


Repetitive Sequence Transcription in Breast Cancer

Walter Arancio  and Claudia Coronello * 

Advanced Data Analysis Group, Fondazione Ri.MED, 90133 Palermo, Italy

* Correspondence: ccoronello@fondazionerimed.com

Abstract: Repetitive sequences represent about half of the human genome. They are actively transcribed and play a role during development and in epigenetic regulation. The altered activity of repetitive sequences can lead to genomic instability and they can contribute to the establishment or the progression of degenerative diseases and cancer transformation. In this work, we analyzed the expression profiles of DNA repetitive sequences in the breast cancer specimens of the HMUCC cohort. Satellite expression is generally upregulated in breast cancers, with specific families upregulated per histotype: in HER2-enriched cancers, they are the human satellite II (HSATII), in luminal A and B, they are part of the ALR family and in triple-negative, they are part of SAR and GSAT families, together with a perturbation in the transcription from endogenous retroviruses and their LTR sequences. We report that the background expression of repetitive sequences in healthy tissues of cancer patients differs from the tissues of non-cancerous controls. To conclude, peculiar patterns of expression of repetitive sequences are reported in each specimen, especially in the case of transcripts arising from satellite repeats.

Keywords: breast cancer; repetitive sequences; HERV; endogenous retrovirus; satellite repeats; centromeres; telomeres; SVA; LINE1; transposons



Citation: Arancio, W.; Coronello, C. Repetitive Sequence Transcription in Breast Cancer. *Cells* **2022**, *11*, 2522. <https://doi.org/10.3390/cells11162522>

Academic Editor: Bolin Liu

Received: 20 July 2022

Accepted: 12 August 2022

Published: 14 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Breast Cancer Classification

Breast cancer is still the leading cause of mortality among the female population in developed countries. In post-menopausal women, it accounts for 23% of all cancer deaths [1]. Breast cancers can be classified following anatomical, histological and molecular features [1], and their classification is a dynamic process, as stated in the last World Health Organization classification of tumors of the breast [2], and novel entities are added to the classification year by year following the increase in the knowledge of the disease [3]. Breast cancer is as a heterogeneous disease with different clinical and pathological features, variable therapeutic approaches and responses and with different outcomes even within the same class of breast cancer, suggesting that the current classifications are far from exhaustive. In order to follow the original classification of the cohort used in this study [4,5], the breast cancer specimens have been classified as: luminal-A, luminal-B (HER2-negative), luminal-B (HER2-positive), HER2-enriched and triple-negative breast cancers. This classification is based on immunohistochemical-relevant markers and was recommended by the St. Gallen Expert Consensus [6] and it has become a standard in routine clinical analysis since then [6–8]. For detail reviews, please refer to Hennigs et al. [8] and Prat et al. [9].

1.2. Non-Coding RNA in Mammals

The mammalian genome is pervasively transcribed and only a small portion of the transcriptional output has protein-coding potential [10]. The non-coding RNAs (ncRNAs) can be categorized using sizes and function, such as small-nuclear RNAs (snRNAs), small-nucleolar RNAs (snoRNAs), long non-coding RNAs (lncRNAs) and many others. The most well-studied class of ncRNAs is probably represented by microRNAs (miRNAs). Many

studies have identified or suggested their role in human health and diseases, aging [11], cancer [12,13], diagnostic or prognostic purposes [14,15] and as therapeutic agents [16], either per se or in complex networks of cross regulation by the name of competing endogenous RNAs (ceRNAs) [17–24]. An abundant class of ncRNAs with variable functions that are still not fully understood is represented by transcripts arising from non-coding DNA sequences that are repeated along the genome in multiple copies. Even if the transcription from a single copy can be negligible, the sum of transcripts arising from thousands or millions of copies can be massive. A detailed description of them is given in the following paragraphs.

1.3. Repetitive DNA Sequence Classification

A multifaceted category of ncRNAs, of growing interest due to their roles in human health and diseases, is represented by the transcripts arising from repetitive DNA sequences (RS), i.e., DNA sequences that are present in multiple copies in the genomes, with low or nonexistent coding potential. RS represent about 45% of the human genome and are differentially transcribed in many tissues [25]. In mammals, RS have many roles in development and epigenetic regulation, but also in diseases such as cancer transformation [26–30] and degenerative diseases [31], but they are notoriously difficult to study [32]. Due to their nature, length and origin, RS can be roughly classified as: (i) Satellite repeats: a tandem array of simple or complex sequence repeats, abundant in heterochromatic regions, including alpha satellite repeats that represent the main DNA component of human centromeres. (ii) Long interspersed nuclear elements (LINEs): retrotransposons devoid of long terminal repeats (non-LTR) including some that are still able to retrotranspose. (iii) Small interspersed nuclear elements (SINEs): non-autonomous retrotransposons including the Alu elements in humans, which are often involved in genomic rearrangements. (iv) Integrated LTR retroviruses, mainly represented by the human endogenous retrovirus (HERV) families. (v) Additionally, the families of DNA transposons, that are usually not active in humans (Figure 1). The role of RS is starting to be properly understood. E.g., in the human brain LINE-1 retrotransposons are actively transcribed and mobilized and they are suggested to play a role in shaping the adult human brain [33], there is also a suggested role of RS in a model of aging of human brain [34].

1.4. Repetitive DNA Sequence and Cancer

Increased levels of heterochromatic repetitive satellite-coded RNAs in mammary glands induce breast tumor formation in mice, altering the BRCA1-associated protein networks that are required for the proper stabilization of DNA replication forks that in turn lead to genomic instability [35]. In humans, patients with breast cancer that express high levels of RNA derived from alpha satellite have an increased risk of developing multiple cancers [36].

It is known that LINE-1-encoded retrotranscription activity is widespread and its inhibition can reduce the rate of proliferation and promote the differentiation of breast cancer cells [37]. LINE-1 (and Alu) hypomethylation, suggesting an increased transcription in cancer cells and thus their mobilization, has been associated with the HER2-enriched subtype of breast cancer with worst prognosis [38–40]. In the transgenic mice of a well-defined model of breast cancer progression, LINE-1 is upregulated at a very early stage of tumorigenesis [41]. Indeed, the altered expression patterns of LINE-1-coded ORF1 and ORF2 proteins, with differences in overall patient survival, have been reported in invasive breast cancers [42]. In specific cases, pesticide exposure induces LINE-1 reactivation, suggesting the role of LINE transcription in pesticide-induced breast cancer progression [43], and MET-LINE-1 chimeric transcripts identify a subgroup of aggressive triple-negative breast cancers [44]. Overall, it has been suggested that LINE-1 may contribute to the origin or progression of breast cancers [45].

There are many reports regarding Alu and other SINE elements within or surrounding *BRCA1* and *BRCA2* genes essential to genomic rearrangements or genetic mutations leading to etiopathogenic, prognostic or predisposing mutations of breast cancers, both in somatic

and germ lines [46–51]; indeed, the demethylation of Alu sequences may induce, at the same time, both transcription and rearrangements of Alu sequences. Thus, Alu transcription is a marker of increased susceptibility to Alu-mediated genomic rearrangement or genetic mutation at Alu sites. Looking for a direct effect of Alu transcription, it is noteworthy that heterogeneous nuclear ribonucleoprotein C (HNRNPC) is essential in breast cancer cell survival by inhibiting the double-stranded-RNA (dsRNA)-induced interferon response. Indeed, dsRNA in this setting is highly enriched in Alu sequences [52], suggesting that an overexpression of Alu sequences is characteristic of many breast cancers and may have lethal effects in cancer cells if not controlled.

There is significant evidence regarding the use of HERV-K-coded proteins as tumor markers and immunologic targets [52–58] and in influencing cancer stemness [59]. It has even been suggested that they could act as etiological agents [60,61]. Indeed, the expression of HERV-K is upregulated and associated with the basal-like breast cancer phenotype [62] and a HERV-derived long non-coding RNAs (namely, TROJAN) promotes triple-negative breast cancer progression [63]. HERV can directly contribute to cancer progression by activating the ERK pathway and inducing migration and invasion [64]; it has been even suggested that the activation of HERV-K may be essential for the tumorigenesis and metastasis of breast cancer [65]. Indeed, HERV-K-derived RNAs and antibodies against HERV-K-coded proteins are elevated in the blood of patients at an early stage of breast cancer [66].

