

Article

Dysregulation of Locus-Specific Repetitive Elements in TCGA Pan-Cancers

Chao Wang  and Chun Liang *

Department of Biology, Miami University, Oxford, OH 45056, USA; wangc90@miamioh.edu

* Correspondence: liangc@miamioh.edu

Abstract: Background: Understanding the role of repetitive elements (REs) in cancer development is crucial for identifying novel biomarkers and therapeutic targets. **Methods:** This study investigated the locus-specific dysregulation of REs, including the differential expression and methylation of REs, across 12 TCGA cancer types stratified by their genomic context (i.e., genic and intergenic REs). **Results:** We found uniquely dysregulated genic REs co-regulated with their corresponding transcripts and associated with distinct biological functions in different cancer types. Uniquely dysregulated intergenic REs were identified in each cancer type and used to cluster different sample types. Recurrently dysregulated REs were identified in several cancer types, with genes associated with up-regulated genic REs involved in cell cycle processes and those associated with down-regulated REs involved in the extracellular matrix. Interestingly, four out of five REs consistently down-regulated in all 12 cancer types were located in the intronic region of the *TMEM252*, a recently discovered tumor suppressor gene. *TMEM252* expression was also down-regulated in 10 of 12 cancer types, suggesting its potential importance across a wide range of cancer types. With the corresponding DNA methylation array data, we found a higher prevalence of hypo-methylated REs in most cancer types (10 out of 12). Despite the slight overlaps between differentially expressed REs and differentially methylated REs, we showed that the methylation of locus-specific REs negatively correlates with their expression in some of these 12 cancer types. **Conclusions:** Our findings highlight the cancer-specific and recurrent deregulation of REs, their functional associations, and the potential role of *TMEM252* as a pan-cancer tumor suppressor, providing new insights into biomarker discovery and therapeutic development.



Academic Editor: Jung-Kyoon Choi

Received: 5 April 2025

Revised: 24 April 2025

Accepted: 27 April 2025

Published: 29 April 2025

Citation: Wang, C.; Liang, C.

Dysregulation of Locus-Specific

Repetitive Elements in TCGA

Pan-Cancers. *Genes* **2025**, *16*, 528.[https://doi.org/10.3390/](https://doi.org/10.3390/genes16050528)[genes16050528](https://doi.org/10.3390/genes16050528)**Copyright:** © 2025 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

[\(https://creativecommons.org/](https://creativecommons.org/licenses/by/4.0/)[licenses/by/4.0/\)](https://creativecommons.org/licenses/by/4.0/).**Keywords:** repetitive elements; dysregulations; pan-cancers; tumor suppressor gene

1. Introduction

1.1. Repetitive Elements (REs) and Their Regulations in the Human Genome

Repetitive elements (REs) are the most abundant type of sequences in the human genome [1] can be classified into satellites (or tandem repeats) and transposable elements (TEs), with the latter further subdivided into RNA transposable elements and DNA transposable elements. REs are typically hierarchically classified with increasing granularity, from class, family, and subfamily (corresponding to repClass, repFamily, and repName hierarchies defined by the human Repeatmasker [2]) to locus-specific elements based on sequence similarities [3]. As arrays of repeated nucleotides, satellites can be further classified based on the increasing size of the repeated unit into microsatellites, minisatellites, satellites, and macrosatellites [4]. Due to their highly variable array size they are also known as variable number of tandem repeats (VNTRs). Most satellites are composed of

either simple repeats or complex repeats with increased repeat unit length and complexity, mainly appearing at centromeres, pericentromeric regions and subtelomeric regions. While the tandem repeats are primarily found in centromeres and telomeres, TEs and their evolutionary relics are scattered throughout the genome [5]. The most common classes of RNA transposable elements in the human genome include long terminal repeat (LTR), long interspersed nuclear element (LINE), short interspersed nuclear element (SINE), and Retroposon (LINE, SINE, and Retroposon belong to non-LTR), and these use RNA as their intermediate for the transposition [3]. DNA transposable elements in the human genome mainly consist of DNA transposons and RC (rolling circles or helitrons) [3]. Instead of using RNA intermediate for their transposition, DNA transposons mostly use the “cut-and-paste” mechanism for their propagation [6], with the RC using the rolling-circle intermediates for their transposition [7].

LINE-1 (a family of LINE class) and Alu (a family of SINE class) have been extensively studied in human genomes. LINE-1 makes up approximately 17% of the human genome, and it has been estimated that there are more than one million Alu repeats in the human genome [8]. The intact sequence for LINE-1 is about 6 kb, containing protein domains encoded by two open reading frames (ORFs), one of which (ORF2) encodes both the endonuclease domain (EN) and the reverse transcriptase domain (RT) that are important for the transposition. Alu is about 300 bp without the protein-coding ability, and it essentially relies on the activity of the ORF2 protein encoded by LINE-1 for its transposition. In addition to Alu, SVA (SINE-VNTR (variable number of tandem repeats)–Alu, a family of retroposon class) also requires the LINE-1 for the transposition [9]. REs are generally believed to be beneficial for the species, as they can help maintain the integrity of centromeres and telomeres. They have also been extensively domesticated in the genome to benefit genome evolution [10]. In particular, most TEs in the human genome were inserted millions of years ago and have accumulated mutations and thus become defective [8]. However, it has been estimated that about 80–100 LINE-1 elements are still fully functional and capable of transposition [11]. Therefore, it is deleterious at the individual level if their activities are not adequately regulated [10]. Given the critical role REs play in chromosome integrity and genome stability, the host has evolved several mechanisms to ensure the proper regulation of REs. In humans, epigenetic silencing, including DNA methylation and histone modification, is extensively studied [6]. Biochemically, DNA methylation involves adding the methyl (-CH₃) group covalently to the five positions of cytosine moiety, mainly within CpG dinucleotides, which often exist in clusters called CpG islands. Enzymes, including methyltransferases (*DNMTs*), are responsible for DNA methylation [12]. Methyl-CpG binding domain proteins (MBD) can associate with methylated DNA, which can induce histone protein deacetylation by recruiting histone deacetylases. The interplay between DNA methylation and histone acetylation is critical for regulating chromatin conformation, which is essential for regulating gene expressions, including RE expression [12,13]. Besides DNA methylation and histone modifications, piwi proteins and their associated piwi RNA complexes have also been extensively studied for their role in the silencing of REs in the germline via either transcriptional gene regulation (e.g., DNA methylation) or posttranscriptional gene regulation (e.g., bind to and degrade TE transcripts in the cytosol) [11].

1.2. Dysregulation of REs in Cancer Genomes

Compared with normal cells, cancer cells typically present an aberrant epigenetic landscape where the hypermethylation of promoter regions in the tumor suppressor genes is coupled with extensive hypomethylation in the intergenic regions [14]. CpG islands (regions with a high frequency of CpG sites) of gene promoters are a relatively small

part of the genome compared with the vast majority of CpG dinucleotides found in the regions of REs [12]. For example, by analyzing Illumina 450K methylation array data of 10 different types of common cancer samples (including LIHC, HNSC, BLCA, LUSC, COAD, BRCA, KIRC, PRAD, LUAD, and KIRP) compared with the corresponding matched normal samples in the TCGA (The Cancer Genome Atlas) dataset, a recent study found that out of 10 tumor types, 5 of them (LIHC, HNSC, BLCA, LUSC, COAD) showed more hypo-methylated CpGs than hyper-methylated CpGs among differentially methylated CpGs. Importantly, when restricting the analysis of CpGs within TEs, 9 out of 10 analyzed cancer types (except KIRP) showed a significantly higher proportion of hypomethylated CpGs than hypermethylated CpGs [3]. This result indicates that the hypo-methylated TEs are the potential driving force for the extensively observed genome-wide hypomethylation in the cancer genome [12]. Furthermore, a negative relationship was observed between the intergenic TE expression and TE methylation (DNA methylation probes within ± 500 bp region around most 5' sites of intergenic TE annotated in RepeatMasker [2]) at the subfamily level [3]. In terms of temporal activities of TEs during tumor progression, by modeling the progression of tumorigenesis via a series of cell transformations in fibroblast cells, a recent study showed that TE expressions at the subfamily level are significantly increased as cells progress through transformation (i.e., an increasing number of TE subfamilies are up-regulated from early passage to the immortalized stage, and to early transformation) [15]. Notably, it was found that genome hypomethylation occurs at an early stage of transformation, and similar to findings in human cancer studies, this hypomethylation is more pronounced in TE regions [15]. Furthermore, it is observed that the methylation of these TEs remains dynamic (e.g., a given subfamily of TEs can change their methylation level) during the transformation [15]. Besides TEs, studies [16,17] also revealed the dysregulation of satellites in different types of cancers. For example, a recent study found that in bladder cancers, Sat-a (satellite- α) and NBL-2 (microsatellite) were hypomethylated, while D4Z4 (macrosatellite) was hypermethylated compared with normal control; on the other hand, in leukemia, DNA methylation was increased in NBL-2 and D4Z4 [16]. Finally, our previous study on osteosarcoma also found significantly higher expression levels of different satellites in osteosarcoma tumor samples compared with normal controls [17]. It is therefore clear that REs are dynamically dysregulated both epigenetically and transcriptionally in different cancer types as the tumor progresses.

Despite these advancements, the dysregulation of REs in most cancer studies is restricted to the subfamily level (i.e., aggregated measurement) analysis, which may limit the effectiveness of REs as biomarkers in cancer research. Recently, increasing amounts of attention [15,18,19] have been shifted to the characterization of REs in the cancer genome at the locus-specific level. However, most studies only focused on a few types of REs (e.g., HERV, human endogenous retrovirus, a member of LTR) or a few types of cancers. For instance, a recent study focusing on the locus-specific HERV in head and neck cancer patients showed that different clusters of patients, grouped based on HERV expression in the tumor-adjacent normal tissues, had significantly different survival probabilities [18].

To explore the dysregulation of locus-specific REs in cancer, we identified differentially expressed REs between tumor and the matched normal samples in 12 TCGA cancer types (including BLCA, BRCA, COAD, ESCA, HNSC, KIRC, KIRP, LIHC, LUAD, PRAD, THCA, and UCEC) [20]. We found that genic REs were co-regulated with their corresponding transcripts, defined as having overlaps in chromosomal coordinates. We identified the uniquely and recurrently dysregulated REs as well as the biological functions with their associated genes. With the recurrently dysregulated REs (REs that are dysregulated in any seven cancer types), we also identified their associated cancer genes. We identified six REs consistently dysregulated across all 12 cancer types: one up-regulated and five

down-regulated. Notably, four of the five down-regulated REs were located within the intronic region of the *TMEM252* gene, which itself was down-regulated in ten out of twelve cancer types. Our analysis of differentially expressed and methylated REs between tumors and their matched normal controls revealed a consistent negative correlation between RE methylation and expression at the locus-specific level for some of these 12 cancer types.

2. Materials and Methods

2.1. Determine Differentially Expressed REs at Locus-Specific Levels Across 12 Cancer Types

This study analyzed twelve cancer types from the Cancer Genome Atlas (TCGA) database—BLCA, BRCA, COAD, ESCA, HNSC, KIRC, KIRP, LIHC, LUAD, PRAD, THCA, and UCEC [20]. We selected cancer types and subjects (patients) with at least five patients per type, and paired RNA-sequencing (RNA-seq) and methylation data from both tumors and matched normal samples for each patient. RNA-seq data in the bam format was downloaded using the GDC Transfer Tool Client (<https://gdc.cancer.gov/access-data/gdc-data-transfer-tool>, accessed on 25 November 2023). The number of subjects analyzed in each cancer type is detailed in Table S1.

The workflow for RE expression analysis is shown in Figure S1a. Specifically, the bam files were first converted to paired-end reads in fastq format with samtools (version: 1.18) [21]. The paired-end reads were cleaned with Trim Galore (version: 0.6.4, <https://github.com/FelixKrueger/TrimGalore>, accessed on 8 December 2023) to remove adaptors and low-quality bases at read ends by Cutadapt (version: 4.5) [22]. The quality of clean reads was assessed with Fastqc (version: v0.11.8) [23] before aligning to the reference human genome hg38 with STAR (version: 2.7.11a) [24]. The gene expression at the transcript isoform level and RE expression at the locus-specific level were then determined via TELocal (version 1.1.1, <https://github.com/mhammell-laboratory/TELocal>, accessed on 5 December 2023). Specifically, the TELocal takes alignment results generated from STAR, the gene annotation file in gtf format downloaded from UCSC Table Browser [25], and pre-built locus-specific RE gtf indices downloaded from https://labshare.cshl.edu/shares/mhammelllab/www-data/TELocal/prebuilt_indices/hg38_rmsk_TE.gtf.locInd.gz (accessed on 5 December 2023), as inputs, and outputs the read count table for annotated transcripts and REs in the reference human genome. To determine the differentially expressed REs in each cancer type compared with matched normal samples at the locus-specific level, DESeq2 [26] was used to analyze the TELocal results with the *design* = ~ *type* (tumor/normal) + *patient_ID* to account for any individual specific effects (i.e., potential confounding effects from the different characteristics of the individual patient). The normal samples were used as the reference for differential expression analysis. Specifically, count tables generated by TELocal for each cancer type were combined into a single table, which was then analyzed using the DESeq function from the DESeq2 package (version: 1.40.2). REs with a $|\log_2\text{FoldChange}| \geq 1$ and the adjusted *p*-values ≤ 0.05 (using the Benjamini–Hochberg procedure implemented in the results function from DESeq) were considered as the differentially expressed REs comparing tumors with matched normal samples. In the downstream analysis, we then focused on the locus-specific REs that belong to one of the following seven classes DNA, LINE, LTR, RC, Retroposon, SINE, and Satellite in terms of the RE annotations (https://labshare.cshl.edu/shares/mhammelllab/www-data/TELocal/annotation_tables/hg38_rmsk_TE.gtf.locInd.locations.gz, accessed on 5 December 2023). Notably, REs with highly similar sequences, which can lead to ambiguous mapping with short reads, were excluded from further analysis. We visualized RE expression changes in each cancer type with the R package EnhancedVolcano (<https://github.com/kevinblighe/EnhancedVolcano>, accessed on 11 December 2023).

