie, Human Endogenous Retrovirus Expression Is Upregulated in the Breast Cancer Microenvironment of HIV Infected Women: A Pilot Study, *Front Oncol* **10** (2020), 553983.

[8] T. Dawson, U. Rentia, J. Sanford, C. Cruchaga, J.S.K. Kauwe and K.A. Crandall, Locus specific endogenous retroviral expression associated with Alzheimer's disease, *Frontiers in Aging Neuroscience* **15** (2023).

[9] M. Dolci, C. Favero, W. Toumi, E. Favi, L. Tarantini, L. Signorini, G. Basile, V. Bollati, S. D'Alessandro, P. Bagnoli, P. Ferrante and S. Delbue, Human Endogenous Retroviruses Long Terminal Repeat Methylation, Transcription, and Protein Expression in Human Colon Cancer, *Front Oncol* **10** (2020), 569015.

[10] M.K. Ediriweera, K.H. Tennekoon and S.R. Samarakoon, Role of the PI3K/AKT/mTOR signaling pathway in ovarian cancer: Biological and therapeutic significance, *Semin Cancer Biol* **59** (2019), 147–160.

[11] M. Garcia-Montojo, T. Doucet-O'Hare, L. Henderson and A. Nath, Human endogenous retrovirus-K (HML-2): a comprehensive review, *Crit Rev Microbiol* **44** (2018), 715–738.

[12] S.J. Jung, M.S. Kang, C.H. Lee, S.H. Hong and H.K. Yoon, Immunohistochemical Findings in 10 Cases of Inflammatory Myofibroblastic Tumor, (1999).

[13] G. Kassiotis and J.P. Stoye, Making a virtue of necessity: the pleiotropic role of human endogenous retroviruses in cancer, *Philos Trans R Soc Lond B Biol Sci* **372** (2017).

[14] P. Langfelder and S. Horvath, WGCNA: an R package for weighted correlation network analysis, *BMC Bioinformatics* **9** (2008), 559.

[15] J.R. Lee, C.H. Kwon, Y. Choi, H.J. Park, H.S. Kim, H.J. Jo, N. Oh and Y. Park do, Transcriptome analysis of paired primary colorectal carcinoma and liver metastases reveals fusion transcripts and similar gene expression profiles in primary carcinoma and liver metastases, *BMC Cancer* **16** (2016), 539.

[16] X. Li, R. Bechara, J. Zhao, M.J. McGeachy and S.L. Gaffen, IL-17 receptor-based signaling and implications for disease, *Nat Immunol* **20** (2019), 1594–1602.

[17] Y. Liao, G.K. Smyth and W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics* **30** (2014), 923–930.

[18] J. Lin, Y.J. Wu, X. Liang, M. Ji, H.M. Ying, X.Y. Wang, X. Sun, C.H. Shao, L.X. Zhan and Y. Zhang, Network-based integration of mRNA and miRNA profiles reveals new target genes involved in pancreatic cancer, *Mol Carcinog* **58** (2019), 206–218.

[19] N.I. Lindeman, P.T. Cagle, D.L. Aisner, M.E. Arcila, M.B. Beasley, E.H. Bernicker, C. Colasacco, S. Dacic, F.R. Hirsch, K. Kerr, D.J. Kwiatkowski, M. Ladanyi, J.A. Nowak, L. Sholl, R. Temple-Smolkin, B. Solomon, L.H. Souter, E. Thunnissen, M.S. Tsao, C.B. Ventura, M.W. Wynes and Y. Yatabe, Updated Molecular Testing Guideline for the Selection of Lung Cancer Patients for Treatment With Targeted Tyrosine Kinase Inhibitors: Guideline From the College of American Pathologists, the International Association for the Study of Lung Cancer, and the Association for Molecular Pathology, *J Thorac Oncol* **13** (2018), 323–358.

[20] M.I. Love, W. Huber and S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol* **15** (2014), 550.

[21] N. Lunet, A. Lacerda-Vieira and H. Barros, Fruit and vegetables consumption and gastric cancer: a systematic review and meta-analysis of cohort studies, *Nutr Cancer* **53** (2005), 1–10.

