

Article

Pan-Cancer Analysis Reveals the Prognostic Potential of the THAP9/THAP9-AS1 Sense–Antisense Gene Pair in Human Cancers

Richa Rashmi  and Sharmistha Majumdar * 

Discipline of Biological Engineering, Indian Institute of Technology Gandhinagar, Gandhinagar 382355, India; richa.rashmi@iitgn.ac.in

* Correspondence: sharmistham@iitgn.ac.in

Abstract: Human THAP9, which encodes a domesticated transposase of unknown function, and lncRNA THAP9-AS1 (THAP9-antisense1) are arranged head-to-head on opposite DNA strands, forming a sense and antisense gene pair. We predict that there is a bidirectional promoter that potentially regulates the expression of THAP9 and THAP9-AS1. Although both THAP9 and THAP9-AS1 are reported to be involved in various cancers, their correlative roles on each other's expression has not been explored. We analyzed the expression levels, prognosis, and predicted biological functions of the two genes across different cancer datasets (TCGA, GTEx). We observed that although the expression levels of the two genes, THAP9 and THAP9-AS1, varied in different tumors, the expression of the gene pair was strongly correlated with patient prognosis; higher expression of the gene pair was usually linked to poor overall and disease-free survival. Thus, THAP9 and THAP9-AS1 may serve as potential clinical biomarkers of tumor prognosis. Further, we performed a gene co-expression analysis (using WGCNA) followed by a differential gene correlation analysis (DGCA) across 22 cancers to identify genes that share the expression pattern of THAP9 and THAP9-AS1. Interestingly, in both normal and cancer samples, THAP9 and THAP9-AS1 often co-express; moreover, their expression is positively correlated in each cancer type, suggesting the coordinated regulation of this H2H gene pair.

Keywords: THAP9; THAP9-AS1; head-to-head genes; pan-cancer; TCGA; survival; co-expression; guilt-by-association



Citation: Rashmi, R.; Majumdar, S. Pan-Cancer Analysis Reveals the Prognostic Potential of the THAP9/THAP9-AS1 Sense–Antisense Gene Pair in Human Cancers. *Non-Coding RNA* **2022**, *8*, 51. <https://doi.org/10.3390/ncrna8040051>

Academic Editors: Maxime Wery and Julien Soudet

Received: 5 April 2022

Accepted: 13 June 2022

Published: 8 July 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

If the 5' ends of two genes are adjacent to one another on opposite DNA strands, and the two genes are transcribed divergently, they are called head-to-head genes. The region between the transcription start sites (TSS) of head-to-head genes can be identified as a putative bidirectional promoter. Eukaryotic genes are sometimes organized in a head-to-head architecture, sharing a bidirectional promoter region for regulating the expression of the two genes [1]. Genome-wide analyses have shown that more than 10% of human genes are arranged in a bidirectional head-to-head architecture with their TSSs located <1 kb apart [2–4].

The transcriptional regulation of genes that share a bidirectional promoter is complex. They can be positively correlated such as human collagen genes COL4A1 and COL4A2 [5,6], negatively correlated such as mouse TO/KF [7], or they can show tissue-specific or condition-specific correlations such as human HSP60 and HSP10, which are coordinated to respond to induction signals [8]. Interestingly, several recent genome-wide studies have reported that the sense gene expression positively correlates with the expression of the corresponding antisense gene in the same tissues or cells. For instance, the expression of [9] 38% of annotated antisense RNA transcripts positively correlated with

sense gene expression in 376 cancer samples comprising nine tissue types. Moreover, several bidirectional gene pairs are associated with human diseases, such as BRCA1/NBR2 [10], ATM/NPAT [11], DHFR/MSH3 [12], and SERPINII/PDCD10 [13].