To understand how RE dysregulation varies across different genome regions, we categorized differentially expressed REs based on their location. First, we downloaded annotations for various genic features (5'UTRs, coding regions, introns, and 3'UTRs) in hg38 bed format from the UCSC Table Browser [25]. These were then merged using BEDTools' mergeBed module [27]. The remaining regions of the hg38 genome were designated as intergenic using BEDTools' complementBed module. Finally, the intersect module assigned each differentially expressed RE to a genic or intergenic category. We categorized differentially expressed REs as either genic (within genes) or intergenic (between genes) to understand how their dysregulation varies across the genome. For genic REs, we investigated the transcriptional regulation of their corresponding genes. We first identified uniquely up- and down-regulated REs specific to each cancer type within genic regions. Next, we explored the biological functions associated with the genes linked to these uniquely dysregulated REs. We used the Python package mygene (<https://github.com/biothings/mygene.py>, accessed on 8 December 2023, version: 3.2.2) to retrieve the genes corresponding to transcripts associated with these REs, followed by functional enrichment analysis using g:Profiler [28]. For intergenic REs, we employed t-distributed Stochastic Neighbor Embedding (t-SNE) from scikit-learn [29] to visualize distinct sample types. As a non-linear dimensionality reduction technique, t-SNE preserves the relationships between samples in high-dimensional space by maintaining these relationships in a lower-dimensional map. This allows us to visualize potential clusters or groupings within the data based on intergenic RE expression patterns.

To identify REs commonly dysregulated across multiple cancer types, we analyzed the union of up- or down-regulated REs in various combinations of cancer types, ranging from two (i.e., the union of up- or down-regulated REs in any two cancer types) to twelve (i.e., the union of up- or down-regulated REs in all twelve cancer types). We focused on REs that were dysregulated in at least seven cancer types (more than half of the total analyzed), referenced them as recurrently up- or down-regulated REs, and visualized them using the Python package tagore (version: 1.1.2) [30]. To assess the biological significance of these recurrently dysregulated REs, we analyzed the genes associated with these genic REs using g:Profiler [28]. We also intersected these genes with a curated list of 2682 well-established cancer-related genes from COSMIC Cancer Gene Census [31], TSGene [32], IntOgen [33], oncogene database [34], and OncoKB Cancer Gene List [35]. The expression changes (log₂Fold) of the relevant transcripts between the tumor and matched normal samples were visualized using the R package pheatmap (version: 1.0.12) [36]. Finally, we employed IGV (Integrative Genomics Viewer) [37] to visually validate the recurrently dysregulated REs across all 12 cancer types.

2.2. Determine Differentially Methylated REs at Locus-Specific Levels Across 12 Cancer Types

The Illumina 450K methylation array offers single-base resolution methylation data for over 450,000 CpG sites across the human genome [38]. While covering 96% of CpG islands and previously identified differentially methylated regions in cancer (https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_humanmethylation450.pdf, accessed on 5 February 2024), it only targets about 1.5% of all CpG sites [39]. Specifically, the 50 bp probes employed by this array can cover 99% of RefSeq genes, including their gene bodies and promoter regions. Similar to RNA-seq data, methylation data (SeSAmE Methylation β Values from Methylation Array Harmonization Workflow) expressed in β values for each subject were downloaded using the GDC Transfer Tool Client (<https://gdc.cancer.gov/access-data/gdc-data-transfer-tool>, accessed on 25 November 2023). The workflow for RE methylation analysis is shown in Figure S1b. Specifically, the hg19 annotation associated with each probe was obtained via the R annotation package IlluminHumanMethylation450kann.ilmn12.hg19 (<http://www>.

bioconductor.org/packages/IlluminaHumanMethylation450kanno.ilmn12.hg19/, accessed on 5 February 2024) [40] and the conversion of the coordinates from hg19 to hg38 was conducted using pyliftover (<https://pypi.org/project/pyliftover/>, version: 0.4, accessed on 5 February 2024), where only the probes that can be converted to hg38 were kept for further analysis. Focusing on the transcription start site (TSS) known to be associated with transcriptional repression in cancer [41], we defined a 1 kb region flanking the 5' end (TSS \pm 500 bp) of each RE in terms of the previous study [3]. We used pybedtools [27,42] to identify probes intersecting these 1 kb regions, ensuring they mapped uniquely to each RE locus. Probes that can be mapped to multiple RE loci were removed to reduce the ambiguity. Average β values were calculated for RE loci mapped by multiple probes to represent their methylation levels.

Similar to RE expression analysis, we focused on REs belonging to the seven RE classes. Differentially methylated REs were identified using the limma R package (version: 3.56.2) [43], with the *design* = \sim *type* (tumor/normal) + *patient_ID* to account for individual effects [3]. β values were converted to M values (M value = $\log_2(\beta \text{ value}/(1 - \beta \text{ value}))$) for the statistical tests given their approximate normal distributional properties. Differentially methylated REs were identified based on the adjusted *p*-value ≤ 0.05 (using the Benjamini–Hochberg procedure implemented in the topTable from limma) and an absolute difference in β values between tumor and matched normal samples ≥ 0.1 , as in the previous study [3]. Due to the limited coverage of the Illumina 450K methylation array on the human genome, methylation data were primarily used for association analysis with RE expression data, as described below.

2.3. Determine the Association Between RE Methylation and Expression Changes at the Locus-Specific Level

While a previous study identified a negative association between the intergenic TE expression and the methylation level at the subfamily level [3], we aimed to examine this relationship at the locus-specific level. In this study, however, only 66,859 of the 4,467,488 locus-specific REs used for expression analysis have corresponding methylation probes (107,634 probes). The limited overlap between differentially expressed and methylated REs necessitates separate analyses. For each cancer type, we compared the expression changes (\log_2 fold changes) between hypo- and hypermethylated REs. Additionally, we compared the average methylation changes (tumor vs. normal) based on M values for up- and down-regulated REs. Finally, the Pearson correlation coefficient between the methylation changes (i.e., tumor–normal) based on M values and expression changes based on \log_2 fold changes of normalized expressions (i.e., $\log_2(\text{tumor}/\text{normal})$) for each of these REs was determined based on all subjects in a given cancer type.

2.4. Statistical Analysis

To assess the statistical significance of differences in expression and methylation changes between tumor and matched normal samples, we employed the non-parametric Wilcoxon rank-sum test implemented in the R package ggpubr (version: 0.6.0) [44]. A *p*-value < 0.05 was considered statistically significant. For the enrichment analysis of gene sets, we used the hypergeometric test implemented in g:Profiler [28]. Only annotated genes were included in the statistical domain scope. Multiple testing correction was performed using the default g:SCS method, and adjusted *p*-values < 0.05 were considered significant. Pearson correlation coefficients were calculated using the pearsonr function implemented in SciPy [45] for the correlation analysis between RE methylation and expression changes.

3. Results

3.1. Differentially Expressed REs Between Tumor and Matched Normal Samples at the Locus-Specific Level

Our analysis focused on locus-specific REs in the human genome, as annotated by TElocal for hg38. There are 4,505,469 locus-specific REs, mostly belonging to SINE, LINE, LTR, and DNA classes (Figure S2a,b). Sequence lengths varied across seven RE classes: RC, SINE, DNA, and Retroposons ranged from 10 to 1000 bps, while LINEs spanned 10 to 10,000 bps, and Satellite and LTR elements ranged from 10 to 100,000 bps (Figure S3a). The number of REs generally corresponded to chromosome size across the 22 autosomes (Figure S3b). About half of the human genome comprises various REs (Figure S3c). Furthermore, most REs associated with protein-coding and non-protein-coding genes (based on RefSeq annotation) reside in intronic or intergenic regions (Figure S3d). Finally, we excluded 37,981 REs with identical genomics sequences to reduce the ambiguity in expression analysis, resulting in 4,467,488 REs for further analysis.

Based on these 4,467,488 locus-specific REs, we analyzed RE dysregulation patterns in 12 cancer types, identifying distinct expression changes for each. Concretely, the differentially expressed REs between tumor and matched normal samples in each cancer type were determined separately. As shown in Figure S4 and Table S2, KIRC exhibited the highest number of up-regulated REs (40,125), followed by BRCA (16,586). Conversely, BRCA had the most down-regulated REs (45,586), followed by THCA (31,099). Interestingly, UCEC and HNSC showed the fewest up- and down-regulated REs, respectively. A detailed breakdown of these differentially expressed REs in either genic or intergenic regions is shown in Figure 1a. Clearly, most of these differentially expressed REs fall into the largest four RE classes (i.e., SINE, LINE, LTR, and DNA) in the genic regions. Interestingly, among all RE classes, HNSC has a relatively higher number of Retroposon being up-regulated in tumor samples compared to the matched normal controls (i.e., 720 genic RE up-regulations + 601 intergenic RE up-regulations vs. 5 genic RE down-regulations + 4 intergenic RE down-regulations; see Table S2 for detailed information in the context of other cancer types).

For both up- and down-regulated genic REs, we observed a strong co-regulation with their corresponding transcripts across all 12 cancer types. As shown in Figure 1b, for up-regulated genic REs, the large proportions of the log₂Fold changes for the associated transcripts show positive values across all 12 cancer types (indicating up-regulation). However, these proportions varied slightly among different cancer types. A similar co-regulation between RE and associated transcripts is also consistently observed for the down-regulated genic REs, with predominantly negative values in log₂ fold change, as shown in Figure 1c.

3.2. Uniquely Differentially Expressed REs in Each Cancer Type at the Locus-Specific Level

To identify REs uniquely differentially expressed in each cancer type, we analyzed the differentially expressed REs identified in each of the 12 cancer types (see the details in Table S3–S14 for uniquely dysregulated REs in BLCA, BRCA, COAD, ESCA, HNSC, KIRC, KIRP, LIHC, LUAD, PRAD, THCA, and UCEC, respectively). This analysis revealed that each cancer type exhibits a distinct pattern of RE dysregulation, with unique sets of REs being up- or down-regulated. Figure 2a,b summarize the number of uniquely up- and down-regulated REs stratified by their genomic context (genic vs. intergenic) in each cancer type. As expected, KIRC, which had the highest overall number of up-regulated REs, also exhibited the most uniquely up-regulated REs, primarily in genic regions. Similarly, BRCA and THCA demonstrated the highest number of uniquely down-regulated REs, predominantly within genic regions.