DNA transposons are the less active and less well-studied class of RS in humans. Nevertheless, few reports suggest their role in breast cancer [67]; however, they were not investigated further. In addition, a mechanism of *BRCA1* mutation in three unrelated French breast/ovarian cancer families, that can be generated by an abortive integration of the human Tigger1 DNA transposon, has been postulated [68].

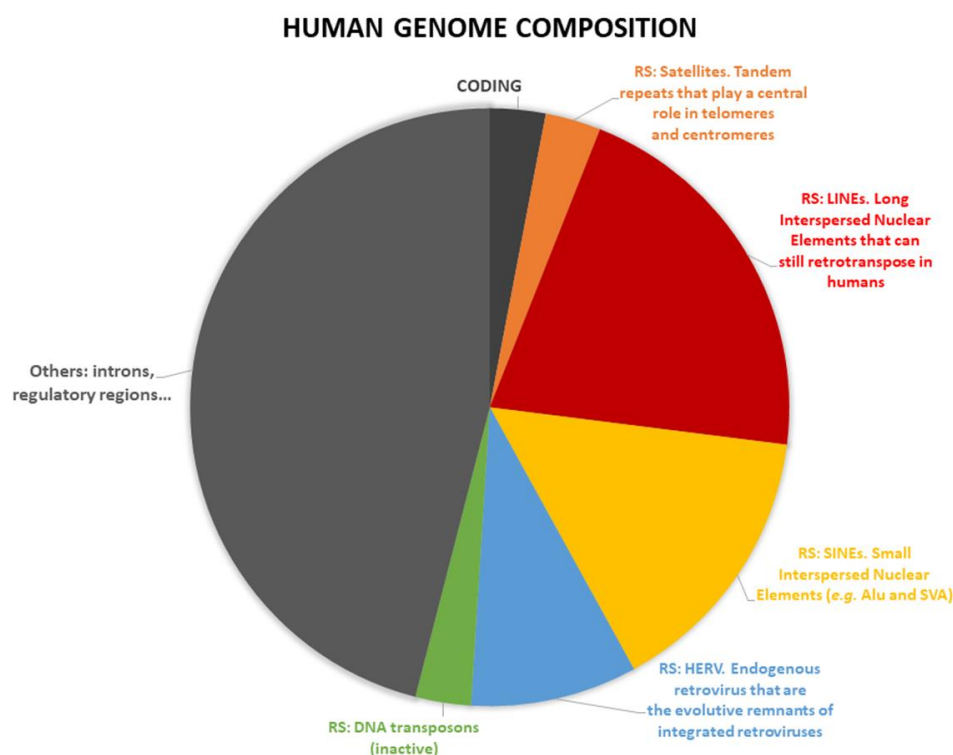


Figure 1. Repetitive sequences (RS) represent about half of the human genome. The panel reports RS activities associated with breast cancer. In orange: Satellite repeats [35,36]. In red: Long interspersed nuclear elements (LINEs) [37,42,45]. In yellow: Small interspersed nuclear elements (SINEs) [46–51]. In blue: Human endogenous retrovirus (HERV) [62–64,66]. In green: DNA transposons [68].

1.5. Main Aim

The main aim of this work was to analyze the transcripts arising from the repetitive sequences in a cohort of breast cancer specimens. Overall, we report peculiar patterns of expression and a diffuse upregulation of satellite transcription specific for each histotype.

2. Materials and Methods

2.1. Identifying and Quantifying the Repetitive Sequence Expression

The method of analysis and positive and negative controls (human beta actin cDNA NM_001101.4, 18S and 5.8 human ribosomal subunits, and the locus EF191515 of *Bacillus subtilis* SMY strain) have previously been described [34].

In brief, the analyses have been performed in a Galaxy environment [69,70]. The FASTQ raw sequences (obtained from the European Nucleotide Archive) have been uploaded, processed by Trimmomatic (Galaxy Version 0.36.5) [71] and quality checked. Bowtie2 aligner (Galaxy Version 2.3.4.2) [72] has been used with very sensitive local parameters to retrieve the expression of RS, oblivious of their genomic localization, aligning the reads against pseudochromosomes containing the reference RS sequences. ‘Very sensitive’ parameter takes into account the intrinsic sequence variability of RS, aiming to retrieve RS sequences that are slightly different from the canonical sequence used as reference. ‘Local’ alignment allows the retrieval of RS sequences attached to other sequences in the same reads, because RS are often embedded in other transcripts. Beta actin cDNA is a positive control of the pipeline used to retrieve the RS: RS are unspliced and the Bowtie2 aligner must efficiently retrieve the cDNA sequence of mature mRNAs, such as NM_001101.4, when treated as a pseudochromosome. Mature rRNA subunit sequences have been used as positive controls, following the same rationale. Their large amount allows their retrieval even after rRNA depletion, which never reaches total efficiency, and they are physiologically unspliced, non-poly-adenylated ncRNAs of different lengths. Instead, EF191515 locus is a negative control that has no significant homology with human sequences and thus must have zero or almost zero reads in every sample. Raw data are reported in Supplementary Table S1.

2.2. Analysis of Coding Gene Expression

The raw FASTQ data have been aligned by the means of HISAT2 aligner (Galaxy Version 2.1.0) [73] using the Galaxy embedded hg38 as a reference. The generated BAM files have been compared using the hg38_GENCODE_GENE_V19.bed as a reference.

2.3. Statistical Analyses

The differential expression analysis was performed with the DESeq2 [74] algorithms implemented in a Galaxy environment (Version 2.11.40.2). The analysis was performed on the merged raw counts dataset, including coding genes, RS and control expression, if not otherwise specified. *p*-value correction for multiple comparisons was performed with the Benjamini and Hochberg method [75].

2.4. Dataset Used

We analyzed the data published in the European Nucleotide Archive (ENA), RRID:SCR_006515, study accession: PRJNA292118 [4,5]. The dataset contains sequencing data of 15 invasive breast cancer specimens (3 each for luminal-A, luminal-B (HER2-negative), luminal-B (HER2-positive), HER2-enriched and triple-negative breast cancer) and 18 controls (15 paired adjacent non-cancerous tissues and 3 healthy tissues). As reported in experiments SRX1135937 to SRX1135969 in the PRJNA292118 project [76], RNA was ribo-depleted via Ribo-Zero™ Gold Kits (human) before using the Illumina TruSeq RNA Sample Prep Kit (Cat#FC-122-1001) for the construction of the sequencing libraries. This kind of library allowed us to analyze both poly-adenylated and non-poly-adenylated transcripts; thus, it is suitable to analyze transcripts arising from RS, whose poly-adenylation status is often unknown.

3. Results

The expression of repetitive sequences has been retrieved and analyzed accordingly to the previously described pipeline [34] with minor adaptations. Raw expression data are reported in Supplementary Table S1. The normalized expression data of the merged raw counts of the coding genes and RS are reported in Supplementary Table S2. Whisker plots for selected RS of interest are reported in Figure 2.

