ORIGINAL ARTICLE

10.2478/AMB-2026-0071

EXPRESSION OF HUMAN ENDOGENOUS RETROVIRUSES (HERVS) IN BREAST CANCER: A TRANSCRIPTOMIC STUDY

D. Bartzi[1], A. Philippou[2], I. Tsatsou[3], M. Koutsilieris[2]

[1]Hellenic Airforce General Hospital & Researcher, Medical School, University of Crete, Herakleion, Crete
[2]Medical School, National & Kapodistrian University – Athens, Greece
[3]Nursing Department, University of West Attica – Athens, Greece

Abstract. Background: *Human endogenous retroviruses (HERVs) are ancient viral elements integrated into the human genome. HERVs, especially HERV-K (HML-2), play a significant role in the onset and advancement of breast cancer. This study examined the expression of endogenous retroviruses in human breast tissues to reveal potential associations between retroviruses and breast cancer.* ***Materials and methods:*** *RNA-Seq datasets from breast cancer tissues (ductal carcinoma and triple-negative subtypes) and normal breast tissues were retrieved from the NCBI Sequence Read Archive (SRA). Bioinformatics analysis included read mapping against the human genome (hg19) and a custom pseudogenome containing HERV sequences. Normalized expression of HERV-human junctions was calculated, and statistical comparisons were performed using independent samples t-tests and Mann-Whitney tests.* ***Results:*** *Breast cancer tissues exhibited significantly higher normalized coverage of HERV integration sites compared to healthy controls (p = 0.009). Among breast cancer cases, patients who had undergone treatment showed significantly lower HERV expression than untreated patients (p = 0.023). Visualization via IGV confirmed increased read accumulation at HERV loci in cancer samples.* ***Conclusion:*** *HERVs were overexpressed in breast cancer cell lines compared to healthy breast tissue samples, with expression levels influenced by therapeutic interventions. These findings support the potential role of HERVs as biomarkers for breast cancer diagnosis, prognosis, and therapeutic monitoring.*

Key words: *HERVs, breast cancer, transcriptomics, RNA-Seq, biomarkers, gene expression, therapy response*

Corresponding author: *Dimitra Bartzi, Hellenic Airforce General Hospital & Researcher, Medical School, University of Crete, Herakleion, Crete, email: demmy345@hotmail.com*

ORCID: *0009-0002-2575-0548*

Received: *13 December 2025;* ***Revised/Accepted:*** *15 May 2025*

INTRODUCTION

Breast cancer remains a leading cause of cancer-related mortality among women globally [1]. In 2024, it was estimated that approximately 310,720 new cases of invasive breast cancer were diagnosed in women in the United States, along with about 56,500 cases of ductal carcinoma in situ (DCIS). Breast cancer continues to account for around 30% of all new cancer diagnoses in U.S.

women. Mortality remains significant, with an expected 42,250 female deaths and 530 male deaths out of an estimated 2,790 new male cases. Despite these numbers, the five-year relative survival rate for breast cancer diagnosed at a localized stage exceeds 99%. This rate declines to 87% for regional-stage disease and drops further to 32% when diagnosed at a distant, metastatic stage. Encouragingly, breast cancer mortality has declined by 44% since 1989, resulting in over 517,900 averted deaths. However, concerning trends include a rising incidence among younger women under 50 years, increasing at an annual rate of 1.4% between 2012 and 2021 [2]. Significant racial disparities also persist, with Black women experiencing a 40% higher mortality rate compared to White women, largely due to later-stage diagnoses and unequal access to quality care [3].

Tumorigenesis is a multistep process, with oncogenic mutations in a normal cell conferring clonal advantage as the initial event. However, despite pervasive somatic mutations and clonal expansion in normal tissues, their transformation into cancer remains a rare event, indicating the presence of additional driver events for progression to an irreversible, highly heterogeneous, and invasive lesion [4]. Advances in molecular biology and genomics have revolutionized our understanding of tumor biology, yet many mechanisms contributing to breast cancer onset and progression remain poorly understood. Researchers are focusing on encompassing diverse aspects, such as tumor stemness, intra-tumoral microbiota, and circadian rhythms [5]. Among the genomic elements receiving increasing attention are endogenous retroelements – particularly human endogenous retroviruses (HERVs) [6].