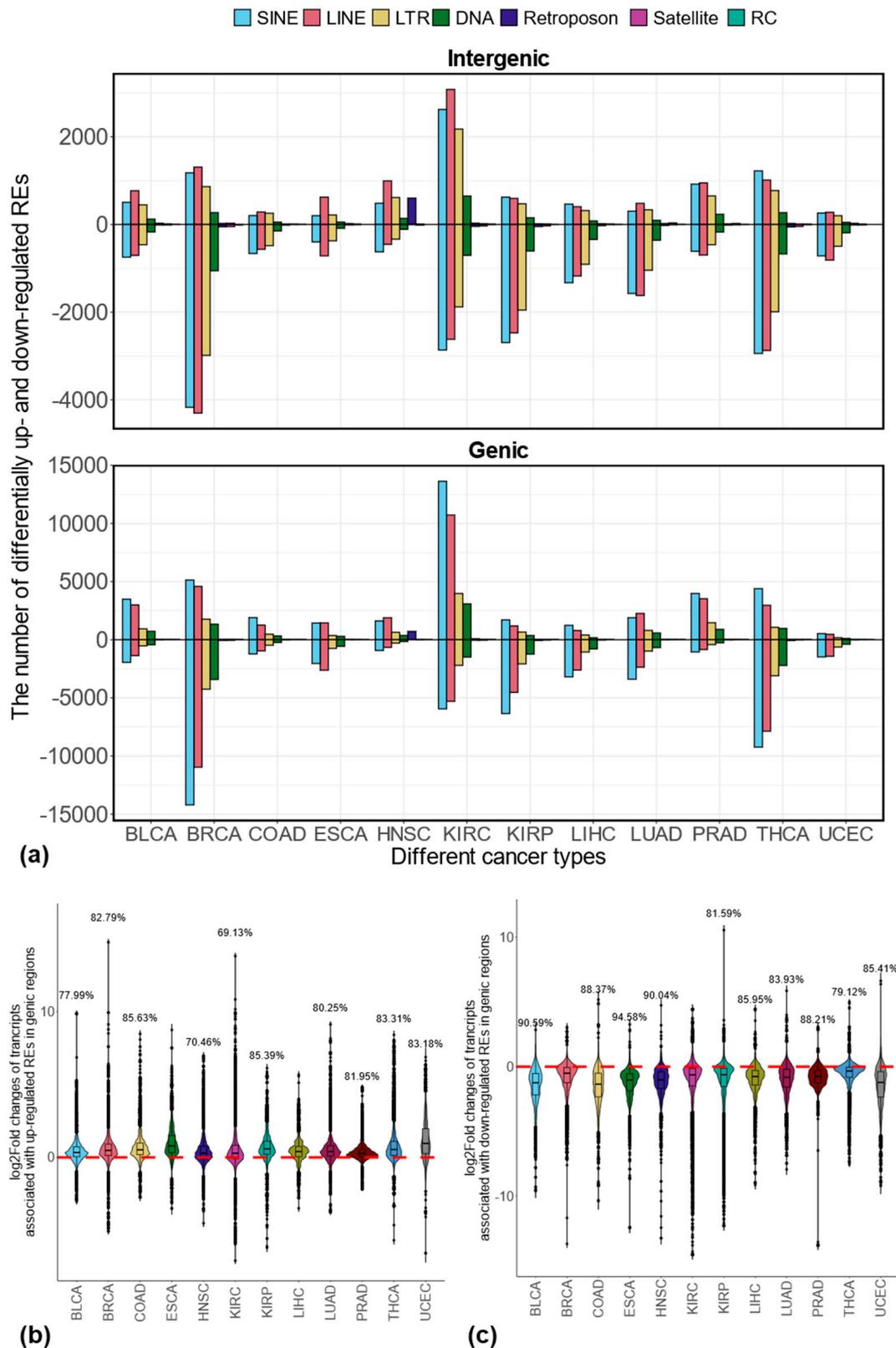


Figure 1. Differentially expressed locus-specific REs across 12 cancer types and co-regulation of transcripts with corresponding genic REs. (a) Number of up- or down-regulated REs stratified by their genomic contexts; (b) expression changes of transcripts associated with corresponding up-regulated genic REs; (c) expression changes of transcripts associated with corresponding down-regulated genic REs.

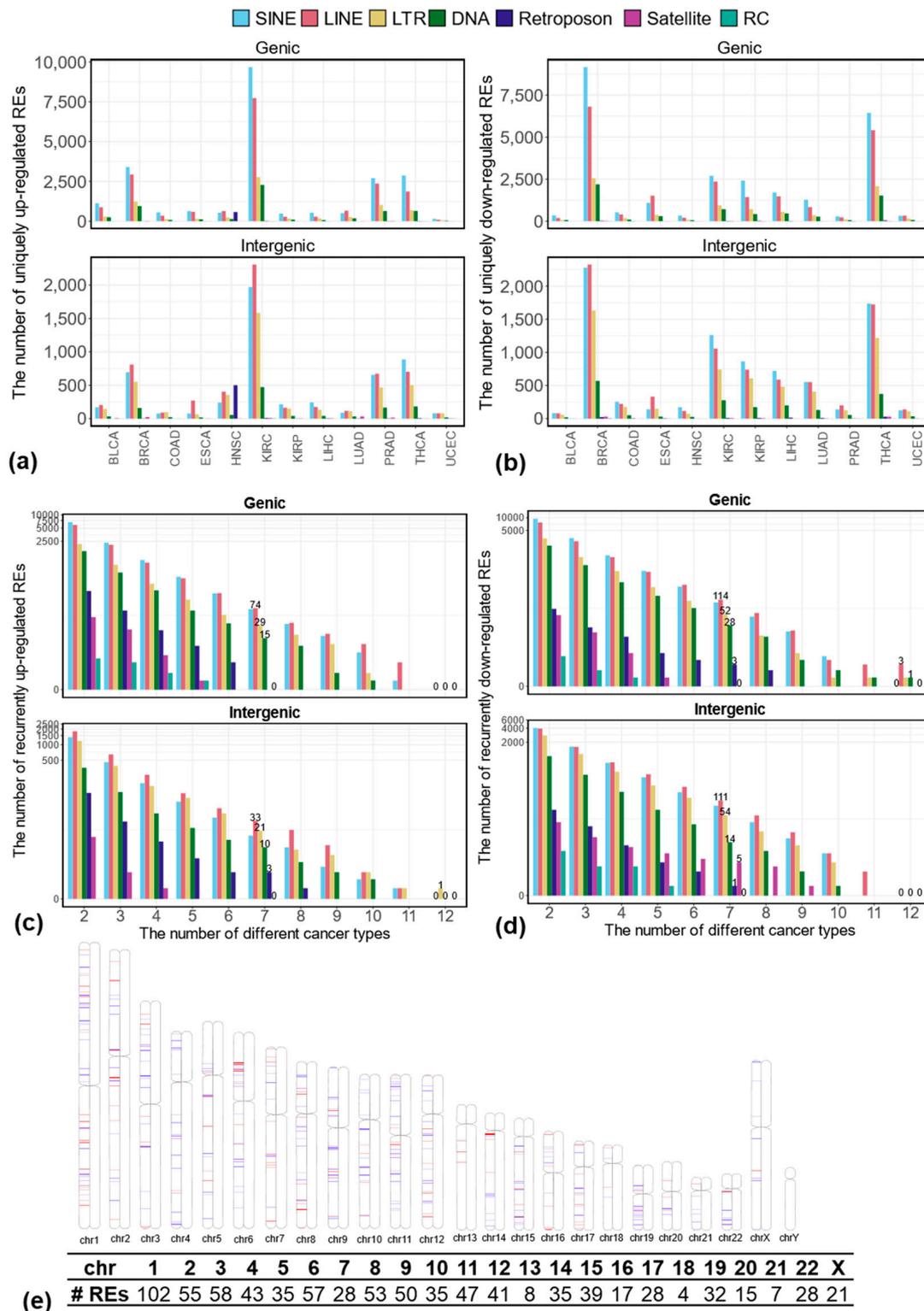


Figure 2. Uniquely and recurrently dysregulated REs across 12 cancer types. (a) Number of uniquely up-regulated REs in a given cancer type stratified by their genomic context; (b) number of uniquely down-regulated REs in a given cancer type stratified by their genomic context; (c) number of recurrently up-regulated REs among any number of 12 cancer types stratified by their genomic context; (d) number of recurrently down-regulated REs among any number of 12 cancer types stratified by their genomic context; (e) genomic locations of recurrently up- (red lines) and down-regulated (blue lines) REs in any seven cancer types (defined as the recurrently dysregulated REs in this study) and their abundance in each human chromosome.

To understand the biological relevance of uniquely differentially expressed REs, we performed gene set enrichment analysis for genes associated with these elements in genic regions. Briefly, we retrieved genes linked to uniquely dysregulated genic REs using the mygene package (<https://github.com/biothings/mygene.py>, version: 3.2.2, accessed on 5 December 2023). The relationship between the number of REs, associated transcripts, and corresponding genes is summarized in Table S15. We then used g:Profiler [28] to analyze the functional terms associated with these genes, including Gene Ontology (GO) terms of GO: MF, GO: BP, and GO: CC and terms related to biological pathways in KEGG, Reactome, and WikiPathways. Figure 3a,b depict the top five enriched functional terms for uniquely up- and down-regulated REs across the 12 cancer types. All terms were statistically significant (adjusted p -values < 0.05), based on the g_SCS correction method. The rich factor is calculated as the intersection size (i.e., the number of genes in the input query annotated to the corresponding term) divided by the term size (i.e., the number of genes annotated to the term in hg38 genome annotation), multiplied by 100 (%).

In terms of uniquely up-regulated genic REs, as shown in Figure 3a, among a total of 38 (i.e., union of the top 5 enriched terms from all 12 cancers) functional terms, 24 are significantly enriched in KIRC, followed by 18 terms in BLCA and 14 terms in BRCA. Notably, UCEC lacked enriched terms. Namely, genes associated with uniquely up-regulated genic REs in UCEC do not contain any enriched terms from GO and Pathway databases. Interestingly, BRCA and ESCA are enriched in the terms related to the cell cycle process, whereas KIRC, BLCA, BRCA, PRAD, HNSC, LUAD, and LIHC show enrichment in intracellular structures, with KIRC and THCA also displaying enriched terms in signal transduction pathways. Among 53 functional terms associated with uniquely down-regulated genic REs, BRCA have the most enriched terms (24), followed by KIRP (21). ESCA, PRAD, and UCEC each have five enriched terms (Figure 3b). BRCA, KIRP, LUAD, BLCA and, to a lesser extent, KIRC, THCA, and ESCA all display enrichment in cell junctions. COAD and LIHC show enrichment in distinct pathways, including cellular glucuronidation, uronic acid metabolic process, pentose and glucuronate interconversion, ascorbate, aldarate, and retinol metabolism.

REs that are uniquely differentially expressed in each cancer may represent the unique characteristics of that cancer. To assess the potential use of uniquely differentially expressed REs as cancer-type representations, we focused on intergenic REs to minimize confounding effects from associated transcripts. We used t-SNE [29] to visualize the normalized expression of these REs, including uniquely up-regulated, down-regulated, and combined sets. Briefly, the normalized expressions of uniquely up- or down-regulated intergenic REs, as well as the combination of them, were used as the feature inputs to t-SNE to cluster different sample types (including different normal samples, different tumor samples, and different normal and tumor samples), as shown in Figure 4. Specifically, totals of 17,541 uniquely up-regulated intergenic REs, 24,351 uniquely down-regulated intergenic REs, as well as 38,899 uniquely dysregulated intergenic REs (i.e., the union of up-regulated and down-regulated REs; there are 2993 REs that overlap between uniquely up-regulated and down-regulated REs because some of the uniquely up-regulated REs in one cancer can be down-regulated in other cancers) were used as the input for t-SNE visualizations (see Figures 4a,b and S5a–f). Clearly, Figure 4 demonstrates the clustering of different sample types (normal, tumor, and combinations) based on these REs. Notably, uniquely up-regulated intergenic REs effectively differentiate sample types with different tissue origins for both normal and tumor samples, as shown in Figure 4a. Interestingly, the tumors and their matched normal samples are close together. Still, they are distinct from other sample types in this two-dimensional latent space when these uniquely up-regulated intergenic REs are used to cluster different normal and tumor samples, as shown in Figure 4b. Similar trends were observed with uniquely down-regulated and combined intergenic REs (Figure S5c–f).