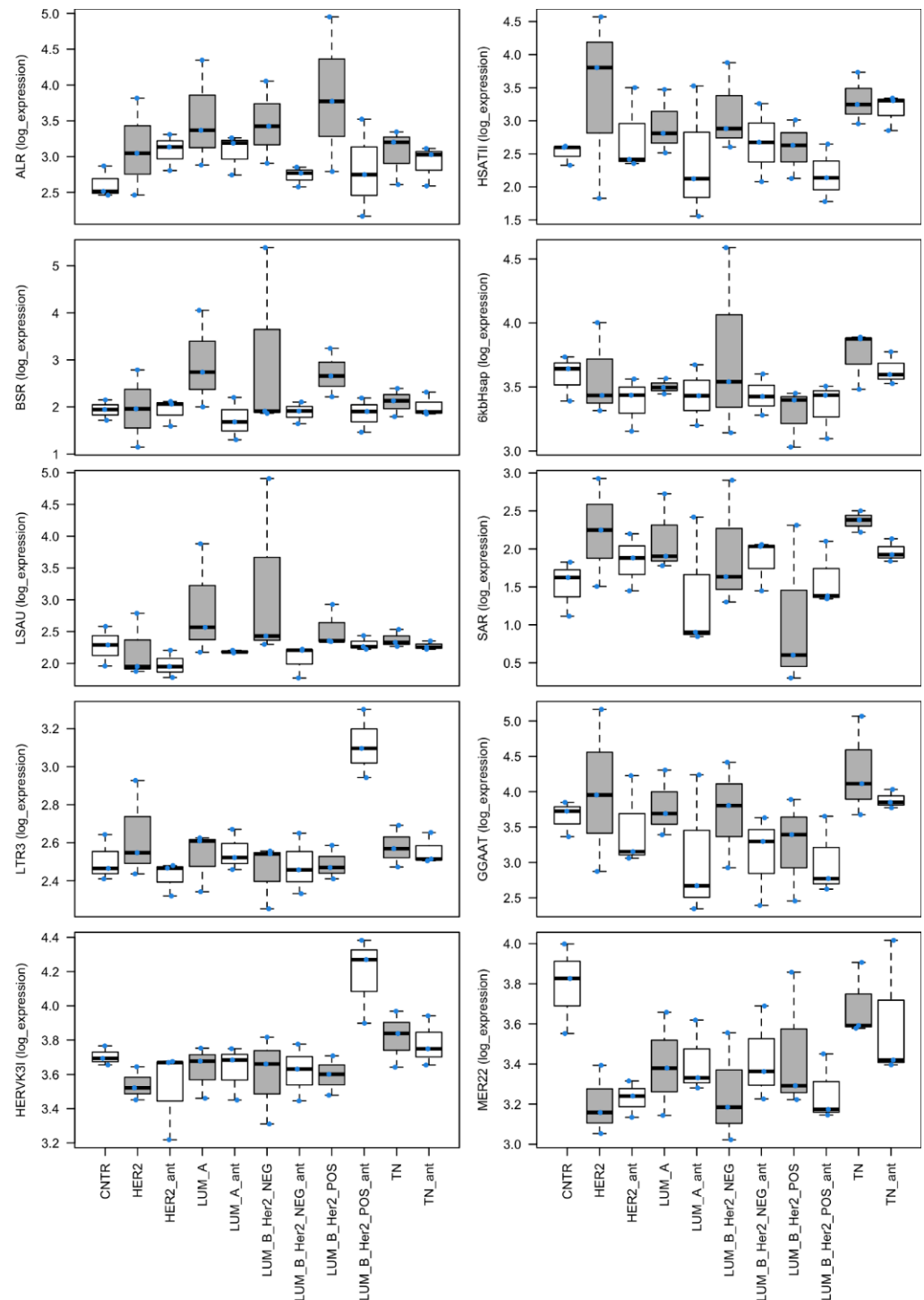


Figure 2. Whisker plots of the expression of selected RS in Log10. The cancerous specimens are plotted in gray.

3.1. Analysis of the Expression of Repetitive Sequences in Cancer Specimens

A comparison between the 15 invasive breast cancer specimens and 18 controls indicated a panel of RS differentially expressed in the two conditions (Table 1, Figure 2 and Supplementary Table S3). ALR, BSR and LSAU repetitive sequences are the most significantly upregulated, i.e., more than two-fold, in cancer specimens compared to the controls ($P\text{-adj} < 0.05$). However, the specimens showed great variability (Supplementary Webpage S4); thus, a comparison between each histotype with its adjacent normal tissues (ant) was performed.

Table 1. A comparison between the 15 invasive breast cancer specimens and 18 controls (cancer vs. normal) and between each histotype with its adjacent non-cancerous tissues highlighted a panel of RS differentially expressed in the two conditions. The RS whose mean expression is above 100 and with a p -value < 0.05 are reported.

GeneID	Base Mean	log ₂ (FC)	StdErr	Wald-Stats	p -Value	P-adj
Cancer vs. normal						
BSR	597.99	1.62	0.28	5.80	6.57×10^{-9}	1.93×10^{-4}
ALR	5004.62	1.52	0.28	5.44	5.24×10^{-8}	7.70×10^{-4}
LSAU	484.17	1.36	0.28	4.91	9.31×10^{-7}	8.20×10^{-3}
ALRb	1358.44	1.03	0.27	3.81	1.37×10^{-4}	1.44×10^{-1}
ALR1	4308.91	0.99	0.27	3.73	1.90×10^{-4}	1.48×10^{-1}
GGAAT	11,375.95	0.91	0.28	3.28	1.05×10^{-3}	3.35×10^{-1}
HSATII	2167.60	0.91	0.28	3.27	1.08×10^{-3}	3.35×10^{-1}
PABL_BI	121.89	0.46	0.16	2.89	3.86×10^{-3}	5.15×10^{-1}
SAR	137.70	0.75	0.28	2.69	7.08×10^{-3}	6.09×10^{-1}
LTR22B2	517.80	-0.32	0.13	-2.52	1.17×10^{-2}	7.00×10^{-1}
LTR72	226.51	-0.30	0.12	-2.41	1.58×10^{-2}	7.74×10^{-1}
LTR12C	39,355.33	-0.51	0.22	-2.36	1.84×10^{-2}	8.01×10^{-1}
MER9B	648.51	-0.26	0.12	-2.27	2.33×10^{-2}	8.33×10^{-1}
LTR7B	2761.48	-0.35	0.16	-2.21	2.70×10^{-2}	8.50×10^{-1}
LTR7C	483.73	-0.25	0.12	-2.11	3.46×10^{-2}	8.83×10^{-1}
LTR35	156.02	-0.21	0.10	-2.10	3.61×10^{-2}	8.86×10^{-1}
ALR_	10,312.15	0.49	0.24	2.00	4.51×10^{-2}	9.16×10^{-1}
HER2 vs. ant						
ZAPHOD	160.72	-0.86	0.41	-2.07	3.81×10^{-2}	1.00
HSATII	6250.25	1.10	0.56	1.96	4.98×10^{-2}	1.00
LumA vs. ant						
ALR	5513.55	1.11	0.44	2.52	1.16×10^{-2}	1.00
LTR38B	472.48	-1.04	0.45	-2.32	2.04×10^{-2}	1.00
LumB_Her2Neg vs. ant						
ALR	2804.92	1.75	0.48	3.64	2.74×10^{-4}	1.29×10^{-1}
ALR1	3720.86	1.36	0.46	2.96	3.07×10^{-3}	3.40×10^{-1}
ALRb	1210.79	1.13	0.45	2.52	1.19×10^{-2}	6.05×10^{-1}
MER57C1	196.12	1.00	0.43	2.31	2.08×10^{-2}	7.23×10^{-1}
6kbHsap	7701.13	1.06	0.49	2.18	2.94×10^{-2}	8.01×10^{-1}

Table 1. Cont.

GeneID	Base Mean	log2(FC)	StdErr	Wald-Stats	p-Value	P-adj
LumB_Her2Pos vs. ant						
LTR3	853.33	−1.87	0.45	−4.14	3.45×10^{-5}	1.33×10^{-2}
HERVK3I	10,531.29	−1.76	0.47	−3.72	1.99×10^{-4}	4.53×10^{-2}
BSR	436.11	2.09	0.60	3.49	4.92×10^{-4}	8.70×10^{-2}
ALR1	20,223.11	2.06	0.64	3.21	1.32×10^{-3}	1.56×10^{-1}
ALR	14,122.78	1.97	0.64	3.06	2.24×10^{-3}	2.10×10^{-1}
LTR1B0	5449.61	−1.94	0.64	−3.02	2.50×10^{-3}	2.23×10^{-1}
LTR12C	61,931.23	−1.35	0.55	−2.45	1.45×10^{-2}	5.91×10^{-1}
MER122	128.05	1.31	0.64	2.06	3.97×10^{-2}	9.10×10^{-1}
ALR_	13,301.52	1.18	0.59	1.99	4.61×10^{-2}	9.55×10^{-1}
TN vs. ant						
SAR	171.13	0.84	0.28	3.01	2.63×10^{-3}	2.41×10^{-1}
LTR72B	423.96	−0.56	0.22	−2.57	1.03×10^{-2}	4.98×10^{-1}
MER87B	580.38	0.43	0.19	2.23	2.59×10^{-2}	7.70×10^{-1}
GSAT	291.10	0.58	0.29	1.98	4.72×10^{-2}	9.7×10^{-1}