HERVs are sequences derived from ancient retroviral infections that have been integrated into the human genome over millions of years. Although most are defective and transcriptionally silent due to mutations and epigenetic repression, certain HERV families, especially human endogenous retrovirus type K [HERV-K (HML-2)], retain the ability to be transcribed under specific physiological and pathological conditions. In cancer biology, aberrant reactivation of HERVs has been reported in several malignancies, including melanoma, testicular cancer, and breast cancer [6-8].

The HERV-K (HML-2) subgroup is the most biologically active and has been implicated in oncogenesis through mechanisms such as insertional mutagenesis, induction of inflammation, and modulation of the immune microenvironment. The transcriptional activity of HERVs in tumor tissues may provide insights into novel oncogenic pathways and serve as a source of tumor-specific antigens [10, 11]. The HERV-K (HML-2) is found to be expressed in various subtypes of breast cancer. This retrovirus was integrated into the primate genome approximately 55 million years ago, presenting a challenge for potential immunotherapy approaches for breast cancer. Recent studies have indicated that several endogenous retroviruses (ERVs) have been reactivated in tumors, with many exhibiting overexpression in cancerous tissues while showing low or undetectable levels in normal tissues [11, 12].

This study focuses on evaluating the differential expression of HERVs in breast cancer versus normal breast tissue using publicly available RNA-Seq datasets. We also assessed the effect of treatment on HERV expression levels, with the ultimate goal of exploring HERVs as potential biomarkers and therapeutic targets in breast cancer.

MATERIALS AND METHODS

Dataset Acquisition

RNA-Seq datasets were obtained from the NCBI Sequence Read Archive (SRA) [13]. The study cohort was divided into two distinct groups to facilitate a comparative analysis: (1) Breast cancer cell lines, consisting of 11 samples representing various subtypes, including ductal carcinoma, adenocarcinoma, and triple-negative breast cancer; and (2) Healthy breast tissue samples, consisting of 6 samples, which served as the control group. All data has been securely downloaded into password-protected directories of the SRA. The SRA Toolkit was obtained through Linux to facilitate access to SRA data. This toolkit enables the creation of next-generation sequencing files in the specified format and cloud storage.

RNA-Seq data for breast ductal carcinoma (BAM files), along with their corresponding clinical information, were sourced from the NCBI SRA. The paired-end FASTQ files for each sample were downloaded in SRA format and subsequently converted to FASTQ format using the fastq-dump command from the SRA Toolkit. Samples included 11 breast cancer cell lines (various subtypes, including ductal carcinoma and triple-negative) and 6 healthy breast tissues.

Bioinformatics Analysis

The paired-end reads in FASTQ format were aligned to the human reference genome, hg19, using bowtie2 with default settings. The paired-end reads in FASTQ format were also aligned to the pseudogenome, which was made using the 5' and 3' LTRs (long terminal repeat) coordinates of HERV-K(HML-2), as described by Subramanian et al. [14]. Mapping to this

pseudogenome allowed for quantification of reads aligning to HERV-host junctions.

The resulting alignments were subsequently converted to the standard BAM file format utilizing Samtools with its default configurations. The expression of HERV-K-human junctions was obtained from raw reads through BedTools Multicov, also employing default settings. Following this, the mapped reads were quantified and normalized using the specified formula:

$$\frac{\text{multicov pseudogenome reads}}{\text{alignment rate \%*reads}}$$

A pseudogenome was constructed using the coordinates of LTRs from HERV-K(HML-2) to identify the expression of LTR-host junctions [15]. The LTR coordinates utilized for the development of the pseudogenome were detailed by Subramanian et al. The incorporation of HERVs into the human genome leads to areas that consist of both partial LTR sequences and partial human sequences, known as LTR-host junctions. Various tools were employed to analyze these datasets, including SRA-toolkit, Samtools, Bowtie, IGV, and BedTools, which is a software specifically designed to quantify the number of reads associated with each gene.

To incorporate the integration sites for the pseudogenome, partial human sequences were utilized, commencing and concluding at a defined distance of bases from the beginning and end of LTRs (the coordinates for the start and end of LTRs are provided in the attached Excel file). Consequently, a Bed file was generated by adjusting the length of reads by 66% to encompass junctions. The formula is outlined below:

chr(1,2,3, etc) LTR start-66% length LTR start+66% length

chr(1,2,3, etc) LTR end-66% length LTR end+66% length

for any LTR of HERV.