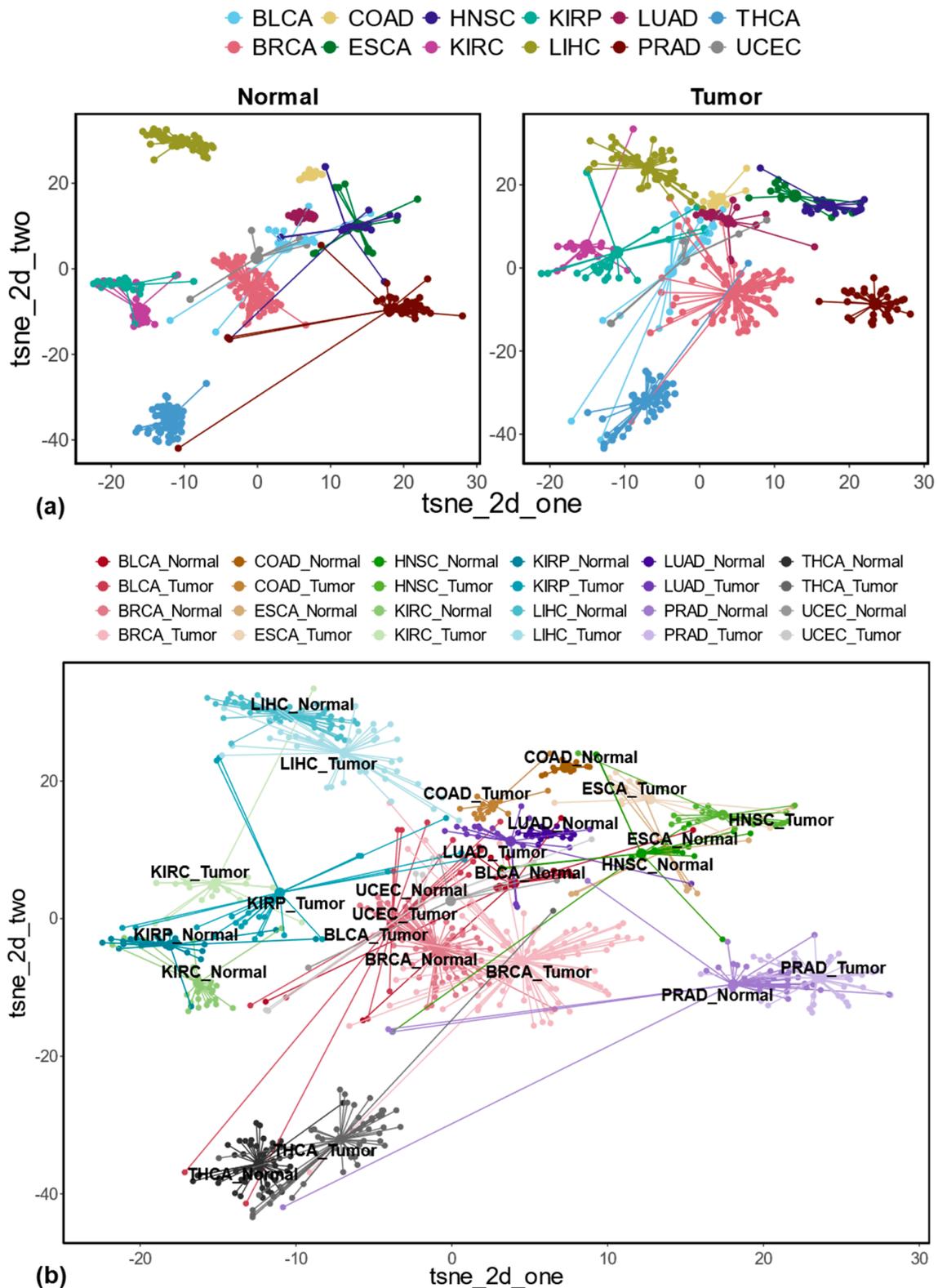


Figure 4. Sample clustering with uniquely up-regulated intergenic REs identified across 12 cancer types (tsne_2d_one, first dimension; tsne_2d_two, second dimension). (a) t-SNE plots based on uniquely up-regulated intergenic REs for normal and tumor sample clustering separately; (b) t-SNE plot based on uniquely up-regulated intergenic REs for different sample type clustering (tumor with matched normal samples).

3.3. Recurrently Differentially Expressed REs in Multiple Cancer Types at the Locus-Specific Level

To identify REs commonly dysregulated across 12 cancer types, we analyzed the overlap of up- and down-regulated REs in various combinations. We focused on those appearing in at least 7 cancer types (more than half of the 12 cancer types), which we termed “recurrently differentially expressed REs” (see the details in Table S16). As shown in Figure 2c, we identified a total of 272 recurrently up-regulated REs, including 188 in genic regions (15 DNA, 74 LINE, 29 LTR, 70 SINE) and 84 in intergenic regions (10 DNA, 33 LINE, 21 LTR, 17 SINE, 3 Retroposon). Similarly, we identified 566 recurrently down-regulated REs (Figure 2d), with a similar distribution between genic and intergenic regions, namely, 295 in genic regions (consisting of 28 DNA, 114 LINE, 52 LTR, 98 SINE, 3 Retroposon REs) and 271 in intergenic regions (composed of 14 DNA, 111 LINE, 54 LTR, 5 Satellite, 86 SINE, 1 Retroposon). Clearly, most of these recurrently dysregulated REs belong to LINE, SINE, LTR, and DNA classes. Figure 2e illustrates the genomic locations of these recurrently up- and down-regulated REs within each chromosome. Notably, chromosome 1, the largest human chromosome, contains the highest number of recurrently dysregulated REs (102 out of 838 or 12.17%).

Moreover, we found that among the transcripts associated with recurrently up-regulated genic REs, a high percentage exhibited increased expression in tumor samples compared to matched normal controls across all 12 cancer types (Figure 5a). Specifically, 188 recurrently up-regulated genic REs are associated with 588 transcripts. Among these transcripts, 90.14% in BLCA show a positive log₂ fold change in expression, while BRCA has the lowest percentage, at 79.08%. The remaining cancer types fall between these two values. Similarly, transcripts associated with recurrently down-regulated REs display decreased expression in tumors. We detected 295 recurrently down-regulated genic REs that correspond to 863 transcripts, most of which are down-regulated, as shown in Figure 5b. Among these transcripts, 90.96% in ESCA exhibit a negative log₂ fold change when comparing tumor and matched normal samples, while PRAD has the lowest percentage at 71.15%. The remaining cancer types fall between these two.

Using the mygene package, we retrieved genes for transcripts associated with recurrently up- and down-regulated genic REs. Among the 588 transcripts associated with recurrently up-regulated genic REs, 450 correspond to 121 genes, while the remaining 138 transcripts lack corresponding reference genes. Similarly, among the 863 transcripts associated with recurrently down-regulated genic REs, 658 correspond to 158 genes, with the remaining 205 transcripts lacking reference genes. These two gene sets were submitted to g:Profiler [28] for functional analysis. Interestingly, the functional analysis revealed distinct patterns for up- vs. down-regulated REs. Genes associated with recurrently up-regulated REs were enriched for functions in the cell cycle and DNA replication (Figure 5c), processes crucial for cancer cell proliferation. Conversely, genes associated with recurrently down-regulated REs displayed enrichment for extracellular matrix (ECM) and ECM proteoglycans (Figure 5d), suggesting potential tumor microenvironment remodeling.

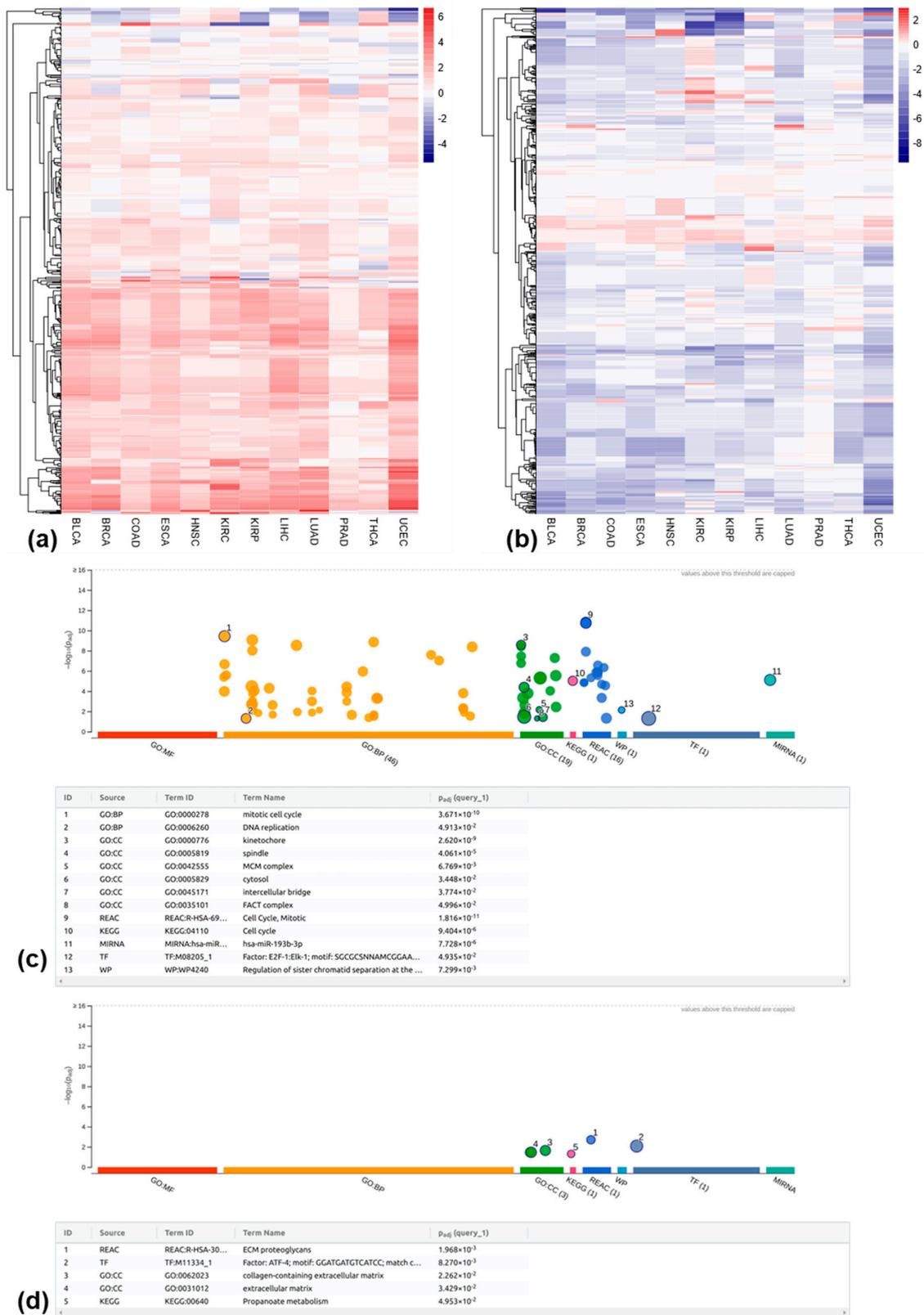
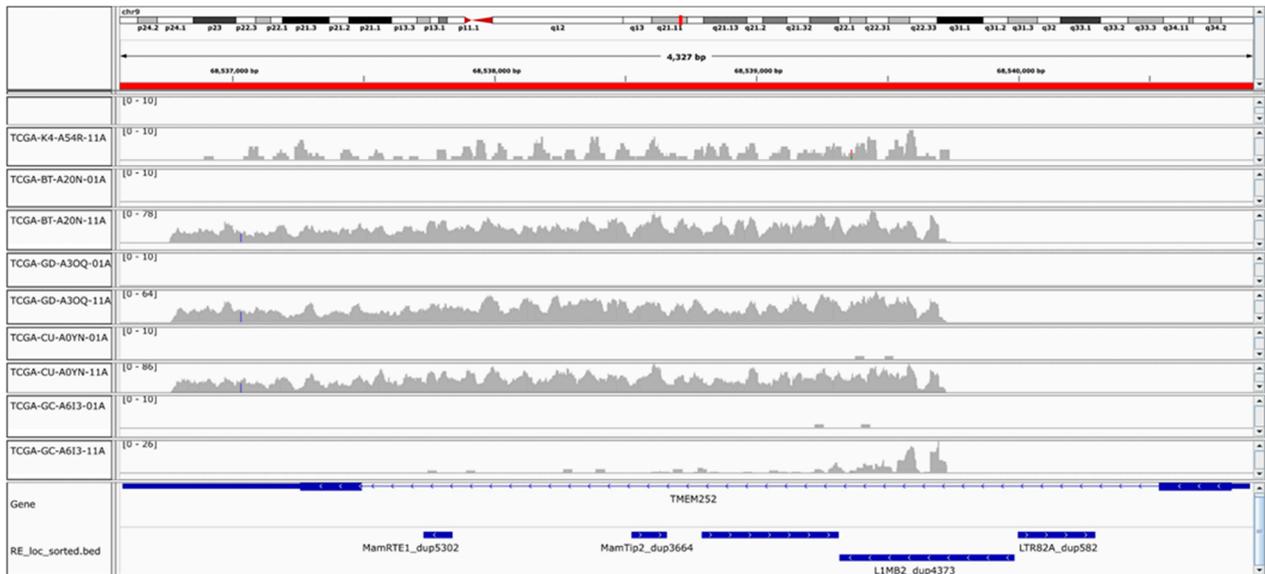


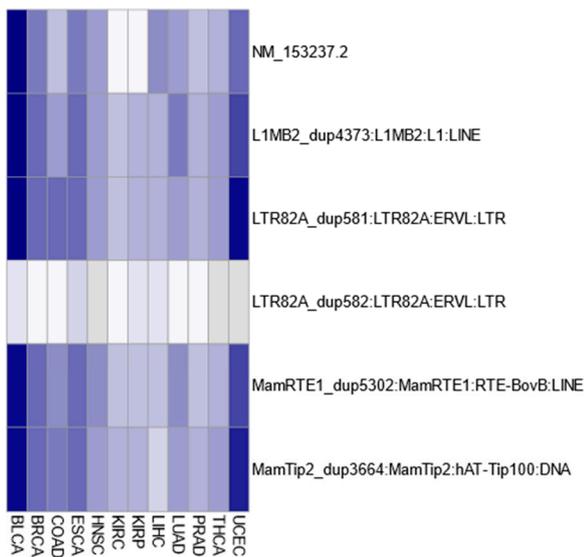
Figure 5. Regulation and functions of transcripts associated with recurrently dysregulated genes. (a) log₂ expression changes between tumor and matched normal samples for transcripts corresponding to recurrently up-regulated genes; (b) log₂ expression changes between tumor and matched normal samples for transcripts corresponding to recurrently down-regulated genes; (c) enriched biological functions associated with the genes in (a); (d) enriched biological functions associated with the genes in (b).