In ‘HER2-enriched’ breast cancer, an upregulation of HSATII satellite expression is reported (Table 1; Supplementary Table S5). Regarding ‘luminal-A’ histotype, the expression of ALR is strongly upregulated (Table 1; Supplementary Table S6). In the case of ‘luminal-B HER2-negative’, the expression of ALR satellite family and “6 kb tandem repeat sequence in Homo sapiens” is upregulated (Table 1; Supplementary Table S7). We also report a striking upregulation of BSR satellite sequences between a specific luminal-B HER2-negative specimen and its adjacent non-cancerous tissues (namely, LUM_B_Her2_NEG_0). Its normalized expression is 208203 reads in the cancer specimen against 88 in the adjacent tissue (Supplementary Table S2). Thus, this specimen has the worst classification in TNM staging in the cohort (IIIb, together with another specimen) and many lymph node metastases (36 positives out of the 42 inspected). Analyzing ‘luminal-B HER2-positive’ specimens, we report a generalized upregulation of several satellite-derived transcripts together with a downregulation of endogenous retroviruses and their LTR sequences (Table 1; Supplementary Table S8). In ‘triple-negative breast cancer’, there is an upregulation of satellite sequences and a perturbation in the transcription of endogenous retroviruses and their LTR sequences (Table 1; Supplementary Table S9).

3.2. Analysis of Expression Background

A detailed analysis of the comparison between the expression of ANT and the normal controls is reported in Table 2 and Supplementary Table S10. The most differentially expressed RS is MER22 satellite (also called SST1) [77,78], which is downregulated in ANT compared to the controls. Another abundant RS that is differentially expressed in ANT compared to the controls is SVA_A, which is also downregulated. SVAs are SINES that contain Alu sequences. SVAs are still active in humans and may have biological effects [79].

We also report that the vast majority of the specimens analyzed showed an altered expression of GGAAT repeats in comparison with their specific ANT, either increased (GGAAT^{higher}) or decreased (GGAAT^{lower}).

Table 2. Comparison between the expression of RS in ANT and the normal controls (ANT vs. normal). The RS whose mean expression is above 100 and with a *p*-value < 0.05 are reported.

GeneID	Base Mean	log2(FC)	StdErr	Wald-Stats	<i>p</i> -Value	P-adj
MER22	3245.14	−1.03	0.35	−2.95	3.18×10^{-3}	6.42×10^{-1}
PABL_BI	101.04	0.67	0.24	2.82	4.78×10^{-3}	6.42×10^{-1}
PTR5	394.09	−1.10	0.41	−2.67	7.67×10^{-3}	6.42×10^{-1}
TAR1	184.59	−0.66	0.26	−2.53	1.15×10^{-2}	6.42×10^{-1}
LTR7C	516.36	−0.55	0.23	−2.44	1.45×10^{-2}	6.42×10^{-1}
GSAT	178.69	−1.22	0.51	−2.37	1.79×10^{-2}	6.42×10^{-1}
LTR46	249.63	−0.69	0.31	−2.18	2.94×10^{-2}	6.42×10^{-1}
LTR7Y	1511.10	−0.76	0.35	−2.16	3.05×10^{-2}	6.42×10^{-1}
LTR1E	413.79	−0.59	0.28	−2.15	3.19×10^{-2}	6.42×10^{-1}
LTR7A	2067.24	−0.77	0.36	−2.14	3.20×10^{-2}	6.42×10^{-1}
LTR22B2	563.94	−0.50	0.23	−2.14	3.27×10^{-2}	6.42×10^{-1}
LTR44	116.12	0.81	0.39	2.08	3.79×10^{-2}	6.42×10^{-1}
LTR27C	315.21	−0.59	0.28	−2.08	3.79×10^{-2}	6.42×10^{-1}
ALR	1028.61	1.00	0.48	2.07	3.83×10^{-2}	6.42×10^{-1}
HARLEQUINLTR	3541.69	−0.45	0.22	−2.07	3.84×10^{-2}	6.42×10^{-1}
MER51C	493.52	−0.45	0.22	−2.03	4.19×10^{-2}	6.42×10^{-1}
SVA_A	65,198.32	−0.57	0.28	−2.02	4.32×10^{-2}	6.42×10^{-1}
MER54B	165.03	−0.57	0.28	−2.00	4.54×10^{-2}	6.42×10^{-1}

4. Discussion

High-throughput RNA sequencing helps one analyze the pervasive transcription of the human genome, but it bears the burden of processing a huge amount of data. In the routinely used pipelines of analysis, the transcription of RS is often overlooked due to the intrinsic difficulties to be analyzed by the most common means. Nevertheless, RS represent about half of the human genome and the source of a fair amount of transcriptional output. Indeed, whenever analyzed, transcripts arising from RS showed biological and medical properties far beyond the role of simple bystanders or byproducts [31,32].

While several studies on breast cancers highlighted a potential role of specific RS as etiopathogenic agents or as diagnostic or prognostic tools, this is the first study analyzing the expression of RS as a whole in a panel of breast cancer specimens classified by their molecular characteristics.

Overall, it is evident that the cancerous specimens showed an increased expression from satellite repeats, suggesting centromeric and telomeric loss of heterochromatinization and thus chromosomal instability [80,81]. In particular, it has been previously demonstrated that the overexpression of alpha satellite transcripts leads to chromosomal instability in breast cancer via segregation errors [82].

It is interesting that each histotype and, more generally, each specimen showed a specific altered pattern of transcripts arising from RS and in detail from satellite repeats, suggesting that the altered transcription of RS could be something more than an epiphenomenon and may indicate the peculiar characteristic of the specimens, such as an effect of a previous viral infection [83]. The specificity of RS transcription is supported by a case of ‘luminal-B HER2-negative’ in which the transcripts derived from BSR (beta satellite repeats) are thousands of times more upregulated in comparison with its background (Figure 2), a unique case among all the cases analyzed in this paper and others [84]. It is noteworthy that this case has the worst classification in TNM staging and diffuse lymph node metastases. The increase may be due to a true increase in BSR expression or a significant increase in BSR-derived transcript polyadenylation and stability. Indeed, satellite polyadenylation has been postulated in humans, following studies on other organisms [85], and evidence is now arising and being consolidated [86]. The analysis of RS transcription suggests that the current molecular classification of breast cancers, even if functional in defining the therapeutic course [9], is far from being exhaustive in defining their molecular characteristics.

This is in line with the great variability of prognosis and clinical course in the same class of breast cancers [86].

We also report that the vast majority of the specimens analyzed showed an altered expression of GGAAT repeats in comparison with their specific ANT, either increased (GGAAT^{higher}) or decreased (GGAAT^{lower}). The GGAAT^{higher} specimens have a higher ki67 staining [87], a widely used marker of cell proliferation, pointing to a faster growing tumor mass. Indeed, the two specimens that had the worst TNM classification (IIIb) and an evident lymph node involvement are GGAAT^{higher}.

Regarding the transcriptional background, comparing the healthy adjacent tissues of cancer specimens with tissues from healthy donors, there is a generalized downregulation of MER22 satellite expression. MER22 has been implied in meiotic instability [76,88], and its methylation status has been found relevant to multiple cancer types [89–92].

Interestingly, ALR satellites are upregulated in ANTs. Considering that ALR expression is also upregulated in tumors in comparison with non-cancerous specimens, this suggests the role of these satellites in the progression of the disease.

Overall, the altered transcriptional landscape of RS in the background of patients may suggest either a genetic predisposition and increased susceptibility to cancer transformation or it could be the result of epigenetic alteration due to environmental factors (e.g., exposition to chemicals or previous infections), which ultimately need to be investigated further.