[22] S. Nakagawa, M. Kawashima, Y. Miyatake, K. Kudo, R. Kotaki, K. Ando and A. Kotani, Expression of ERV3-1 in leukocytes of acute myelogenous leukemia patients, *Gene* **773** (2021), 145363.

[23] C. Oliveira, H. Pinheiro, J. Figueiredo, R. Seruca and F. Carneiro, Familial gastric cancer: genetic susceptibility, pathology, and implications for management, *Lancet Oncol* **16** (2015), e60–70.

[24] D.-W. Rhyu, Y.-J. Kang, M.-S. Ock, J.-W. Eo, Y.-H. Choi, W.-J. Kim, S.-H. Leem, J.-M. Yi, H.-S. Kim and H.-J. Cha, Expression of Human Endogenous Retrovirus env Genes in the Blood of Breast Cancer Patients, *International Journal of Molecular Sciences* **15** (2014), 9173–9183.

[25] Z. Salavatiha, R. Soleimani-Jelodar and S. Jalilvand, The role of endogenous retroviruses-K in human cancer, *Rev Med Virol* **30** (2020), 1–13.

[26] J. Siren, N. Valimaki and V. Makinen, Indexing Graphs for Path Queries with Applications in Genome Research, *IEEE/ACM Trans Comput Biol Bioinform* **11** (2014), 375–388.

[27] Z. Sun, L. Wang, B.W. Eckloff, B. Deng, Y. Wang, J.A. Wampfler, J. Jang, E.D. Wieben, J. Jen, M. You and P. Yang, Erratum to: Conserved recurrent gene mutations correlate with pathway deregulation and clinical outcomes of lung adenocarcinoma in never-smokers, *BMC Med Genomics* **10** (2017), 1.

[28] A.P. Thrift and H.B. El-Serag, Burden of Gastric Cancer, *Clin Gastroenterol Hepatol* **18** (2020), 534–542.

[29] E. Van Cutsem, X. Sagaert, B. Topal, K. Haustermans and H. Prenen, Gastric cancer, *Lancet* **388** (2016), 2654–2664.

[30] D. Wei, W. Zhan, Y. Gao, L. Huang, R. Gong, W. Wang, R. Zhang, Y. Wu, S. Gao and T. Kang, RAB31 marks and controls an ESCRT-independent exosome pathway, *Cell Res* **31** (2021), 157–177.

[31] T.J. Wheeler, J. Clements, S.R. Eddy, R. Hubley, T.A. Jones, J. Jurka, A.F. Smit and R.D. Finn, Dfam: a database of repetitive DNA based on profile hidden Markov models, *Nucleic Acids Res* **41** (2013), D70–82.

[32] T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo and G. Yu, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data, *Innovation (Camb)* **2** (2021), 100141.

[33] H. Yu, T. Liu, Z. Zhao, Y. Chen, J. Zeng, S. Liu and F. Zhu, Mutations in 3'-long terminal repeat of HERV-W family in chromosome 7 upregulate syncytin-1 expression in urothelial

cell carcinoma of the bladder through interacting with c-Myb, *Oncogene* **33** (2014), 3947–3958.

[34] M. Zapatka, I. Borozan, D.S. Brewer, M. Iskar, A. Grundhoff, M. Alawi, N. Desai, H. Sultmann, H. Moch, P. Pathogens, C.S. Cooper, R. Eils, V. Ferretti, P. Lichter and P. Consortium, The landscape of viral associations in human cancers, *Nat Genet* **52** (2020), 320–330.

[35] M. Zare, S. Mostafaei, A. Ahmadi, S. Azimzadeh Jamalkandi, A. Abedini, Z. Esfahani-Monfared, R. Dorostkar and M.

Saadati, Human endogenous retrovirus env genes: Potential blood biomarkers in lung cancer, *Microbial Pathogenesis* **115** (2018), 189–193.

[36] Y. Zhu, S. Wang, X. Xi, M. Zhang, X. Liu, W. Tang, P. Cai, S. Xing, P. Bao, Y. Jin, W. Zhao, Y. Chen, H. Zhao, X. Jia, S. Lu, Y. Lu, L. Chen, J. Yin and Z.J. Lu, Integrative analysis of long extracellular RNAs reveals a detection panel of noncoding RNAs for liver cancer, *Theranostics* **11** (2021), 181–193.