Command “bedtools getfasta” [16] was used with default settings, in order to create a fasta file including the LTR-host junction sequences that was used as the pseudogenome. As far as it concerns the alignment of the pseudogenome is concerned, use the command:

“bowtie2 -t -x bowtie2/hg19 ncbi/public/sra/....fastq -S bowtie2/....sam”

Furthermore, downloading SAMtools was used for mapping the pseudogenome via the following commands:

1. ”wget https://github.com/samtools/samtools/releases/download/1.3.1/samtools-1.3.1.tar.bz2”

2. “tar -jvxf samtools-1.3.1.tar.bz2.”

3. ”cd samtools-1.3.1.”

4. “make”

“.sam file” was converted to “.bam file” by means of the commands:

“samtools view -S -b bowtie2/....sam > bowtie2/.... bam”

After that, files were sorted with samtools by means of the commands:

“samtools sort bowtie2/....bam -o bowtie2/..._sorted. bam”

“samtools index bowtie2/..._sorted.bam” [17].

Statistical Analysis

Data analysis was performed in IBM SPSS 24. Chromosomes were represented using frequencies. Normalized coverage of integration sites was calculated using BEDTools and custom scripts. Statistical significance between groups was assessed using independent samples t-tests and Mann-Whitney U tests, with significance set at $p < 0.05$.

RESULTS

RNASeq Datasets

Transcriptome data were examined from 11 patients with invasive ductal carcinoma breast cancer, as detailed in Table 1, alongside six healthy breast tissue samples, presented in Table 2. In Table 1, there are some duplicate HCC38 entries; while the cell line is the same, each SRR ID represents a distinct RNA-Seq dataset retrieved from the SRA.

The 11 breast cancer samples encompass a heterogeneous range of clinical profiles, including Luminal A (MCF7), Luminal B (ZR75-1), and Triple-Negative (HCC38, HCC39, BT549, MDA-MB-231) subtypes.

The inclusion of both primary tumor-derived lines (e.g., HCC38) and those from metastatic sites (e.g., MCF7, MDA-MB-231) ensures that the observed HERV-K (HML-2) overexpression reflects broad pathological states in breast cancer biology. By incorporating both primary and metastatic-derived lines, our study captures a diverse landscape of HERV-K (HML-2) expression relevant to clinical breast cancer biology.

In Tables 3 and 4, the raw reads and overall alignment rate from the alignment of any RNA-seq to the Human genome (hg19) are presented for breast cancer patients and healthy tissue, respectively. Pseudogenome reads came from the alignment of any RNA sequences to the Pseudogenome. Normalized coverage of integration sites:

Table 1. RNA-seq data: Instrument ILUMINA HiSeq 2500, strategy: RNA-seq, source: Transcriptomic, Selection: cDNA, Layout: Single (breast carcinoma)

SRR Experiment	Cell Line	Molecular Subtype	Histological Type	Clinical Origin
SRR8743332	MCF7	Luminal A	Ductal Carcinoma	Metastatic (Pleural Effusion)
SRR10502311	MDA-MB-231	Triple-Negative (Basal)	Adenocarcinoma	Metastatic (Pleural Effusion)
SRR10612305	HCC38	Triple-Negative (Basal)	Ductal Carcinoma	Primary Tumor
SRR10247656	HCC38	Triple-Negative (Basal)	Ductal Carcinoma	Primary Tumor
SRR7509758	HCC38	Triple-Negative (Basal)	Ductal Carcinoma	Primary Tumor
SRR7509733	HCC38	Triple-Negative (Basal)	Ductal Carcinoma	Primary Tumor
SRR7509732	HCC38	Triple-Negative (Basal)	Ductal Carcinoma	Primary Tumor
SRR7509754	HCC39	Triple-Negative (Basal)	Ductal Carcinoma	Primary Tumor
SRR12180250	HCC1428	Luminal	Adenocarcinoma	Metastatic (Pleural Effusion)
SRR7509721	BT549	Triple-Negative (Basal)	Ductal Carcinoma	Primary Tumor
SRR7500823	ZR75-1	Luminal B	Ductal Carcinoma	Metastatic (Ascites)

Table 2. RNA-seq data: Instrument: Next seq 500, strategy: RNA-seq, source: Transcriptomic, Selection: PolyA, Layout: Single, (healthy breast tissue sample)

SRR Experiment	Cell_Line	Disease_state
SRR7687788	Normal left Stroma	normal breast tissue
SRR7687782	Normal left epi	normal breast tissue
SRR7687798	Normal left Stroma	normal breast tissue
SRR7687780	Normal left Stroma	normal breast tissue
SRR7687781	Normal left epi	normal breast tissue
SRR7687797	Normal right epi	normal breast tissue

$$\frac{\text{multicov pseudogenome reads}}{\text{alignment rate \%*reads}}$$

Table 3. Raw reads and overall alignment rate came from alignment of any RNA-seq to the human genome (hg19)