5. Conclusions

The analysis of transcripts arising from RS in breast cancer specimens classified following the current molecular markers showed a great interspecimen variability, with peculiar patterns of altered RS transcription. Overall, there is an evident alteration in the transcripts arising from satellite repeats and, in specific cases, from SINE and endogenous retrovirus sequences. The expression from healthy adjacent tissues of cancer specimens showed an altered expression of RS transcription when compared to the transcription of healthy donors. If the data presented here are confirmed and extended in a larger population, RS expression may play a role in the molecular classification and stratification of patients or even be potentially adopted as a biomarker in liquid biopsy.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cells11162522/s1>, Table S1: Raw expression data; Table S2: Normalized expression data of the merged raw counts; Table S3: A comparison between the 15 invasive breast cancer specimens and 18 controls; Webpage S4: Deseq2 results of the comparison between the 15 invasive breast cancer specimens and 18 controls; Table S5: HER2-enriched analysis; Table S6: Luminal-A analysis; Table S7: Luminal-B HER2-negative analysis; Table S8: Luminal-B HER2-positive analysis; Table S9: triple-negative breast cancer analysis; Table S10: Comparison between the expression of ANT and normal controls.

Author Contributions: Conceptualization, W.A.; methodology, W.A. and C.C.; resources, C.C.; writing—original draft preparation, W.A.; writing—review and editing, C.C.; supervision, C.C.; project administration, C.C.; funding acquisition, C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Sicilian Region funded by ERDF (FESR) 2014/2020–Action 1.1.5–Project “OBIND”–CUP G29J18000700007.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The analyzed data are published in the European Nucleotide Archive (ENA), RRID:SCR_006515, study accession: PRJNA292118 [4,5].

Acknowledgments: Not Applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Akram, M.; Iqbal, M.; Daniyal, M.; Khan, A.U. Awareness and Current Knowledge of Breast Cancer. *Biol. Res.* **2017**, *50*, 33. [[CrossRef](#)] [[PubMed](#)]
2. Tan, P.H.; Ellis, I.; Allison, K.; Brogi, E.; Fox, S.B.; Lakhani, S.; Lazar, A.J.; Morris, E.A.; Sahin, A.; Salgado, R.; et al. The 2019 World Health Organization Classification of Tumours of the Breast. *Histopathology* **2020**, *77*, 181–185. [[CrossRef](#)] [[PubMed](#)]
3. Di Napoli, A.; Jain, P.; Duranti, E.; Margolskee, E.; Arancio, W.; Facchetti, F.; Alobeid, B.; Santanelli di Pompeo, F.; Mansukhani, M.; Bhagat, G. Targeted next Generation Sequencing of Breast Implant-Associated Anaplastic Large Cell Lymphoma Reveals Mutations in JAK/STAT Signalling Pathway Genes, TP53 and DNMT3A. *Br. J. Haematol.* **2018**, *180*, 741–744. [[CrossRef](#)] [[PubMed](#)]
4. Xu, S.; Kong, D.; Chen, Q.; Ping, Y.; Pang, D. Oncogenic Long Noncoding RNA Landscape in Breast Cancer. *Mol. Cancer* **2017**, *16*, 129. [[CrossRef](#)] [[PubMed](#)]
5. Pang, B.; Wang, Q.; Ning, S.; Wu, J.; Zhang, X.; Chen, Y.; Xu, S. Landscape of Tumor Suppressor Long Noncoding RNAs in Breast Cancer. *J. Exp. Clin. Cancer Res.* **2019**, *38*, 79. [[CrossRef](#)] [[PubMed](#)]
6. Goldhirsch, A.; Wood, W.C.; Coates, A.S.; Gelber, R.D.; Thürlimann, B.; Senn, H.J. Strategies for Subtypes-Dealing with the Diversity of Breast Cancer: Highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann. Oncol.* **2011**, *22*, 1736–1747. [[CrossRef](#)]
7. Goldhirsch, A.; Winer, E.P.; Coates, A.S.; Gelber, R.D.; Piccart-Gebhart, M.; Thürlimann, B.; Senn, H.J.; Albain, K.S.; André, F.; Bergh, J.; et al. Personalizing the Treatment of Women with Early Breast Cancer: Highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann. Oncol.* **2013**, *24*, 2206–2223. [[CrossRef](#)] [[PubMed](#)]
8. Hennigs, A.; Riedel, F.; Gondos, A.; Sinn, P.; Schirmacher, P.; Marmé, F.; Jäger, D.; Kauczor, H.U.; Stieber, A.; Lindel, K.; et al. Prognosis of Breast Cancer Molecular Subtypes in Routine Clinical Care: A Large Prospective Cohort Study. *BMC Cancer* **2016**, *16*, 734. [[CrossRef](#)]
9. Prat, A.; Pineda, E.; Adamo, B.; Galván, P.; Fernández, A.; Gaba, L.; Díez, M.; Viladot, M.; Arance, A.; Muñoz, M. Clinical Implications of the Intrinsic Molecular Subtypes of Breast Cancer. *Breast* **2015**, *24*, S26–S35. [[CrossRef](#)] [[PubMed](#)]
10. Clark, M.B.; Amaral, P.P.; Schlesinger, F.J.; Dinger, M.E.; Taft, R.J.; Rinn, J.L.; Ponting, C.P.; Stadler, P.F.; Morris, K.V.; Morillon, A.; et al. The Reality of Pervasive Transcription. *PLoS Biol.* **2011**, *9*, e1000625. [[CrossRef](#)] [[PubMed](#)]
11. Arancio, W.; Coronello, C. Repetitive Sequences in Aging. *Aging* **2021**, *13*, 10816–10817. [[CrossRef](#)] [[PubMed](#)]
12. Hirschberger, S.; Hinske, L.C.; Kretz, S. MiRNAs: Dynamic Regulators of Immune Cell Functions in Inflammation and Cancer. *Cancer Lett.* **2018**, *431*, 11–21. [[CrossRef](#)] [[PubMed](#)]
13. Li, M.; Cui, X.; Guan, H. MicroRNAs: Pivotal Regulators in Acute Myeloid Leukemia. *Ann. Hematol.* **2020**, *99*, 399–412. [[CrossRef](#)] [[PubMed](#)]
14. Liu, J.; Ke, F.; Chen, T.; Zhou, Q.; Weng, L.; Tan, J.; Shen, W.; Li, L.; Zhou, J.; Xu, C.; et al. MicroRNAs That Regulate PTEN as Potential Biomarkers in Colorectal Cancer: A Systematic Review. *J. Cancer Res. Clin. Oncol.* **2020**, *146*, 809–820. [[CrossRef](#)] [[PubMed](#)]
15. Arancio, W.; Calogero Amato, M.; Magliozzo, M.; Pizzolanti, G.; Vesco, R.; Giordano, C. Serum MiRNAs in Women Affected by Hyperandrogenic Polycystic Ovary Syndrome: The Potential Role of MiR-155 as a Biomarker for Monitoring the Estroprogestinic Treatment. *Gynecol. Endocrinol.* **2018**, *34*, 704–708. [[CrossRef](#)] [[PubMed](#)]
16. Hashemi, A.; Gorji-bahri, G. MicroRNA: Promising Roles in Cancer Therapy. *Curr. Pharm. Biotechnol.* **2020**, *21*, 1186–1203. [[CrossRef](#)] [[PubMed](#)]
17. Lou, W.; Ding, B.; Fu, P. Pseudogene-Derived LncRNAs and Their MiRNA Sponging Mechanism in Human Cancer. *Front. Cell Dev. Biol.* **2020**, *8*, 85. [[CrossRef](#)] [[PubMed](#)]
18. Arancio, W.; Carina, V.; Pizzolanti, G.; Tomasello, L.; Pitrone, M.; Baiamonte, C.; Amato, M.C.; Giordano, C. Anaplastic Thyroid Carcinoma: A CeRNA Analysis Pointed to a Crosstalk between SOX2, TP53, and MicroRNA Biogenesis. *Int. J. Endocrinol.* **2015**, *2015*, 439370. [[CrossRef](#)] [[PubMed](#)]
19. Poliseno, L.; Pandolfi, P.P. PTEN CeRNA Networks in Human Cancer. *Methods* **2015**, *77–78*, 41–50. [[CrossRef](#)] [[PubMed](#)]
20. Arancio, W.; Genovese, S.I.; Bongiovanni, L.; Tripodo, C. A CeRNA Approach May Unveil Unexpected Contributors to Deletion Syndromes, the Model of 5q-Syndrome. *Oncoscience* **2015**, *2*, 872–879. [[CrossRef](#)] [[PubMed](#)]
21. Arancio, W.; Giordano, C.; Pizzolanti, G. A CeRNA Analysis on LMNA Gene Focusing on the Hutchinson-Gilford Progeria Syndrome. *J. Clin. Bioinform.* **2013**, *3*, 2. [[CrossRef](#)] [[PubMed](#)]
22. Arancio, W. A Bioinformatics Analysis of Lamin-A Regulatory Network: A Perspective on Epigenetic Involvement in Hutchinson-Gilford Progeria Syndrome. *Rejuvenation Res.* **2012**, *15*, 123–127. [[CrossRef](#)] [[PubMed](#)]
23. Arancio, W. CeRNA Analysis of SARS-CoV-2. *Arch. Virol.* **2021**, *166*, 271–274. [[CrossRef](#)] [[PubMed](#)]
24. Bertolazzi, G.; Cipollina, C.; Benos, P.V.; Tumminello, M.; Coronello, C. MiR-1207-5p Can Contribute to Dysregulation of Inflammatory Response in COVID-19 via Targeting SARS-CoV-2 RNA. *Front. Cell. Infect. Microbiol.* **2020**, *10*, 586592. [[CrossRef](#)] [[PubMed](#)]
25. Faulkner, G.J.; Kimura, Y.; Daub, C.O.; Wani, S.; Plessy, C.; Irvine, K.M.; Schroder, K.; Cloonan, N.; Steptoe, A.L.; Lassmann, T.; et al. The Regulated Retrotransposon Transcriptome of Mammalian Cells. *Nat. Genet.* **2009**, *41*, 563–571. [[CrossRef](#)] [[PubMed](#)]
26. Clayton, E.A.; Wang, L.; Rishishwar, L.; Wang, J.; McDonald, J.F.; Jordan, I.K. Patterns of Transposable Element Expression and Insertion in Cancer. *Front. Mol. Biosci.* **2016**, *3*, 76. [[CrossRef](#)]