To further investigate the potential importance of these genes in tumor development, we compared these two sets of genes to a curated list of 2682 well-established cancer genes, as mentioned in the method section. We identified 12 cancer genes associated with recurrently up-regulated REs (*CDK1*, *ESCO2*, *GAS5*, *GTSE1*, *H2BC12*, *KIFC1*, *MMS22L*, *NF1*, *NIT2*, *PVT1*, *STAT1*, *UBE2C*) and 19 cancer genes associated with recurrently down-regulated REs (*ADAMTS9-AS2*, *ATF3*, *CASP8*, *DCN*, *DUSP1*, *ECT2L*, *EMP1*, *FANCC*, *JDP2*, *KLF6*, *NDRG2*, *NR4A1*, *RHOBTB2*, *SOCS2*, *SPARCL1*, *STARD13*, *SYNPO2*, *TAGLN*, *TIMP3*). The expression changes of the cancer-related transcripts associating with recurrently up- and down-regulated genic REs between tumor and matched normal controls are shown in Figure S6a,, respectively. While most genes associated with recurrently dysregulated REs are transcriptionally regulated in a similar manner (up-regulated REs associated with up-regulated genes, and vice versa, see Figure 1b,c), some genes, such as *CASP8*, *FANCC*, *ECT2L*, and *SPARCL1*, exhibit opposing expression patterns. For example, in the case of the *CASP8* annotated with nine transcripts, 105 REs were found in its genic regions. Although some of these REs (e.g., *SVA_E_dup106*, *SVA_E_dup107*) are recurrently down-regulated in most of the 12 cancer types, the vast majority of these 105 REs are up-regulated, as shown in Figure S6c. A similar scenario also holds for the *FANCC*, as shown in Figure S6d. On the other hand, for gene *ECT2L* and especially *SPARCL1*, their up-regulation shown in Figure S6b is indeed consistent with the up-regulation of the corresponding genic REs shown in Figure S6e,, respectively. This suggests complex regulatory mechanisms might be involved in RE expression and regulations.

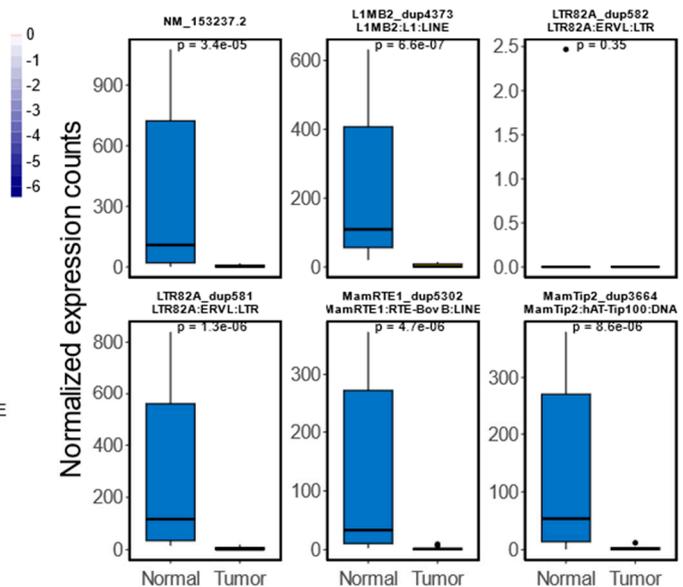
Beyond the recurrently dysregulated REs identified in at least seven cancer types, we also identified a subset of REs consistently dysregulated in all 12 cancer types. As shown in Figure 2c, one intergenic RE (*MLT1D_dup1540*), which belongs to the *MLT1D* subfamily, *ERVL-MaLR* family, and *LTR* class, is consistently up-regulated in all 12 cancers when comparing tumors with matched normal controls. Furthermore, five genic REs (i.e., *MamTip2_dup3664*, *L1MB2_dup4373*, *L1ME3Cz_dup10031*, *MamRTE1_dup5302*, and *LTR82A_dup581*) are consistently down-regulated in all 12 cancer types, as shown in Figure 2d. We validated these findings using IGV, comparing read coverage between tumor and matched normal samples. An example of read coverage corresponding to *MLT1D_dup1540* between 5 randomly selected tumors and matched normal samples from BLCA (out of total 17 paired tumor–normal samples in BLCA) is shown in Figure S7a. The comparison of normalized expression levels across all 12 cancer types is shown in Figure S7b, which is consistent with the IGV results. Interestingly, four of these five consistently down-regulated REs (except for *L1ME3Cz_dup10031*) are located in the same intronic region of the *TMEM252* gene (with transcripts ID: *NM_153237.2*, consisting of two exons). For example, in Figure 6a, we show the reads coverage corresponding to these four REs between the same randomly selected tumor and matched normal samples from BLCA (note that no reads are mapped to the fifth RE, *LTR82A_dup582*). The \log_2 Fold changes of *TMEM252*, as well as the five associated REs (displayed in Figure 6a), across 12 cancer types, are depicted in Figure 6b, and the corresponding normalized expression comparison between tumor and matched normal samples in BLCA is shown in Figure 6c. A similar comparison across all 12 cancer types is shown in Figure S8. Among these five consistently down-regulated genic REs, *L1ME3Cz_dup10031* is located in the intronic region of genes *GCOM1* and *MYZAP*. The IGV genomic context and the normalized expression comparison for *L1ME3Cz_dup10031* between tumor and matched normal samples across all 12 cancer types are shown in Figure S9.



(a)



(b)



(c)

Figure 6. Four consistent down-regulated REs and their corresponding gene *TMEM252* among 12 cancer types. (a) Reads coverage corresponding to these 4 REs (down-regulated in all 12 cancer types) and their associated gene, *TMEM252* (with transcript ID: NM_153237.2), between the 5 randomly selected tumor and matched normal samples in BLCA (same set of samples as used in Figure S7a, tumor sample ends with 01A while the matched normal sample ends with 11A). (b) The log₂Fold changes of *TMEM252* as well as the associated REs between each tumor type and corresponding normal controls, the darker blue represents the increased downregulation of *TMEM252* or REs in tumor samples in terms of their expressions. (c) Normalized expression comparison between tumor and matched normal samples in BLCA for *TMEM252* and associated REs.

3.4. Differentially Methylated REs Between Tumor and Matched Normal Samples at the Locus-Specific Level

To analyze RE methylation, we focused on probes that mapped uniquely to the 1 kb region surrounding the transcription start site (TSS) of locus-specific REs. Among the 4,467,488 locus-specific REs available for expression analysis, 131,322 were covered by at

least one probe within their 1 kb TSS. After excluding probes that mapped to multiple RE loci (due to the overlapping REs in the genome), we identified 66,859 unique REs associated with a total of 107,634 probes (multiple probes can be mapped to the 1 kb TSS of the same RE). These unique REs consist of 33,590 SINE, 17,348 LINE, 7427 LTR, 7988 DNA, 245 Retroposon, 241 Satellite, and 20 RC, as shown in Figure S10a. The breakdown of these REs at the family level is shown in Figure S10b.

To identify differentially methylated REs between tumors and matched normal controls for these 66,859 REs, the calculated M values for each RE were used as the input for the limma R package [43]. Table S17 summarizes the number of hypomethylated REs (i.e., reduced methylation level in tumors) and hypermethylated REs (i.e., increased methylation level in tumors) in each cancer type. For 12 cancer types, 10 (excluding KIRP and PRAD) exhibited significantly more hypomethylated REs than hypermethylated ones. Like differentially expressed REs, most differentially methylated REs belonged to SINE, LINE, LTR, and DNA classes.

3.5. Relationship Between RE Methylation and Expression at the Locus-Specific Level

Given the limited coverage of Illumina 450K methylation array data, we focused on assessing the association between RE methylation and expression changes among the differentially methylated REs (see Table S17). We observed a slight overlap between hypomethylated REs and differentially expressed REs across all 12 cancer types (Figure S11a). Similar patterns were found for hypermethylated REs and differentially expressed REs (Figure S11b). Furthermore, we also identified the differentially methylated REs that are up- or down-regulated in their expression. As shown in Figure S11c,d, similar small overlaps were also observed in this case.

To examine the relationship between methylation and expression changes, we compared expression changes (\log_2 fold changes) of hypo- and hypermethylated REs, assuming hypo-methylated REs are expected to have a higher expression change compared to hypermethylated REs if DNA methylations negatively regulate the expression of REs. As shown in Figure 7a, four cancer types (BRCA, KIRC, KIRP, PRAD) showed this expected pattern, while the remaining eight had non-significant results. Furthermore, we compared the averaged methylation changes (i.e., tumor–normal) based on M values between up- and down-regulated REs with the assumption that averaged methylation changes should be lower for up-regulated REs compared to down-regulated ones. As shown in Figure 7b, six cancer types (BRCA, COAD, KIRC, KIRP, LUAD, and PRAD) showed lower averaged methylation changes when these REs were up-regulated. Therefore, based on this two-complementary analysis, BRCA, KIRC, KIRP, and PRAD demonstrated a consistent negative relationship between RE methylation and expression at the locus-specific level. For recurrently dysregulated REs, we found limited overlap with methylation data due to array coverage. Unfortunately, the six consistently dysregulated REs identified in this study are not covered by the Illumina 450K methylation arrays. Therefore, we focused on those recurrently dysregulated in any seven cancer types. Among 272 recurrently up-regulated REs, only 16 were covered by the DNA methylation array; among 566 recurrently down-regulated REs, only 10 were covered. The Pearson correlation between methylation and expression changes for recurrently up-regulated REs is shown in Figure 7c, while Figure 7d shows the results for recurrently down-regulated REs. While the correlation varied slightly between cancer types, a clear negative correlation was observed for some specific REs (e.g., L1M4_dup16795 in Figure 7d).

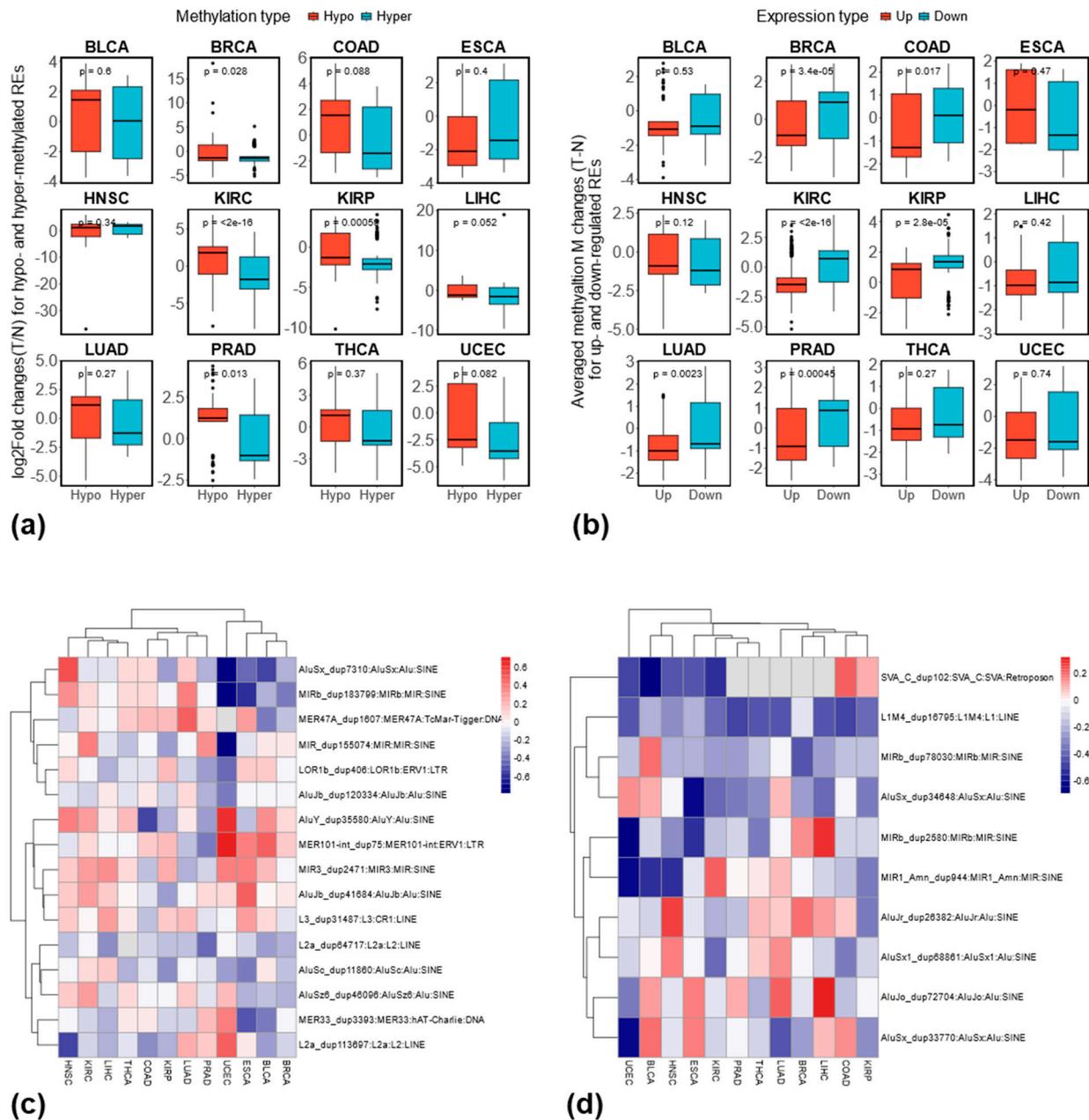


Figure 7. Association between changes of RE methylation and RE expression across 12 cancer types. (a) Comparison of expression changes for differentially expressed REs that are either hypo- or hypermethylated; (b) comparison of methylation changes for differentially methylated REs that are either up- or down-regulated; (c) Pearson correlation coefficient between methylation changes (i.e., tumor–normal) based on M values and expression changes based on log₂ fold changes of normalized expressions (i.e., log₂(tumor/normal)) for recurrently up-regulated REs that are covered by DNA methylation array; (d) Pearson correlation coefficient between methylation changes and expression changes for recurrently down-regulated REs that are covered by the DNA methylation array.