Breast cancer	Treatment	Alignment Rate (%)	Raw Reads	Pseudogenome Reads	Normalized coverage of integration sites
SRR8743332 ductal	Yes	93.13	26333636	160	0.0000065
SRR10502311 Triple negative	Yes	91.25	8590444	39	0.0000050
SRR10612305 ductal	Yes	91	21682035	13	0.0000007
SRR7509754 ductal	No	95.58	1087054	23	0.0000221
SRR7509758 ductal	No	96.04	1347370	17	0.0000131
SRR12180250 adenocarcinoma	Yes	91.13	13538984	67	0.0000054
SRR7500823 ductal	Yes	85.92	13765655	12	0.0000010
SRR10247656 ductal	Yes	93.17	60339431	108	0.0000019
SRR7509721 ductal	Yes	93.2	2075305	52	0.0000269
SRR7509732 ductal	No	94.12	1464373	38	0.0000276
SRR7509733 ductal	No	94.32	1442350	47	0.0000345

Table 4. Raw Reads and overall alignment rate came from alignment of any RNA-seq to the Human genome (hg19)

Health Controls	Alignment rate (%)	Raw Reads	Pseudogenome Reads	Normalized coverage of integration sites
SRR7687780	82.47	43714492	42	0.0000012
SRR7687798	80.74	43150648	44	0.0000013
SRR7687782	84.92	63226124	43	0.0000008
SRR7687788	82.18	59081485	54	0.0000011
SRR7687781	85.98	37692344	40	0.0000012
SRR7687797	85.58	46520403	57	0.0000014

Elevated HERV Expression in Breast Cancer

Mean normalized coverage of HERV integration sites was significantly higher in breast cancer cell lines (M = 0.000013) than in healthy tissues (M = 0.000001, p = 0.009) (Table 5).

Effect of Treatment

Patients who had not received treatment exhibited significantly higher HERV expression (mean rank = 9.00) compared to treated patients (mean rank = 4.29, p = 0.023) (Table 6).

In this study, the variable of treatment refers to the clinical status (treated vs. untreated) as documented in the original metadata of the retrieved SRA datasets. Specific details regarding the type of therapeutic agents, regimens, or treatment durations were not available in the public records. Consequently, our analysis compares the retroviral expression profiles between patients classified as 'treated' versus those who were treatment-naive at the time of sampling.

Visualization

The visualization of files was conducted using IGV, a software application that enables the observation of reads on chromosomes [17, 18]. The associated IGV web application enables users to upload FASTQ files and receive visualized outcomes. IGV genome browser confirmed increased read accumulation at specific HERV loci in cancer samples versus healthy controls (Figures 1-4).

Table 5. Independent samples t-test for normalized coverage of integration sites between health controls and breast cancers

Factor	Group	N	M	df	t	p-value
Normalized coverage of integration sites	Health controls	6	0.000001	10.010	-3.210	0.009
	Breast cancers	11	0.000013			

Table 6. Mann-Whitney test for normalized coverage of integration sites in cases of treatment for breast cancer

Factor	Treatment	N	Mean Rank	U	p-value
Normalized coverage of integration sites	No	4	9.00	2	0.023
	Yes	7	4.29		

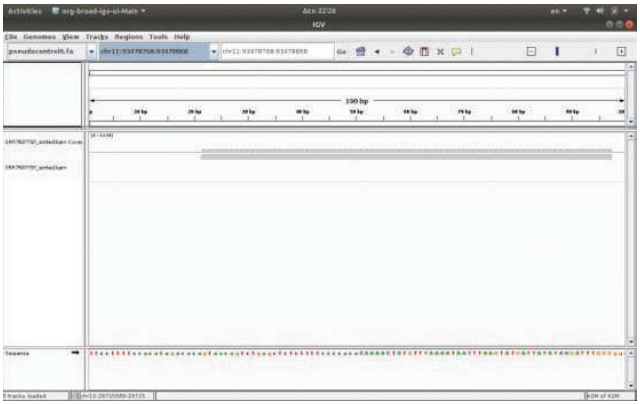

Fig. 1. Visualization of a BAM file of healthy breast tissue (healthy control) in chromosome 11 (chr11: 93478768-93478868) in the IGV viewer