27. Sciamanna, I.; De Luca, C.; Spadafora, C. The Reverse Transcriptase Encoded by LINE-1 Retrotransposons in the Genesis, Progression, and Therapy of Cancer. *Front. Chem.* **2016**, *4*, 6. [[CrossRef](#)]
28. Ohms, S.; Lee, S.H.; Rangasamy, D. LINE-1 Retrotransposons and Let-7 MiRNA: Partners in the Pathogenesis of Cancer? *Front. Genet.* **2014**, *5*, 338. [[CrossRef](#)]
29. Tubio, J.M.C.; Li, Y.; Ju, Y.S.; Martincorena, I.; Cooke, S.L.; Tojo, M.; Gundem, G.; Pipinikas, C.P.; Zamora, J.; Raine, K.; et al. Extensive Transduction of Nonrepetitive DNA Mediated by L1 Retrotransposition in Cancer Genomes. *Science* **2014**, *345*, 1251343. [[CrossRef](#)] [[PubMed](#)]
30. Di Ruocco, F.; Basso, V.; Rivoire, M.; Mehlen, P.; Ambati, J.; De Falco, S.; Tarallo, V. Alu RNA Accumulation Induces Epithelial-to-Mesenchymal Transition by Modulating MiR-566 and Is Associated with Cancer Progression. *Oncogene* **2018**, *37*, 627–637. [[CrossRef](#)]
31. Padeken, J.; Zeller, P.; Gasser, S.M. Repeat DNA in Genome Organization and Stability. *Curr. Opin. Genet. Dev.* **2015**, *31*, 12–19. [[CrossRef](#)]
32. Treangen, T.J.; Salzberg, S.L. Repetitive DNA and Next-Generation Sequencing: Computational Challenges and Solutions. *Nat. Rev. Genet.* **2012**, *13*, 36–46. [[CrossRef](#)] [[PubMed](#)]
33. Coufal, N.G.; Garcia-Perez, J.L.; Peng, G.E.; Yeo, G.W.; Mu, Y.; Lovci, M.T.; Morell, M.; O’Shea, K.S.; Moran, J.V.; Gage, F.H. L1 Retrotransposition in Human Neural Progenitor Cells. *Nature* **2009**, *460*, 1127–1131. [[CrossRef](#)] [[PubMed](#)]
34. Arancio, W. Progerin Expression Induces a Significant Downregulation of Transcription from Human Repetitive Sequences in iPSC-Derived Dopaminergic Neurons. *GeroScience* **2019**, *41*, 39–49. [[CrossRef](#)] [[PubMed](#)]
35. Zhu, Q.; Hoong, N.; Aslanian, A.; Hara, T.; Benner, C.; Heinz, S.; Miga, K.H.; Ke, E.; Verma, S.; Soroczynski, J.; et al. Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. *Mol. Cell* **2018**, *70*, 842–853.e7. [[CrossRef](#)]
36. Kakizawa, N.; Suzuki, K.; Abe, I.; Endo, Y.; Tamaki, S.; Ishikawa, H.; Watanabe, F.; Ichida, K.; Saito, M.; Futsuhara, K.; et al. High Relative Levels of Satellite Alpha Transcripts Predict Increased Risk of Bilateral Breast Cancer and Multiple Primary Cancer in Patients with Breast Cancer and Lacking BRCA-Related Clinical Features. *Oncol. Rep.* **2019**, *42*, 857–865. [[CrossRef](#)]
37. Patnala, R.; Lee, S.H.; Dahlstrom, J.E.; Ohms, S.; Chen, L.; Dheen, S.T.; Rangasamy, D. Inhibition of LINE-1 Retrotransposon-Encoded Reverse Transcriptase Modulates the Expression of Cell Differentiation Genes in Breast Cancer Cells. *Breast Cancer Res. Treat.* **2014**, *143*, 239–253. [[CrossRef](#)]
38. Park, S.Y.; Seo, A.N.; Jung, H.Y.; Gwak, J.M.; Jung, N.; Cho, N.Y.; Kang, G.H. Alu and LINE-1 Hypomethylation Is Associated with HER2 Enriched Subtype of Breast Cancer. *PLoS ONE* **2014**, *9*, e100429. [[CrossRef](#)]
39. Van Hoesel, A.Q.; Van De Velde, C.J.H.; Kuppen, P.J.K.; Liefers, G.J.; Putter, H.; Sato, Y.; Elashoff, D.A.; Turner, R.R.; Shamoni, J.M.; De Kruijf, E.M.; et al. Hypomethylation of LINE-1 in Primary Tumor Has Poor Prognosis in Young Breast Cancer Patients: A Retrospective Cohort Study. *Breast Cancer Res. Treat.* **2012**, *134*, 1103–1114. [[CrossRef](#)]
40. Harris, C.R.; Normart, R.; Yang, Q.; Stevenson, E.; Haffty, B.G.; Ganesan, S.; Cordon-Cardo, C.; Levine, A.J.; Tang, L.H. Association of Nuclear Localization of a Long Interspersed Nuclear Element-1 Protein in Breast Tumors with Poor Prognostic Outcomes. *Genes and Cancer* **2010**, *1*, 115–124. [[CrossRef](#)]
41. Gualtieri, A.; Andreola, F.; Sciamanna, I.; Sinibaldi-Vallebona, P.; Serafino, A.; Spadafora, C. Increased Expression and Copy Number Amplification of LINE-1 and SINE B1 Retrotransposable Elements in Murine Mammary Carcinoma Progression. *Oncotarget* **2013**, *4*, 1882–1893. [[CrossRef](#)] [[PubMed](#)]
42. Chen, L.; Dahlstrom, J.E.; Chandra, A.; Board, P.; Rangasamy, D. Prognostic Value of LINE-1 Retrotransposon Expression and Its Subcellular Localization in Breast Cancer. *Breast Cancer Res. Treat.* **2012**, *136*, 129–142. [[CrossRef](#)]
43. Miret, N.; Zappia, C.D.; Altamirano, G.; Pontillo, C.; Zárata, L.; Gómez, A.; Lasagna, M.; Cocca, C.; Kass, L.; Monczor, F.; et al. AhR Ligands Reactivate LINE-1 Retrotransposon in Triple-Negative Breast Cancer Cells MDA-MB-231 and Non-Tumorigenic Mammary Epithelial Cells NMuMG. *Biochem. Pharmacol.* **2020**, *175*, 113904. [[CrossRef](#)] [[PubMed](#)]
44. Miglio, U.; Berrino, E.; Panero, M.; Ferrero, G.; Coscujuela Tarrero, L.; Miano, V.; Dell’Aglia, C.; Sarotto, I.; Annaratone, L.; Marchiò, C.; et al. The Expression of LINE1-MET Chimeric Transcript Identifies a Subgroup of Aggressive Breast Cancers. *Int. J. Cancer* **2018**, *143*, 2838–2848. [[CrossRef](#)] [[PubMed](#)]
45. Bratthauer, G.L.; Cardiff, R.D.; Fanning, T.G. Expression of LINE-1 Retrotransposons in Human Breast Cancer. *Cancer* **1994**, *73*, 2333–2336. [[CrossRef](#)]
46. Wang, Y.; Bernhardt, A.J.; Nacson, J.; Kraus, J.J.; Tan, Y.F.; Nicolas, E.; Radke, M.R.; Handorf, E.; Llop-Guevara, A.; Balmaña, J.; et al. BRCA1 Intronic Alu Elements Drive Gene Rearrangements and PARP Inhibitor Resistance. *Nat. Commun.* **2019**, *10*, 5661. [[CrossRef](#)] [[PubMed](#)]
47. Staaf, J.; Glodzik, D.; Bosch, A.; Vallon-Christersson, J.; Reuterswärd, C.; Häkkinen, J.; Degasperi, A.; Amarante, T.D.; Saal, L.H.; Hegardt, C.; et al. Whole-Genome Sequencing of Triple-Negative Breast Cancers in a Population-Based Clinical Study. *Nat. Med.* **2019**, *25*, 1526–1533. [[CrossRef](#)]
48. Felicio, P.S.; Alemar, B.; Coelho, A.S.; Berardinelli, G.N.; Melendez, M.E.; Lengert, A.V.H.; Miche Ili, R.D.; Reis, R.M.; Fernandes, G.C.; Ewald, I.P.; et al. Screening and Characterization of BRCA2 c.156_157insAlu in Brazil: Results from 1380 Individuals from the South and Southeast. *Cancer Genet.* **2018**, *228–229*, 93–97. [[CrossRef](#)]
49. Rizza, R.; Hackmann, K.; Paris, I.; Minucci, A.; De Leo, R.; Schrock, E.; Urbani, A.; Capoluongo, E.; Gelli, G.; Concolino, P. Novel BRCA1 Large Genomic Rearrangements in Italian Breast/Ovarian Cancer Patients. *Mol. Diagnosis Ther.* **2019**, *23*, 121–126. [[CrossRef](#)] [[PubMed](#)]