4. Discussion and Conclusions

Previous research has extensively explored the dysregulation of repetitive elements (REs), particularly transposable elements (TEs), in various adult and pediatric cancers [17]. Differentially expressed TEs were investigated comprehensively among 13 cancer types [3]. Although these studies primarily analyzed REs at the subfamily level, few studies investigated the dysregulation of specific REs at the more detailed, locus-specific level. For example, using Telescope [46], a recent study compared the expression levels of locus-

specific HERVs between prostate, breast, and colon cancers and their matched normal controls [19]. They found that 155 HERV loci were differentially expressed in all three cancer types, and 114 were differentially expressed in the same direction. Focusing on head and neck cancer and using Telescope, expression levels of HERVs in the tumor-adjacent normal tissue were shown to help cluster patients with different survival probabilities [18].

To better understand the dysregulation landscape of REs comprehensively at the locus-specific level in cancer genomes, we used TElocal (<https://github.com/mhammell-laboratory/TElocal>, accessed on 5 December 2023). Unlike Telescope, designed primarily for locus-specific HERV analysis [46], TElocal performed better in a recent benchmarking comparison among TE RNA-Seq analysis tools [47]. Our study comprehensively investigated the dysregulation of repetitive elements (REs) at the locus-specific level across 12 cancer types, offering a more detailed picture of RE dysregulation in cancer genomes. We observed uniquely dysregulated REs in each cancer type and commonly dysregulated REs across multiple types. The dysregulation of genic REs, specifically those situated in introns, may influence the expression of corresponding genes. Notably, several genes associated with these dysregulated REs are well-known tumor suppressors or oncogenes, highlighting the potential role of REs in cancer development.

Our analysis identified distinct patterns of RE dysregulation across the 12 cancer types. After accounting for the potential confounding effects of individual subjects in our expression analysis, our results (see Table S2) show that among the 12 cancer types analyzed, 5 (BLCA, COAD, HNSC, KIRC, PRAD) showed more up-regulated REs than down-regulated ones at the locus-specific level. In comparison, despite focusing on the intergenic TE dysregulations at the subfamily levels, the previous study using the TCGA dataset also showed the over-expression of TEs in these five cancer types [3]. On the other hand, the remaining seven cancer types (BRCA, ESCA, KIRP, LIHC, LUAD, THCA, and UCEC) showed more down-regulated REs than up-regulated ones. Among these seven cancer types, four (KIRP, LUAD, BRCA, and THCA) also showed more down-regulated intergenic TE expressions when comparing tumors with matched normal samples from the same previous study [3], where ESCA and UCEC were not analyzed. Interestingly, in the previous study [3], LIHC displayed more up-regulated intergenic TE expressions in tumor samples compared to matched normal samples. This discrepancy is possibly due to the differences in analysis methods (e.g., subfamily level versus locus-specific level), the inclusion of genic REs in our study, and the consideration of potential individual effects in our data analysis. Despite the small differences, our results are, in general, consistent with the previous findings in terms of the amount of dysregulated REs identified from different cancer types (the vast majority of REs are TE; see Figure S2). Compared with the previous study [3], we also included the genic region REs in our analysis, which can help us to better characterize the biological effects of RE dysregulations in cancers by the functional analysis of the associated genes. Most of these genic REs are located in the intronic regions (see Figure S3), and we showed that genic REs are primarily regulated similarly to their corresponding transcripts (see Figure 1b,c and Figure S6c–f). This co-regulation may occur because these dysregulated genic REs could be read-through transcriptions of host genes, as suggested by the previous study [3]. However, it is also possible that these REs are independent entities, regulated in a manner similar to their corresponding transcripts and genes.

To understand the unique characteristics of each cancer type, we identified REs that were exclusively up- or down-regulated in specific cancers. By analyzing the genes associated with these REs, we discovered distinct biological functions enriched in each cancer type. For example, genes linked to up-regulated REs in BRCA and ESCA were enriched in the mitotic cell cycle process, while those related to down-regulated REs in COAD and

LIHC were enriched in cellular glucuronidation. This highlights how the dysregulation of genic REs, likely reflecting the dysregulation of their corresponding transcripts, can contribute to the unique features of different cancer types. In addition to the uniquely dysregulated genic REs, by projecting the uniquely dysregulated intergenic REs to a lower dimensional space with t-SNE, we demonstrated their usefulness in clustering different sample types (cancer types). Notably, these low-dimensional representations captured both the critical difference between different sample types (e.g., clustering different normal sample types as well as different tumor types) as well as the difference for different tissues (i.e., tumor and matched normal samples tend to cluster together, see Figure 4b and Figure S5d,f). The importance of the information contained in dysregulated REs has also been recently shown to lead to improved cancer classifications for liver and esophagus cancer when differentially expressed REs measured at the subfamily level from the blood plasma of different cancer patients are included as features in a logistic regression model [48]. Compared with aggregated measurement at subfamily levels, the measurement of locus-specific REs can provide richer and more informative insights into cancer studies. For example, the uniquely dysregulated intergenic REs identified in this study across different cancer types could help us better characterize each cancer type.

In addition to the uniqueness of each cancer type, we also explored their commonality by determining the commonly dysregulated REs at locus-specific levels in multiple cancer types. With the dysregulation in any of the seven cancer types analyzed, we defined the recurrently differentially expressed REs (see Figure 2c,d). Interestingly, genes corresponding to the recurrently up-regulated genic REs are enriched in DNA replication and mitotic cell cycle. In contrast, genes corresponding to the recurrently down-regulated genic REs are enriched in the extracellular matrix. Notably, among these genes, 12 cancer-related genes, including *CDK1* (oncogene [49]), *ESCO2* (oncogene [50]), *GAS5* (tumor suppressor non-coding gene [51]), *GTSE1* (oncogene [52]), *H2BC12* (potential oncogene [53]), *KIFC1* (oncogene [54]), *MMS22L* (oncogene [55]), *NF1* (tumor suppressor gene [56]), *NIT2* (tumor suppressor gene [57]), *PVT1* (non-long coding RNA with oncogenic effects [58]), *STAT1* (tumor suppressor gene [59]) and *UBE2C* (oncogene [60]), are consistently up-regulated (see Figure S6a). At the same time, 17 out of 19 cancer-related genes, including *ADAMTS9-AS2* (tumor suppressor long non-coding gene [61]), *ATF3* (tumor suppressor gene [62]), *DCN* (tumor suppressor gene [63]), *DUSP1* (promote carcinogenesis in some cancers and inhibits carcinogenesis in other cancers [64]), *ECT2L* (oncogene [65]), *EMP1* (oncogene [66]), *JDP2* (tumor suppressor gene [67]), *KLF6* (tumor suppressor gene [68]), *NDRG2* (tumor suppressor gene [69]), *NR4A1* (tumor suppressor in some cancers oncogene in other cancers [70]), *RHOBTB2* (tumor suppressor gene [71]), *SOCS2* (tumor suppressor gene [72]), *SPARCL1* (tumor suppressor gene [73]), *STARD13* (tumor suppressor gene [74]), *SYNPO2* (tumor suppressor gene [75]), *TAGLN* (oncogene [76]), and *TIMP3* (tumor suppressor gene [77]), are mostly down-regulated in most of the 12 cancer types (see Figure S6b). Furthermore, since one gene typically consists of many transcripts and each transcript can be associated with many genic REs, we found that the regulation direction (i.e., up or down-regulation) for a given gene is similar to the regulation direction of the majority of the corresponding genic REs, as shown in Figure S6c–f.

Among the five consistently down-regulated genic REs in all 12 cancer types, four are in the same intronic region of gene *TMEM252* (with transcripts ID NM_153237.2), as shown in Figure 6a. Interestingly, one of these five RE elements (i.e., LTR82A_dup582) is also located in the same intronic region, but with little to no read coverage in almost all 12 cancer types (see Figure 6b,c and Figure S8), indicating the less likely event of read-through transcription. *TMEM252* (Human transmembrane protein 252), a member of the transmembrane protein family, showed significantly reduced expression in the majority of

the 12 cancer types (including BLCA, BRCA, COAD, ESCA, HNSC, LIHC, LUAD, PRAD, THCA) with non-significant reductions in UCEC when comparing tumors with matched normal samples, as shown in Figure S8. Interestingly, even with all four corresponding genic REs being significantly down-regulated in KIRC and KIRP, the expression level of *TMEM252* showed comparable expression between the tumor and matched normal samples (see Figure S8). *TMEM252* has recently been identified as a tumor suppressor gene in triple-negative breast cancer, inhibiting its progression by suppressing *STAT3* activation [78]. Furthermore, a recent study on papillary thyroid carcinoma demonstrated that the overexpression of *TMEM252* can suppress cell proliferation by repressing the expressions of p53, p21, and p16 through the inhibition of the Notch pathway, and consequently epithelial-mesenchymal transition. The overexpression of *TMEM252* also inhibited cell migration and invasion [79]. Although not being shown to be down-regulated in KIRC and KIRP in this study, evidence has shown that the higher expression levels of *TMEM252* in KIRC (<https://www.proteinatlas.org/ENSG00000181778-TMEM252/pathology/renal+cancer/KIRC>, accessed on 5 May 2024) and KIRP (<https://www.proteinatlas.org/ENSG00000181778-TMEM252/pathology/renal+cancer/KIRP>, accessed on 5 May 2024) are associated with a higher survival probability for the cancer patients. Given the consistent down-regulation across the vast majority of 12 cancer types analyzed in this study and the recently reported tumor-suppressing effects, *TMEM252* could potentially act as a tumor suppressor gene in a wide range of cancer types, thus warranting further in-depth investigation.

To investigate the epigenetic dysregulations of REs, we also analyzed the DNA methylation changes of REs that can be uniquely identified in their 1kb TSS by the corresponding methylation probes. Compared with RNA-seq data, only about 1.5% of REs (66,859 out of 4,467,488) are available for the analysis of methylation changes due to the low coverage of the Illumina 450K methylation array in the human genome (i.e., only 1.5% CpG coverage in the human genome [39]). In contrast to differentially expressed REs (see Table S2), 10 out of 12 cancer types (BLCA, BRCA, COAD, ESCA, HNSC, LIHC, LUAD, THCA, and UCEC) showed a higher number of hypomethylated REs at locus-specific levels than hypermethylated ones (see Table S17). This is consistent with the TE methylation changes observed at the subfamily level [3], further indicating the validity of the RE methylation analysis conducted in this study at the locus-specific level.

Despite the general belief that DNA methylation negatively regulates RE expression [1, 12], our analysis revealed a discrepancy between the prevalence of hypomethylated REs and up-regulated REs in cancer types, namely, the larger number of cancer types with higher levels of hypomethylated REs and the smaller number of cancer types showing more up-regulated REs. This discrepancy may be partially attributed to the limited coverage of the Illumina 450K methylation array. To overcome the small overlaps between differentially methylated REs and differentially expressed REs shown in Figure S11, and to explore the association between RE methylation and expression changes, we compared the expression changes between hypo- and hypermethylated REs, as well the averaged methylation changes between up- and down-regulated REs shown in Figure 7a,b, respectively. We also checked the correlation between methylation and expression changes for recurrently dysregulated REs covered by the DNA methylation array (see Figure 7c,d). In general, expression changes are higher for hypo-methylated REs than hyper-methylated ones, and averaged methylation changes are lower for up-regulated REs than down-regulated ones. Although the Illumina 450K methylation array's limited coverage restricted our analysis of the association between RE methylation and expression, our results are broadly consistent with the belief that DNA methylations restrict RE activities.