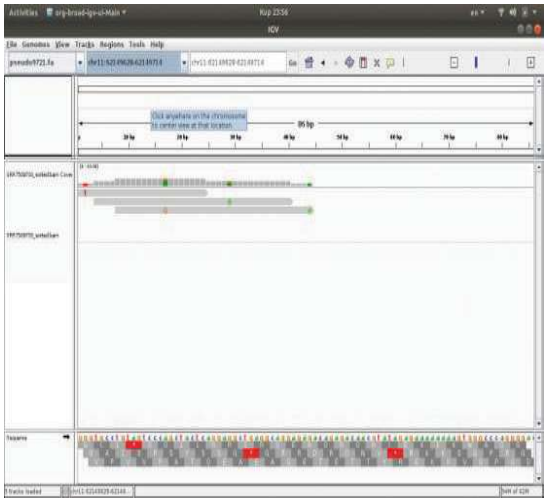

Fig. 2. Visualization of a BAM file of breast cancer tissue in chromosome 11 (chr11:62149628-62149714) in the IGV viewer

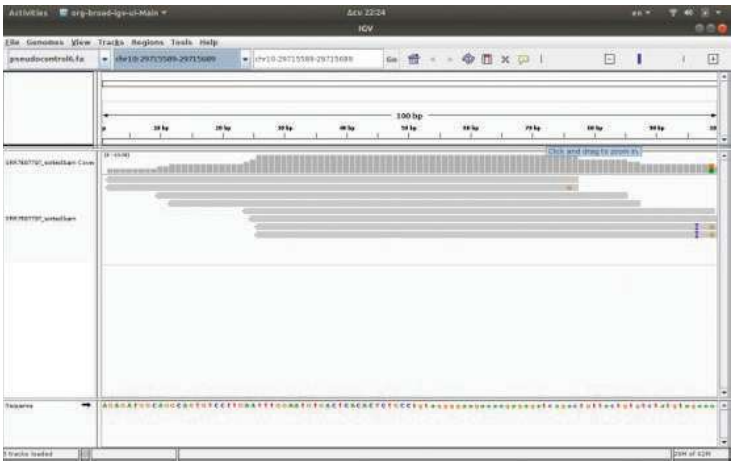

Fig. 3. Visualization of a BAM file of healthy breast tissue (healthy control) in chromosome 10 (chr10:29715589-29715689) in the IGV viewer

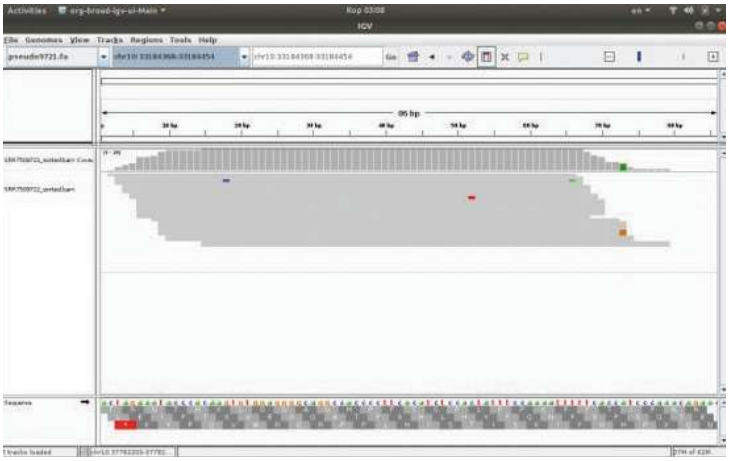

Fig. 4. Visualization of a BAM file of breast cancer tissue in chromosome 10 (chr10:33184368-33184454) in the IGV viewer

DISCUSSION

The present analysis investigated the expression of HERVs in breast cancer tissue compared to normal breast tissue. Furthermore, it explored the variations in normalized coverage of integration sites between breast cancer patients and healthy controls, as well as the differences between treated and untreated breast cancer cases.

The results provide compelling evidence that HERVs, specifically HERV-K (HML-2), are significantly overexpressed in breast cancer cell lines compared to normal breast tissues. This differential expression pattern supports a growing body of literature suggesting that HERVs play a role in the etiology and progression of cancer [19-21].