50. Gallegos-Arreola, M.P.; Figuera, L.E.; Flores-Ramos, L.G.; Puebla-Pérez, A.M.; Zúñiga-González, G.M. Association of the Alu Insertion Polymorphism in the Progesterone Receptor Gene with Breast Cancer in a Mexican Population. *Arch. Med. Sci.* **2015**, *11*, 551–560. [[CrossRef](#)] [[PubMed](#)]
51. Machado, P.M.; Brandão, R.D.; Cavaco, B.M.; Eugénio, J.; Bento, S.; Nave, M.; Rodrigues, P.; Fernandes, A.; Vaz, F. Screening for a BRCA2 Rearrangement in High-Risk Breast/Ovarian Cancer Families: Evidence for a Founder Effect and Analysis of the Associated Phenotypes. *J. Clin. Oncol.* **2007**, *25*, 2027–2034. [[CrossRef](#)] [[PubMed](#)]
52. Wu, Y.; Zhao, W.; Liu, Y.; Tan, X.; Li, X.; Zou, Q.; Xiao, Z.; Xu, H.; Wang, Y.; Yang, X. Function of HNRNPC in Breast Cancer Cells by Controlling the DsRNA-induced Interferon Response. *EMBO J.* **2018**, *37*, e99017. [[CrossRef](#)] [[PubMed](#)]
53. Grabski, D.F.; Hu, Y.; Sharma, M.; Rasmussen, S.K. Close to the Bedside: A Systematic Review of Endogenous Retroviruses and Their Impact in Oncology. *J. Surg. Res.* **2019**, *240*, 145–155. [[CrossRef](#)]
54. Zhao, J.; Rycaj, K.; Geng, S.; Li, M.; Plummer, J.B.; Yin, B.; Liu, H.; Xu, X.; Zhang, Y.; Yan, Y.; et al. Expression of Human Endogenous Retrovirus Type K Envelope Protein Is a Novel Candidate Prognostic Marker for Human Breast Cancer. *Genes Cancer* **2011**, *2*, 914–922. [[CrossRef](#)] [[PubMed](#)]
55. Wang-Johanning, F.; Radvanyi, L.; Rycaj, K.; Plummer, J.B.; Yan, P.; Sastry, K.J.; Piyathilake, C.J.; Hunt, K.K.; Johanning, G.L. Human Endogenous Retrovirus K Triggers an Antigen-Specific Immune Response in Breast Cancer Patients. *Cancer Res.* **2008**, *68*, 5869–5877. [[CrossRef](#)]
56. Golan, M.; Hizi, A.; Resau, J.H.; Yaal-Hahoshen, N.; Reichman, H.; Keydar, I.; Tsarfaty, I. Human Endogenous Retrovirus (HERV-K) Reverse Transcriptase as a Breast Cancer Prognostic Marker. *Neoplasia* **2008**, *10*, 521–533. [[CrossRef](#)]
57. Wang-Johanning, F.; Frost, A.R.; Jian, B.; Epp, L.; Lu, D.W.; Johanning, G.L. Quantitation of HERV-K Env Gene Expression and Splicing in Human Breast Cancer. *Oncogene* **2003**, *22*, 1528–1535. [[CrossRef](#)] [[PubMed](#)]
58. Wang-Johanning, F.; Frost, A.R.; Johanning, G.L.; Khazaeli, M.B.; LoBuglio, A.F.; Shaw, D.R.; Strong, T.V. Expression of Human Endogenous Retrovirus K Envelope Transcripts in Human Breast Cancer. *Clin. Cancer Res.* **2001**, *7*, 1553–1560. [[PubMed](#)]
59. Matteucci, C.; Balestrieri, E.; Argaw-Denboba, A.; Sinibaldi-Vallebona, P. Human Endogenous Retroviruses Role in Cancer Cell Stemness. *Semin. Cancer Biol.* **2018**, *53*, 17–30. [[CrossRef](#)]
60. Salmons, B.; Lawson, J.S.; Günzburg, W.H. Recent Developments Linking Retroviruses to Human Breast Cancer: Infectious Agent, Enemy within or Both? *J. Gen. Virol.* **2014**, *95*, 2589–2593. [[CrossRef](#)] [[PubMed](#)]
61. Nguyen, T.D.; Davis, J.; Eugenio, R.A.; Liu, Y. Female Sex Hormones Activate Human Endogenous Retrovirus Type K Through the OCT4 Transcription Factor in T47D Breast Cancer Cells. *AIDS Res. Hum. Retrovir.* **2019**, *35*, 348–356. [[CrossRef](#)]
62. Johanning, G.L.; Malouf, G.G.; Zheng, X.; Esteva, F.J.; Weinstein, J.N.; Wang-Johanning, F.; Su, X. Expression of Human Endogenous Retrovirus-K Is Strongly Associated with the Basal-like Breast Cancer Phenotype. *Sci. Rep.* **2017**, *7*, 41960. [[CrossRef](#)] [[PubMed](#)]
63. Jin, X.; Xu, X.E.; Jiang, Y.Z.; Liu, Y.R.; Sun, W.; Guo, Y.J.; Ren, Y.X.; Zuo, W.J.; Hu, X.; Huang, S.L.; et al. The Endogenous Retrovirus-Derived Long Noncoding RNA TROJAN Promotes Triple-Negative Breast Cancer Progression via ZMYND8 Degradation. *Sci. Adv.* **2019**, *5*, eaat9820. [[CrossRef](#)] [[PubMed](#)]
64. Lemaître, C.; Tsang, J.; Bireau, C.; Heidmann, T.; Dewannieux, M. 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*, e1006451. [[CrossRef](#)] [[PubMed](#)]
65. Zhou, F.; Li, M.; Wei, Y.; Lin, K.; Lu, Y.; Shen, J.; Johanning, G.L.; Wang-Johanning, F. Activation of HERV-K Env Protein Is Essential for Tumorigenesis and Metastasis of Breast Cancer Cells. *Oncotarget* **2016**, *7*, 84093–84117. [[CrossRef](#)] [[PubMed](#)]
66. Wang-Johanning, F.; Li, M.; Esteva, F.J.; Hess, K.R.; Yin, B.; Rycaj, K.; Plummer, J.B.; Garza, J.G.; Ambs, S.; Johanning, G.L. Human Endogenous Retrovirus Type K Antibodies and mRNA as Serum Biomarkers of Early-Stage Breast Cancer. *Int. J. Cancer* **2014**, *134*, 587–595. [[CrossRef](#)] [[PubMed](#)]
67. Jiang, J.C.; Upton, K.R. Human Transposons Are an Abundant Supply of Transcription Factor Binding Sites and Promoter Activities in Breast Cancer Cell Lines. *Mob. DNA* **2019**, *10*, 16. [[CrossRef](#)] [[PubMed](#)]
68. Presneau, N.; Laplace-Marieze, V.; Sylvain, V.; Lortholary, A.; Hardouin, A.; Bernard-Gallon, D.; Bignon, Y.J. New Mechanism of BRCA-1 Mutation by Deletion/Insertion at the Same Nucleotide Position in Three Unrelated French Breast/Ovarian Cancer Families. *Hum. Genet.* **1998**, *103*, 334–339. [[CrossRef](#)]
69. Jalili, V.; Afgan, E.; Gu, Q.; Clements, D.; Blankenberg, D.; Goecks, J.; Taylor, J.; Nekrutenko, A. The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2020 Update. *Nucleic Acids Res.* **2020**, *48*, W395–W402. [[CrossRef](#)]
70. Afgan, E.; Baker, D.; Batut, B.; Van Den Beek, M.; Bouvier, D.; Ech, M.; Chilton, J.; Clements, D.; Coraor, N.; Grünig, B.A.; et al. The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2018 Update. *Nucleic Acids Res.* **2018**, *46*, W537–W544. [[CrossRef](#)]
71. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
72. Langmead, B.; Salzberg, S.L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)] [[PubMed](#)]
73. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A Fast Spliced Aligner with Low Memory Requirements. *Nat. Methods* **2015**, *12*, 357–360. [[CrossRef](#)]
74. Love, M.I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)] [[PubMed](#)]