In terms of the regulatory effects of genic REs on their associated genes, studies have shown that genes containing more highly methylated REs tend to display significantly

reduced expression compared to genes that contain less methylated REs or genes without genic REs [80]. Therefore, it is possible that the progressive hypomethylation of REs during cancer development can drive the expression of their associated genes. The recent study also showed that cryptic regulatory elements within REs can be frequently co-opted by cancer cells to drive the expression of oncogenes [81]. In *Arabidopsis thaliana*, it has been found that genic REs/TEs are often less methylated than intergenic TEs, and the maintenance of the heterochromatic state of genic REs is important for proper host gene expression [80]; however, it is unclear whether this is also true in human cancer genomes. Studies in mammals demonstrated that repressive histone H3K9 methylation can be deposited on genic LINE1 within the transcriptionally active chromatin for RE repression [82]. Therefore, besides DNA methylation, histone modifications can also affect RE activities, which could then, in turn, affect the surrounding genes.

To our knowledge, this is the first study that has comprehensively characterized the dysregulation of locus-specific REs among multiple common cancer types. With the increasing adaptation to long-read sequencing [83], we expect a higher interest in studying REs at locus-specific levels in cancer research. Due to the limited coverage of the Illumina 450K methylation array, a comprehensive understanding of methylation changes for REs at locus-specific levels in the common cancer is still lacking. With the increasing application of whole genome bisulfite sequencing in cancer studies [84], we expect a high-resolution map of DNA methylation changes for different cancer types. Future studies utilizing whole-genome bisulfite sequencing will provide a more comprehensive picture of DNA methylation patterns for REs at the locus-specific level. Considering the rich information contained in REs at the locus-specific level and complementary information offered by RE expression and methylation changes, we expect to see the development of RE-based biomarkers for cancer type classification and, eventually, for potential therapeutic targets in the near future.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes16050528/s1>, Figure S1: The workflow to identify dysregulated REs across 12 cancer types; Figure S2: Summary of locus-specific REs based on hg38 annotation for TElocal; Figure S3: Distribution of locus-specific REs in the human genome based on hg38 annotation for TElocal; Figure S4: Expression changes of locus-specific REs in each of the 12 cancer types; Figure S5: Sample clustering with uniquely dysregulated intergenic REs identified across 12 cancer types; Figure S6: Expression changes of transcripts corresponding to cancer genes associated with recurrently dysregulated genic REs; Figure S7: Consistent up-regulated intergenic RE; Figure S8: Expression comparison between tumor and matched normal samples for *TMEM252* gene (with transcripts ID: NM_153237.2) and its associated REs across 12 cancer types; Figure S9: Consistent down-regulated intronic RE (i.e., L1ME3Cz_dup10031) across 12 cancer types; Figure S10: Number of REs that can be uniquely mapped with DNA methylation probes; Figure S11: Number of REs that are differentially expressed and differentially methylated; Table S1: The number subjects from different cancer types used in this study; Table S2: The number of differentially expressed REs in each cancer type; Table S3: The uniquely dysregulated REs in BLCA; Table S4: The uniquely dysregulated REs in BRCA; Table S5: The uniquely dysregulated REs in COAD; Table S6: The uniquely dysregulated REs in ESCA; Table S7: The uniquely dysregulated REs in HNSC; Table S8: The uniquely dysregulated REs in KIRC; Table S9: The uniquely dysregulated REs in KIRP; Table S10: The uniquely dysregulated REs in LIHC; Table S11: The uniquely dysregulated REs in LUAD; Table S12: The uniquely dysregulated REs in PRAD; Table S13: The uniquely dysregulated REs in THCA; Table S14: The uniquely dysregulated REs in UCEC; Table S15: The number of uniquely up- and down-regulated genic REs with their associated transcripts and genes; Table S16: The recurrently dysregulated REs in any 7 cancer types; Table S17: The number of differentially methylated REs in each cancer type.

Author Contributions: Conceptualization, C.W.; formal analysis, C.W.; methodology, C.W.; resources, C.L.; supervision, C.L.; visualization, C.W.; writing—original draft, C.W.; writing—review and editing, C.L. All authors have read and agreed to the published version of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This project was partially funded by the College of Arts and Science, the Office for the Advancement of Research & Scholarship, and the Biology Department at Miami University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study are available within the article and/or its Supplementary Materials.

Conflicts of Interest: The authors declare no conflicts of interests.

References

1. Rodić, N.; Burns, K.H. Long Interspersed Element-1 (LINE-1): Passenger or Driver in Human Neoplasms? *PLoS Genet.* **2013**, *9*, e1003402. [CrossRef]
2. Smit, A.F.A.; Hubley, R.; Green, P. RepeatMasker Open-4.0. 2013. Available online: <https://www.repeatmasker.org/> (accessed on 20 November 2023).
3. Kong, Y.; Rose, C.M.; Cass, A.A.; Williams, A.G.; Darwish, M.; Lianoglou, S.; Haverty, P.M.; Tong, A.-J.; Blanchette, C.; Albert, M.L.; et al. Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. *Nat. Commun.* **2019**, *10*, 5228. [CrossRef]
4. Thakur, J.; Packiaraj, J.; Henikoff, S. Sequence, Chromatin and Evolution of Satellite DNA. *Int. J. Mol. Sci.* **2021**, *22*, 4309. [CrossRef] [PubMed]
5. Liao, X.; Zhu, W.; Zhou, J.; Li, H.; Xu, X.; Zhang, B.; Gao, X. Repetitive DNA sequence detection and its role in the human genome. *Commun. Biol.* **2023**, *6*, 954. [CrossRef]
6. Lerat, E.; Casacuberta, J.; Chaparro, C.; Vieira, C. On the Importance to Acknowledge Transposable Elements in Epigenomic Analyses. *Genes* **2019**, *10*, 258. [CrossRef] [PubMed]
7. Barro-Trastoy, D.; Köhler, C. Helitrons: Genomic parasites that generate developmental novelties. *Trends Genet.* **2024**, *40*, 437–448. [CrossRef]
8. Burns, K.H. Transposable elements in cancer. *Nat. Rev. Cancer* **2017**, *17*, 415–424. [CrossRef] [PubMed]
9. Cajuso, T.; Sulo, P.; Tanskanen, T.; Katainen, R.; Taira, A.; Hänninen, U.A.; Kondelin, J.; Forsström, L.; Välimäki, N.; Aavikko, M.; et al. Retrotransposon insertions can initiate colorectal cancer and are associated with poor survival. *Nat. Commun.* **2019**, *10*, 4022. [CrossRef]
10. Rebollo, R.; Romanish, M.T.; Mager, D.L. Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes. *Annu. Rev. Genet.* **2012**, *46*, 21–42. [CrossRef]
11. Levine, A.J.; Ting, D.T.; Greenbaum, B.D. P53 and the defenses against genome instability caused by transposons and repetitive elements. *BioEssays* **2016**, *38*, 508–513. [CrossRef]
12. Pappalardo, X.G.; Barra, V. Losing DNA methylation at repetitive elements and breaking bad. *Epigenetics Chromatin* **2021**, *14*, 25. [CrossRef] [PubMed]
13. Kulis, M.; Esteller, M. DNA Methylation and Cancer. In *Advances in Genetics*; Elsevier: Amsterdam, The Netherlands, 2010; Volume 70, pp. 27–56. ISBN 978-0-12-380866-0.
14. Kang, S.; Li, Q.; Chen, Q.; Zhou, Y.; Park, S.; Lee, G.; Grimes, B.; Krysan, K.; Yu, M.; Wang, W.; et al. CancerLocator: Non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. *Genome Biol.* **2017**, *18*, 53. [CrossRef]
15. Kanholm, T.; Rentia, U.; Hadley, M.; Karlow, J.A.; Cox, O.L.; Diab, N.; Bendall, M.L.; Dawson, T.; McDonald, J.I.; Xie, W.; et al. Oncogenic Transformation Drives DNA Methylation Loss and Transcriptional Activation at Transposable Element Loci. *Cancer Res.* **2023**, *83*, 2584–2599. [CrossRef]
16. Choi, S.H.; Worswick, S.; Byun, H.; Shear, T.; Soussa, J.C.; Wolff, E.M.; Douer, D.; Garcia-Manero, G.; Liang, G.; Yang, A.S. Changes in DNA methylation of tandem DNA repeats are different from interspersed repeats in cancer. *Int. J. Cancer* **2009**, *125*, 723–729. [CrossRef] [PubMed]
17. Wang, C.; Liang, C. The insertion and dysregulation of transposable elements in osteosarcoma and their association with patient event-free survival. *Sci. Rep.* **2022**, *12*, 377. [CrossRef]
18. Kolbe, A.R.; Bendall, M.L.; Pearson, A.T.; Paul, D.; Nixon, D.F.; Pérez-Losada, M.; Crandall, K.A. Human Endogenous Retrovirus Expression Is Associated with Head and Neck Cancer and Differential Survival. *Viruses* **2020**, *12*, 956. [CrossRef]

19. Steiner, M.C.; Marston, J.L.; Iñiguez, L.P.; Bendall, M.L.; Chiappinelli, K.B.; Nixon, D.F.; Crandall, K.A. Locus-Specific Characterization of Human Endogenous Retrovirus Expression in Prostate, Breast, and Colon Cancers. *Cancer Res.* **2021**, *81*, 3449–3460. [[CrossRef](#)] [[PubMed](#)]
20. The Cancer Genome Atlas Research Network Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **2008**, *455*, 1061–1068. [[CrossRef](#)]
21. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
22. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *7*, 10–12. [[CrossRef](#)]
23. Andrews, S.; FastQC. A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 25 November 2023).
24. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [[CrossRef](#)] [[PubMed](#)]
25. Karolchik, D. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **2004**, *32*, D493–D496. [[CrossRef](#)]
26. 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](#)]
27. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [[CrossRef](#)]
28. Kolberg, L.; Raudvere, U.; Kuzmin, I.; Adler, P.; Vilo, J.; Peterson, H. g:Profiler—Interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* **2023**, *51*, W207–W212. [[CrossRef](#)]
29. Pedregosa, F. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
30. Rishishwar, L.; Conley, A.B.; Wigington, C.H.; Wang, L.; Valderrama-Aguirre, A.; King Jordan, I. Ancestry, admixture and fitness in Colombian genomes. *Sci. Rep.* **2015**, *5*, 12376. [[CrossRef](#)]
31. Tate, J.G.; Bamford, S.; Jubb, H.C.; Sondka, Z.; Beare, D.M.; Bindal, N.; Boutselakis, H.; Cole, C.G.; Creatore, C.; Dawson, E.; et al. COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **2019**, *47*, D941–D947. [[CrossRef](#)] [[PubMed](#)]
32. Zhao, M.; Kim, P.; Mitra, R.; Zhao, J.; Zhao, Z. TSGene 2.0: An updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Res.* **2016**, *44*, D1023–D1031. [[CrossRef](#)]
33. Gonzalez-Perez, A.; Perez-Llamas, C.; Deu-Pons, J.; Tamborero, D.; Schroeder, M.P.; Jene-Sanz, A.; Santos, A.; Lopez-Bigas, N. IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods* **2013**, *10*, 1081–1082. [[CrossRef](#)]
34. Liu, Y.; Sun, J.; Zhao, M. ONGene: A literature-based database for human oncogenes. *J. Genet. Genomics* **2017**, *44*, 119–121. [[CrossRef](#)]
35. Chakravarty, D.; Gao, J.; Phillips, S.; Kundra, R.; Zhang, H.; Wang, J.; Rudolph, J.E.; Yaeger, R.; Soumerai, T.; Nissan, M.H.; et al. OncoKB: A Precision Oncology Knowledge Base. *JCO Precis. Oncol.* **2017**, *1*, 1–16. [[CrossRef](#)]
36. Kolde, R. Pheatmap: Pretty Heatmaps. 2015. Available online: <https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf> (accessed on 25 November 2023).
37. Negri, G.L.; Grande, B.M.; Delaidelli, A.; El-Naggar, A.; Cochrane, D.; Lau, C.C.; Triche, T.J.; Moore, R.A.; Jones, S.J.; Montpetit, A.; et al. Integrative genomic analysis of matched primary and metastatic pediatric osteosarcoma. *J. Pathol.* **2019**, *249*, 319–331. [[CrossRef](#)] [[PubMed](#)]
38. Vrba, L.; Futscher, B.W. A suite of DNA methylation markers that can detect most common human cancers. *Epigenetics* **2018**, *13*, 61–72. [[CrossRef](#)]
39. Fan, S.; Tang, J.; Li, N.; Zhao, Y.; Ai, R.; Zhang, K.; Wang, M.; Du, W.; Wang, W. Integrative analysis with expanded DNA methylation data reveals common key regulators and pathways in cancers. *Npj Genom. Med.* **2019**, *4*, 2. [[CrossRef](#)] [[PubMed](#)]
40. Hansen, K. IlluminaHumanMethylation450kanno.ilmn12.hg19: Annotation for Illumina’s 450k Methylation Arrays. 2016. Available online: <https://bioconductor.org/packages/release/data/annotation/html/IlluminaHumanMethylation450kanno.ilmn12.hg19.html> (accessed on 5 February 2024).
41. Ando, M.; Saito, Y.; Xu, G.; Bui, N.Q.; Medetgul-Ernar, K.; Pu, M.; Fisch, K.; Ren, S.; Sakai, A.; Fukusumi, T.; et al. Chromatin dysregulation and DNA methylation at transcription start sites associated with transcriptional repression in cancers. *Nat. Commun.* **2019**, *10*, 2188. [[CrossRef](#)]
42. Dale, R.K.; Pedersen, B.S.; Quinlan, A.R. Pybedtools: A flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* **2011**, *27*, 3423–3424. [[CrossRef](#)] [[PubMed](#)]
43. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [[CrossRef](#)]
44. Alboukadel Kassambara. Ggpubr: “ggplot2” Based Publication Ready Plots. 2020. Available online: <https://rpkgs.datanovia.com/ggpubr/> (accessed on 5 February 2024).
45. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [[CrossRef](#)]