Several mechanisms may explain the upregulation of HERVs in malignancies. Hypomethylation of HERV sequences and surrounding genomic regions, a common epigenetic alteration in cancer, may reactivate normally silenced proviral elements. Additionally, the tumor microenvironment, characterized by oxidative stress, cytokine production, and immune dysregulation, can further promote HERV expression [22, 23]. Specifically in breast cancer, the envelope protein of HERV-K (HML-2) serves as the most effective diagnostic tool for breast cancer, as it is found in breast cancer cell lines while being absent or present at significantly lower levels in normal breast tissue [24]. Also, the presence of HERV-K (HML-2) antibodies and mRNA serves as an indicator of early-stage breast cancer, with a subsequent rise suggesting the occurrence of metastasis [20, 25].

The observed reduction in HERV expression among patients who received treatment is particularly noteworthy. It suggests that chemotherapy or other therapeutic regimens may suppress HERV activity, potentially by restoring epigenetic silencing mechanisms or by targeting proliferative cancer cells with elevated HERV transcription [19, 26, 27]. This raises important questions regarding the therapeutic targeting of HERVs, both directly and indirectly [28].

The inclusion of the MCF-7 cell line in our study provides a critical perspective on the utility of HERV-K (HML-2) as a biomarker. While MCF-7 is classically categorized as a Luminal A subtype – a classification typically associated with hormone receptor positivity, lower invasiveness, and a more favorable clinical prognosis – it was notably established from a metastatic site. The significant expression of HERVs observed in this line suggests that retroviral reactivation is not exclusive to the highly aggressive Triple-Negative (Basal-like) subtypes, such as MDA-MB-231 or HCC38, but also persists in metastatic hormone-responsive disease. This 'MCF-7 paradox' highlights the complexity of HERV-K (HML-2) expression. It may serve not only as a marker of cellular aggression, but also as an indicator of advanced disease stage and metastatic potential, regardless of the initial molecular classification. Consequently, these findings reinforce the potential of HERV-K (HML-2) as a versatile biomarker for monitoring disease progression and metastasis across diverse breast cancer landscapes.

Nevertheless, these preliminary findings must be interpreted with caution. The number of samples categorized as treated (n = 7) versus untreated (n = 4) is modest, and the specific therapeutic regimens administered to the treated patients, including the type of chemotherapy, targeted therapy, or hormonal intervention, were not detailed in the available public metadata. Consequently, we cannot distinguish between the effects of specific therapeutic agents or treatment durations on HERV transcript levels. Our results should, therefore, be viewed as preliminary, suggesting a potential correlation that requires validation in larger, longitudinal clinical studies with standardized treatment protocols to establish a causal relationship between therapeutic intervention and the modulation of HERV expression.

While our transcriptomic analysis demonstrates a clear overexpression of HERVs in breast cancer cell lines, interpreting these results requires careful consideration of the transition from *in vitro* models to the *in vivo* environment. Cell lines represent simplified models that lack the complex interactions of the tumor microenvironment, including immune cell infiltration, cytokine signaling, and stromal influences, all of which are known to promote HERV expression.

Furthermore, the causality of HERV activation in malignancy remains a subject of ongoing debate. It is currently unclear whether HERV reactivation is a primary oncogenic driver or a secondary byproduct of the profound epigenetic dysregulation, such as global hypomethylation, that characterizes tumorigenesis. Given that tumor progression often involves oxidative stress and immune dysregulation, it is plausible that HERVs are induced by these factors, subsequently contributing to further genomic instability and tumor heterogeneity. Consequently, while our findings strengthen the potential of HERV-K (HML-2) as a diagnostic and prognostic biomarker, the exact contribution of these retroelements to the onset versus the progression of breast cancer requires further functional validation in patient-derived tissues and longitudinal cohorts.

While this study offers valuable insights, limitations exist. Firstly, our research relies on the analysis of public transcriptomic datasets, which, while valuable, lack the controlled experimental validation required to confirm functional consequences. A critical limitation is the reliance on permanent cell lines for transcriptomic analysis. While widely accepted as models, these systems have limitations that must be acknowledged. First, prolonged laboratory cultivation can induce phenotypic, genetic, and epigenetic drift, potentially distancing the cell line from the original patient tumor. Second, the use of 2D monolayer culture strategies distances these cells from the three-dimensional architecture and microenvironment of the original tumor. Third, due to the profound heterogeneity of breast cancer and continuous clonal evolution, investigations based on a small number of cell lines make it difficult to generalize conclusions to the disease as a whole.