75. Haynes, W. Benjamini–Hochberg Method. In *Encyclopedia of Systems Biology*; Springer: New York, NY, USA, 2013.
76. Available online: from <https://www.ncbi.nlm.nih.gov/sra/SRX1135937> to <https://www.ncbi.nlm.nih.gov/sra/SRX1135969>; (accessed on 19 July 2022).
77. Hoyt, S.J.; Storer, J.M.; Hartley, G.A.; Grady, P.G.S.; Gershman, A.; de Lima, L.G.; Limouse, C.; Halabian, R.; Wojenski, L.; Rodriguez, M.; et al. From Telomere to Telomere: The Transcriptional and Epigenetic State of Human Repeat Elements. *Science* **2022**, *376*, eabk3112. [[CrossRef](#)]
78. Fatyol, K.; Illes, K.; Diamond, D.C.; Janish, C.; Szalay, A.A. Mer22-Related Sequence Elements Form Pericentric Repetitive DNA Families in Primates. *Mol. Gen. Genet.* **2000**, *262*, 931–939. [[CrossRef](#)]
79. Hancks, D.C.; Kazazian, H.H. SVA Retrotransposons: Evolution and Genetic Instability. *Semin. Cancer Biol.* **2010**, *20*, 234–245. [[CrossRef](#)] [[PubMed](#)]
80. Black, E.M.; Giunta, S. Repetitive Fragile Sites: Centromere Satellite DNA as a Source of Genome Instability in Human Diseases. *Genes* **2018**, *9*, 615. [[CrossRef](#)]
81. Ferreira, D.; Meles, S.; Escudeiro, A.; Mendes-da-Silva, A.; Adegas, F.; Chaves, R. Satellite Non-Coding RNAs: The Emerging Players in Cells, Cellular Pathways and Cancer. *Chromosom. Res.* **2015**, *23*, 479–493. [[CrossRef](#)] [[PubMed](#)]
82. Ichida, K.; Suzuki, K.; Fukui, T.; Takayama, Y.; Kakizawa, N.; Watanabe, F.; Ishikawa, H.; Muto, Y.; Kato, T.; Saito, M.; et al. Overexpression of Satellite Alpha Transcripts Leads to Chromosomal Instability via Segregation Errors at Specific Chromosomes. *Int. J. Oncol.* **2018**, *52*, 1685–1693. [[CrossRef](#)]
83. Nogalski, M.T.; Solovyov, A.; Kulkarni, A.S.; Desai, N.; Oberstein, A.; Levine, A.J.; Ting, D.T.; Shenk, T.; Greenbaum, B.D. A Tumor-Specific Endogenous Repetitive Element Is Induced by Herpesviruses. *Nat. Commun.* **2019**, *10*, 90. [[CrossRef](#)] [[PubMed](#)]
84. Yandım, C.; Karakülah, G. Dysregulated Expression of Repetitive DNA in ER+/HER2- Breast Cancer. *Cancer Genet.* **2019**, *239*, 36–45. [[CrossRef](#)] [[PubMed](#)]
85. Rošić, S.; Köhler, F.; Erhardt, S. Repetitive Centromeric Satellite RNA Is Essential for Kinetochore Formation and Cell Division. *J. Cell Biol.* **2014**, *207*, 335–349. [[CrossRef](#)] [[PubMed](#)]
86. Kalinowski, L.; Saunus, J.M.; McCart Reed, A.E.; Lakhani, S.R. Breast Cancer Heterogeneity in Primary and Metastatic Disease. *Adv Exp Med Biol.* **2019**, *1152*, 75–104. [[CrossRef](#)] [[PubMed](#)]
87. Li, L.T.; Jiang, G.; Chen, Q.; Zheng, J.N. Predict Ki67 Is a Promising Molecular Target in the Diagnosis of Cancer (Review). *Mol. Med. Rep.* **2015**, *11*, 1566–1572. [[CrossRef](#)]
88. Tremblay, D.C.; Alexander, G.; Moseley, S.; Chadwick, B.P. Expression, Tandem Repeat Copy Number Variation and Stability of Four Macrosatellite Arrays in the Human Genome. *BMC Genom.* **2010**, *11*, 632. [[CrossRef](#)] [[PubMed](#)]
89. Nagai, H.; Kim, Y.S.; Yasuda, T.; Ohmachi, Y.; Yokouchi, H.; Monden, M.; Emi, M.; Konishi, N.; Nogami, M.; Okumura, K.; et al. A Novel Sperm-Specific Hypomethylation Sequence Is a Demethylation Hotspot in Human Hepatocellular Carcinomas. *Gene* **1999**, *237*, 15–20. [[CrossRef](#)]
90. Suzuki, K.; Suzuki, I.; Leodolter, A.; Alonso, S.; Horiuchi, S.; Yamashita, K.; Perucho, M. Global DNA Demethylation in Gastrointestinal Cancer Is Age Dependent and Precedes Genomic Damage. *Cancer Cell* **2006**, *9*, 199–207. [[CrossRef](#)]
91. Igarashi, S.; Suzuki, H.; Niinuma, T.; Shimizu, H.; Nojima, M.; Iwaki, H.; Nobuoka, T.; Nishida, T.; Miyazaki, Y.; Takamaru, H.; et al. A Novel Correlation between LINE-1 Hypomethylation and the Malignancy of Gastrointestinal Stromal Tumors. *Clin. Cancer Res.* **2010**, *16*, 5114–5123. [[CrossRef](#)]
92. Samuelsson, J.K.; Dumbovic, G.; Polo, C.; Moreta, C.; Alibés, A.; Ruiz-Larroya, T.; Giménez-Bonafé, P.; Alonso, S.; Forcales, S.V.; Perucho, M. Helicase Lymphoid-Specific Enzyme Contributes to the Maintenance of Methylation of Sst1 Pericentromeric Repeats That Are Frequently Demethylated in Colon Cancer and Associate with Genomic Damage. *Epigenomes* **2017**, *1*, 2. [[CrossRef](#)]