46. Bendall, M.L.; De Mulder, M.; Iñiguez, L.P.; Lecanda-Sánchez, A.; Pérez-Losada, M.; Ostrowski, M.A.; Jones, R.B.; Mulder, L.C.F.; Reyes-Terán, G.; Crandall, K.A.; et al. Telescope: Characterization of the retrotranscriptome by accurate estimation of transposable element expression. *PLoS Comput. Biol.* **2019**, *15*, e1006453. [[CrossRef](#)]
47. Savytska, N.; Heutink, P.; Bansal, V. Transcription start site signal profiling improves transposable element RNA expression analysis at locus-level. *Front. Genet.* **2022**, *13*, 1026847. [[CrossRef](#)] [[PubMed](#)]
48. Reggiardo, R.E.; Maroli, S.V.; Peddu, V.; Davidson, A.E.; Hill, A.; LaMontagne, E.; Aaraj, Y.A.; Jain, M.; Chan, S.Y.; Kim, D.H. Profiling of repetitive RNA sequences in the blood plasma of patients with cancer. *Nat. Biomed. Eng.* **2023**, *7*, 1627–1635. [[CrossRef](#)]
49. Talatam, A.; Reddy, P.K.; Motohashi, N.; Vanam, A.; Gollapudi, R. Targeting Overexpressed Cyclin Dependent Kinase 1 (CDK1) in Human Cancers: Kamalachalcone A Emerged as Potential Inhibitor of CDK1 Kinase Through in Silico Docking Study. *Oncogen* **2023**, *6*, 25. [[CrossRef](#)]
50. Huang, Y.; Chen, D.; Bai, Y.; Zhang, Y.; Zheng, Z.; Fu, Q.; Yi, B.; Jiang, Y.; Zhang, Z.; Zhu, J. ESCO2's oncogenic role in human tumors: A pan-cancer analysis and experimental validation. *BMC Cancer* **2024**, *24*, 452. [[CrossRef](#)]
51. Mourtada-Maarabouni, M.; Pickard, M.R.; Hedge, V.L.; Farzaneh, F.; Williams, G.T. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene* **2009**, *28*, 195–208. [[CrossRef](#)]
52. Wu, X.; Wang, H.; Lian, Y.; Chen, L.; Gu, L.; Wang, J.; Huang, Y.; Deng, M.; Gao, Z.; Huang, Y. GTSE1 promotes cell migration and invasion by regulating EMT in hepatocellular carcinoma and is associated with poor prognosis. *Sci. Rep.* **2017**, *7*, 5129. [[CrossRef](#)] [[PubMed](#)]
53. Zhou, J.; Xing, Z.; Xiao, Y.; Li, M.; Li, X.; Wang, D.; Dong, Z. The Value of H2BC12 for Predicting Poor Survival Outcomes in Patients With WHO Grade II and III Gliomas. *Front. Mol. Biosci.* **2022**, *9*, 816939. [[CrossRef](#)]
54. Han, J.; Wang, F.; Lan, Y.; Wang, J.; Nie, C.; Liang, Y.; Song, R.; Zheng, T.; Pan, S.; Pei, T.; et al. KIFC1 regulated by miR-532-3p promotes epithelial-to-mesenchymal transition and metastasis of hepatocellular carcinoma via gankyrin/AKT signaling. *Oncogene* **2019**, *38*, 406–420. [[CrossRef](#)]
55. Nguyen, M.-H.; Ueda, K.; Nakamura, Y.; Daigo, Y. Identification of a novel oncogene, MMS22L, involved in lung and esophageal carcinogenesis. *Int. J. Oncol.* **2012**, *41*, 1285–1296. [[CrossRef](#)]
56. Mo, J.; Moye, S.L.; McKay, R.M.; Le, L.Q. Neurofibromin and suppression of tumorigenesis: Beyond the GAP. *Oncogene* **2022**, *41*, 1235–1251. [[CrossRef](#)]
57. Lin, C.; Chung, M.; Chen, W.; Chien, C. Growth inhibitory effect of the human *NIT2* gene and its allelic imbalance in cancers. *FEBS J.* **2007**, *274*, 2946–2956. [[CrossRef](#)]
58. Onagoruwa, O.T.; Pal, G.; Ochu, C.; Ogunwobi, O.O. Oncogenic Role of PVT1 and Therapeutic Implications. *Front. Oncol.* **2020**, *10*, 17. [[CrossRef](#)]
59. Wang, W.; Lopez McDonald, M.C.; Kim, C.; Ma, M.; Pan, Z. (Tommy); Kaufmann, C.; Frank, D.A. The complementary roles of STAT3 and STAT1 in cancer biology: Insights into tumor pathogenesis and therapeutic strategies. *Front. Immunol.* **2023**, *14*, 1265818. [[CrossRef](#)]
60. Wang, L.; Zhao, S.; Wang, Y.; Liu, J.; Wang, X. UBE2C promotes the proliferation of acute myeloid leukemia cells through PI3K/AKT activation. *BMC Cancer* **2024**, *24*, 497. [[CrossRef](#)] [[PubMed](#)]
61. Yao, J.; Zhou, B.; Zhang, J.; Geng, P.; Liu, K.; Zhu, Y.; Zhu, W. A new tumor suppressor lncRNA ADAMTS9-AS2 is regulated by DNMT1 and inhibits migration of glioma cells. *Tumor Biol.* **2014**, *35*, 7935–7944. [[CrossRef](#)] [[PubMed](#)]
62. Ku, H.-C.; Cheng, C.-F. Master Regulator Activating Transcription Factor 3 (ATF3) in Metabolic Homeostasis and Cancer. *Front. Endocrinol.* **2020**, *11*, 556. [[CrossRef](#)]
63. Zhang, W.; Ge, Y.; Cheng, Q.; Zhang, Q.; Fang, L.; Zheng, J. Decorin is a pivotal effector in the extracellular matrix and tumour microenvironment. *Oncotarget* **2018**, *9*, 5480–5491. [[CrossRef](#)] [[PubMed](#)]
64. Shen, J.; Zhang, Y.; Yu, H.; Shen, B.; Liang, Y.; Jin, R.; Liu, X.; Shi, L.; Cai, X. Role of DUSP1/MKP1 in tumorigenesis, tumor progression and therapy. *Cancer Med.* **2016**, *5*, 2061–2068. [[CrossRef](#)]
65. Hirata, D.; Yamabuki, T.; Miki, D.; Ito, T.; Tsuchiya, E.; Fujita, M.; Hosokawa, M.; Chayama, K.; Nakamura, Y.; Daigo, Y. Involvement of Epithelial Cell Transforming Sequence-2 Oncoantigen in Lung and Esophageal Cancer Progression. *Clin. Cancer Res.* **2009**, *15*, 256–266. [[CrossRef](#)]
66. Ahmat Amin, M.K.B.; Shimizu, A.; Zankov, D.P.; Sato, A.; Kurita, S.; Ito, M.; Maeda, T.; Yoshida, T.; Sakaue, T.; Higashiyama, S.; et al. Epithelial membrane protein 1 promotes tumor metastasis by enhancing cell migration via copine-III and Rac1. *Oncogene* **2018**, *37*, 5416–5434. [[CrossRef](#)]
67. Heinrich, R.; Livne, E.; Ben-Izhak, O.; Aronheim, A. The c-Jun Dimerization Protein 2 Inhibits Cell Transformation and Acts as a Tumor Suppressor Gene. *J. Biol. Chem.* **2004**, *279*, 5708–5715. [[CrossRef](#)] [[PubMed](#)]
68. DiFeo, A.; Martignetti, J.A.; Narla, G. The role of KLF6 and its splice variants in cancer therapy. *Drug Resist. Updat.* **2009**, *12*, 1–7. [[CrossRef](#)]
69. Von Karstedt, S. NDRG2 programs tumor-associated macrophages for tumor support. *Cell Death Dis.* **2018**, *9*, 294. [[CrossRef](#)] [[PubMed](#)]

70. Deng, S.; Chen, B.; Huo, J.; Liu, X. Therapeutic potential of NR4A1 in cancer: Focus on metabolism. *Front. Oncol.* **2022**, *12*, 972984. [[CrossRef](#)]
71. Choi, Y.M.; Kim, K.B.; Lee, J.H.; Chun, Y.K.; An, I.S.; An, S.; Bae, S. DBC2/RhoBTB2 functions as a tumor suppressor protein via Musashi-2 ubiquitination in breast cancer. *Oncogene* **2017**, *36*, 2802–2812. [[CrossRef](#)] [[PubMed](#)]
72. Liu, J.; Liu, Z.; Li, W.; Zhang, S. SOCS2 is a potential prognostic marker that suppresses the viability of hepatocellular carcinoma cells. *Oncol. Lett.* **2021**, *21*, 399. [[CrossRef](#)]
73. Li, T.; Liu, X.; Yang, A.; Fu, W.; Yin, F.; Zeng, X. Associations of tumor suppressor SPARCL1 with cancer progression and prognosis. *Oncol. Lett.* **2017**, *14*, 2603–2610. [[CrossRef](#)]
74. Jaafar, L.; Chamseddine, Z.; El-Sibai, M. StarD13: A potential star target for tumor therapeutics. *Hum. Cell* **2020**, *33*, 437–443. [[CrossRef](#)]
75. Zheng, Z.; Song, Y. Synaptopodin-2: A potential tumor suppressor. *Cancer Cell Int.* **2023**, *23*, 158. [[CrossRef](#)]
76. Zhao, Z.; Lu, L.; Li, W. TAGLN2 promotes the proliferation, invasion, migration and epithelial-mesenchymal transition of colorectal cancer cells by activating STAT3 signaling through ANXA2. *Oncol. Lett.* **2021**, *22*, 737. [[CrossRef](#)]
77. Su, C.-W.; Lin, C.-W.; Yang, W.-E.; Yang, S.-F. TIMP-3 as a therapeutic target for cancer. *Ther. Adv. Med. Oncol.* **2019**, *11*, 1–17. [[CrossRef](#)] [[PubMed](#)]
78. Bi, J.; Wu, Z.; Zhang, X.; Zeng, T.; Dai, W.; Qiu, N.; Xu, M.; Qiao, Y.; Ke, L.; Zhao, J.; et al. TMEM25 inhibits monomeric EGFR-mediated STAT3 activation in basal state to suppress triple-negative breast cancer progression. *Nat. Commun.* **2023**, *14*, 2342. [[CrossRef](#)]
79. Zhang, S.; Xie, R.; Wang, L.; Fu, G.; Zhang, C.; Zhang, Y.; Yu, J. TMEM252 inhibits epithelial–mesenchymal transition and progression in papillary thyroid carcinoma by regulating Notch1 expression. *Head Neck* **2024**, *47*, 324–338. [[CrossRef](#)] [[PubMed](#)]
80. Le, T.N.; Miyazaki, Y.; Takuno, S.; Saze, H. Epigenetic regulation of intragenic transposable elements impacts gene transcription in *Arabidopsis thaliana*. *Nucleic Acids Res.* **2015**, *43*, 3911–3921. [[CrossRef](#)]
81. Jang, H.S.; Shah, N.M.; Du, A.Y.; Dailey, Z.Z.; Pehrsson, E.C.; Godoy, P.M.; Zhang, D.; Li, D.; Xing, X.; Kim, S.; et al. Transposable elements drive widespread expression of oncogenes in human cancers. *Nat. Genet.* **2019**, *51*, 611–617. [[CrossRef](#)]
82. Saze, H. Epigenetic regulation of intragenic transposable elements: A two-edged sword. *J. Biochem.* **2018**, *164*, 323–328. [[CrossRef](#)] [[PubMed](#)]
83. Rech, G.E.; Radío, S.; Guirao-Rico, S.; Aguilera, L.; Horvath, V.; Green, L.; Lindstadt, H.; Jamilloux, V.; Quesneville, H.; González, J. Population-scale long-read sequencing uncovers transposable elements associated with gene expression variation and adaptive signatures in *Drosophila*. *Nat. Commun.* **2022**, *13*, 1948. [[CrossRef](#)]
84. Vidal, E.; Sayols, S.; Moran, S.; Guillaumet-Adkins, A.; Schroeder, M.P.; Royo, R.; Orozco, M.; Gut, M.; Gut, I.; Lopez-Bigas, N.; et al. A DNA methylation map of human cancer at single base-pair resolution. *Oncogene* **2017**, *36*, 5648–5657. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.