Also, while our findings demonstrate consistent overexpression of HERV-K (HML-2) across multiple cancer-derived lines compared to healthy controls, we cannot definitively rule out that these culture-specific conditions influence our observations. Therefore, our results should be interpreted as a foundational observation of retroviral reactivation, which warrants further validation in primary patient tissues to determine the extent to which these patterns reflect in vivo tumor biology.

Secondly, the sample size remains modest, which limits our ability to generalize findings across all breast cancer subtypes. Finally, as our control group consisted of healthy breast tissue samples rather than matched-normal tissue from the same patients, further longitudinal studies are necessary to definitively characterize the differential expression of HERVs during disease progression. Further research is necessary to elucidate the functional consequences of HERV expression and its interaction with oncogenic signaling pathways.

Challenges, implications for practice, and future directions

The findings contribute to the growing evidence that HERVs, particularly the HERV-K (HML-2) family, are significant entities in breast cancer transcriptomics. Our results support the potential of HERV-K expression levels as non-invasive biomarkers, offering a prospective avenue for both early detection and monitoring of therapeutic efficacy. By demonstrating that HERV expression is modulated by therapeutic intervention, we provide a foundation for future studies aiming to incorporate retroviral expression profiles into personalized oncology strategies.

Despite these insights, research into HERVs and oncogenesis faces significant hurdles. A primary challenge is the ‘driver versus passenger’ dilemma: determining whether HERV reactivation is a direct cause of malignant transformation through insertional mutagenesis or immune dysregulation, or if it is merely a consequence of the global epigenetic hypomethylation inherent to cancer cells. Furthermore, the heterogeneity of breast cancer and the complex interactions within the tumor microenvironment, such as oxidative stress and cytokine signaling, make it difficult to standardize HERV expression markers across diverse patient

populations. Establishing a universal 'HERV signature' remains an ongoing challenge for the field.

The findings of this study have several clinical and research implications. In clinical practice, HERV-K expression could serve as a non-invasive biomarker for early breast cancer detection [24]. The observed decrease in HERV expression following treatment suggests potential utility in tracking therapeutic efficacy. Furthermore, stratifying patients by HERV expression profiles may aid in developing tailored treatment strategies and enhancing personalized medicine approaches. The field of personalized medicine is experiencing a significant transformation due to the incorporation of multi-omics data, which primarily includes genomics, transcriptomics, proteomics, and metabolomics [29].

From a research perspective, laboratory-based validation is essential to confirm the functional role of HERVs in tumor growth and immune modulation. The immunogenic nature of HERV-derived proteins also presents opportunities for the development of novel immunotherapies, including cancer vaccines and CAR-T cell therapy [30]. Lastly, longitudinal studies monitoring HERV expression over time could provide valuable insights into disease progression, recurrence, and metastasis.

CONCLUSION

This transcriptomic analysis demonstrates significant overexpression of HERVs in breast cancer cell lines compared to healthy breast tissue samples, particularly in untreated patients. HERV expression profiling may aid in diagnosis, prognosis, and treatment response monitoring in breast cancer care. With further validation, HERVs could become valuable biomarkers and therapeutic targets in oncology.

Statements: *All authors have read and agreed to the published version of the manuscript.*

Funding: *This research received no external funding.*

Institutional Review Board Statement: *Not applicable.*

Informed Consent Statement: *Not applicable.*

Data Availability Statement: *All data presented in this study are available from the corresponding author upon reasonable request.*

Conflicts of Interest: *The authors declare no conflicts of interest.*

REFERENCES

1. Arnold M, Morgan E, Rumgay H, et al. Current and future burden of breast cancer: global statistics for 2020 and 2040. Breast. 2022;66:15–23.
2. Giaquinto AN, Sung H, Newman LA, et al. Breast cancer statistics 2024. CA Cancer J Clin. 2024;74(5):477–495.
3. Merrill RM, Gibbons IS. Inequality in female breast cancer relative survival rates between white and black women in the United States. J Racial Ethn Health Disparities. 2024;1–10.
4. Zhang S, Xiao X, Yi Y, et al. Tumor initiation and early tumorigenesis: molecular mechanisms and interventional targets. Signal Transduct Target Ther. 2024;9:149.
5. Xiong X, Zheng LW, Ding Y, et al. Breast cancer: pathogenesis and treatments. Signal Transduct Target Ther. 2024;10:49.
6. Mao J, Zhang Q, Cong YS. Human endogenous retroviruses in development and disease. Comput Struct Biotechnol J. 2021;19:5978–5986.
7. Geis FK, Goff SP. Silencing and transcriptional regulation of endogenous retroviruses: an overview. Viruses. 2020;12(8):884.
8. Lemaître C, Tsang J, Bireau C, et al. A human endogenous retrovirus-derived gene that can contribute to oncogenesis by activating the ERK pathway and inducing migration and invasion. PLoS Pathog. 2017;13(1):e1006451.
9. Xue B, Sechi LA, Kelvin DJ. Human endogenous retrovirus K (HML-2) in health and disease. Front Microbiol. 2020;11:1690.
10. Dervan E, Bhattacharyya DD, McAuliffe JD, et al. Ancient adversary - HERV-K (HML-2) in cancer. Front Oncol. 2021;11:658489.
11. Subramanian RP, Wildschutte JH, Russo C, et al. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. Retrovirology. 2011;8:90.
12. Montesion M, Williams ZH, Subramanian RP, et al. Promoter expression of HERV-K (HML-2) provirus-derived sequences is related to LTR sequence variation and polymorphic transcription factor binding sites. Retrovirology. 2018;15:57.
13. National Center for Biotechnology Information (NCBI). The Sequence Read Archive (SRA) [Internet]. Bethesda (MD): NCBI; [cited 2024]. Available from: https://www.ncbi.nlm.nih.gov/sra
14. Subramanian I, Verma S, Kumar S, et al. Multi-omics data integration, interpretation, and its application. Bioinform Biol Insights. 2020;14:1177932219899051.
15. Belshaw R, Pereira V, Katzourakis A, et al. Long-term reinfection of the human genome by endogenous retroviruses. Proc Natl Acad Sci USA. 2004;101(14):4894–4899.
16. BEDTools. getfasta [Internet]. [cited 2024]. Available from: https://bedtools.readthedocs.io/en/latest/content/tools/getfasta.html
17. Bartzi D. Human endogenous retroviruses (HERVs) transcriptomics in human breast cancer [Master's thesis]. Athens: National and Kapodistrian University of Athens; 2021. Available from: https://pergamos.lib.uoa.gr/uoa/dl/frontend/en/browse/2938645#contents
18. Integrative Genomics Viewer (IGV) [Internet]. [cited 2024]. Available from: https://igv.org/
19. Costa B, Vale N. Exploring HERV-K (HML-2) influence in cancer and prospects for therapeutic interventions. Int J Mol Sci. 2023;24(19):14631.
20. Johanning GL, Malouf GG, Zheng X, et al. Expression of human endogenous retrovirus-K is strongly associated with the basal-like breast cancer phenotype. Sci Rep. 2017;7:41960.
21. Rivas SR, Valdez MJM, Govindarajan V, et al. The role of HERV-K in cancer stemness. Viruses. 2022;14(9):2019.
22. Burn A, Roy F, Freeman M, et al. Widespread expression of the ancient HERV-K (HML-2) provirus group in normal human tissues. PLoS Biol. 2022;20(10):e3001826.

23. Cherkasova EA, Chen L, Childs RW. Mechanistic regulation of HERV activation in tumors and implications for translational research in oncology. Front Cell Infect Microbiol. 2024;14:1358470.
24. Wei Y, Wei H, Wei Y, et al. Screening and identification of human endogenous retrovirus-K mRNAs for breast cancer through integrative analysis of multiple datasets. Front Oncol. 2022;12:820883.
25. Wang-Johanning F, Li M, Esteva FJ, et al. Human endogenous retrovirus type K antibodies and mRNA as serum biomarkers of early-stage breast cancer. Int J Cancer. 2014;134(3):587–595.
26. Simula ER, Jasemi S, Cossu D, et al. Human endogenous retroviruses as novel therapeutic targets in neurodegenerative disorders. Vaccines. 2025;13(4):415.
27. Bannert N, Hofmann H, Block A, et al. HERVs new role in cancer: from accused perpetrators to cheerful protectors. Front Microbiol. 2018;9:178.
28. Stricker E, Peckham-Gregory EC, Scheurer ME. HERVs and cancer – a comprehensive review of the relationship of human endogenous retroviruses and human cancers. Biomedicines. 2023;11(3):936.
29. Molla G, Bitew M. Revolutionizing personalized medicine: synergy with multi-omics data generation, main hurdles, and future perspectives. Biomedicines. 2024; 12(12):2750.
30. Logotheti S, Stiewe T, Georgakilas AG. The role of human endogenous retroviruses in cancer immunotherapy of the post-COVID-19 world. Cancers. 2023;15(21):5321.