The THAP9 and THAP9-AS1 genes are a putative bidirectional gene pair; their TSSs are located 166 bases apart, and they are arranged in a head-to-head or divergent manner on opposite DNA strands. THAP9 is a domesticated human DNA transposase, homologous to the widely studied *Drosophila* P-element transposase [14]. The THAP9 protein shares 40% similarity to the P-element transposase and probably does not transpose in vivo due to the absence of terminal inverted repeats and target site duplications. Despite being domesticated, it has retained its catalytic activity [14,15]. Additionally, hTHAP9 belongs to the THAP (Thanatos-associated protein) protein family in humans, containing twelve members (hTHAP0-hTHAP11). All human THAP proteins are characterized by an amino terminal DNA-binding domain called the THAP domain, which is typically 80–90 amino acid residues long and possesses a C2CH-type zinc finger [16,17].

Many THAP family proteins are known to be involved in human diseases. THAP1 has been associated with DYT6 dystonia (a hereditary movement disorder involving sustained involuntary muscle contractions) [18]. Regulation of THAP5 by Omi/HtrA2 has been linked to cell cycle control and apoptosis in cardiomyocytes [19]. THAP1 plays a role in apoptosis by facilitating programmed cell death with the help of the transcription repressor protein Par-4 [20]. The LRRC49/THAP10 bidirectional gene pair is reported to have reduced expression in breast cancer [21]. THAP11 is differentially expressed during human colon cancer progression and acts as a cell growth suppressor by negatively regulating the c-Myc pathway in gastric cancer [22].

THAP9-AS1 is a long non-coding RNA gene located on chromosome 4q21.22. It is up-regulated in nasopharyngeal carcinoma and breast cancer [23,24]. Further, it was also identified as an oncogenic factor promoting cell growth in pancreatic ductal adenocarcinoma and gastric cancer and plays a role in the apoptosis of spontaneous neutrophils [25–27]. THAP9-AS1 knockdown suppresses cell proliferation and enhances apoptosis in esophageal squamous cell carcinoma (ESCC) [28]. A recent study reported that THAP9 and THAP9-AS1 exhibit different gene expression patterns under various stress conditions in the S-phase of the cell cycle. THAP9-AS1 is consistently upregulated under stress, whereas THAP9 exhibits both downregulation and upregulation. Both THAP9 and THAP9-AS1 exhibit periodic gene expression throughout the S-phase, which is characteristic of cell-cycle-regulated genes [29]. Nevertheless, little is known about the biochemical and biological functions of the products encoded by THAP9 and THAP9-AS1 or their combined role in tumorigenesis.

In this study, we were interested in understanding the individual as well as combined roles of THAP9 and THAP9-AS1 in various cancers. We conducted a pan-cancer bioinformatic analysis of THAP9 and THAP9-AS1 focusing on their expression, patient prognosis, and genetic mutations in TCGA [30] and GTEx datasets [31] via TIMER2 [32], GEPIA2 [33] and cBio portal [34]. We used a weighted gene co-expression network analysis (WGCNA) [35] and differential gene correlation analysis (DGCA) [36] to explore the gene expression datasets from TCGA for correlations in expression between THAP9 and THAP9-AS1 and their correlation with other genes. Gene Ontology (GO) and KEGG pathway enrichment analyses were performed to identify the primary biological functions linked to the genes that share the THAP9 and THAP9-AS1 clusters in various tumor and normal samples. Our findings identified genetic mutations and indicated statistical correlations between the expression of THAP9 and THAP9-AS1 and clinical prognosis and several cancer-related pathways, which suggests that the gene pair can serve as a potential prognostic cancer biomarker.

2. Results

2.1. Characterization of THAP9/THAP9-AS1 Promoter

THAP9 is a domesticated transposase [14] and THAP9-AS1 (THAP9-antisense1, ENST00000504520) is a lncRNA [37]. Together they form a sense and antisense gene

pair organized in a “head-to-head” (H2H) orientation (Figure 1a). The sequence between an H2H gene pair, i.e., intra-H2H pair, can act as a bidirectional promoter [3]. It has been reported that bidirectional promoters for H2H gene pairs may regulate the expression of the two genes [1] under specific conditions (e.g., disease).

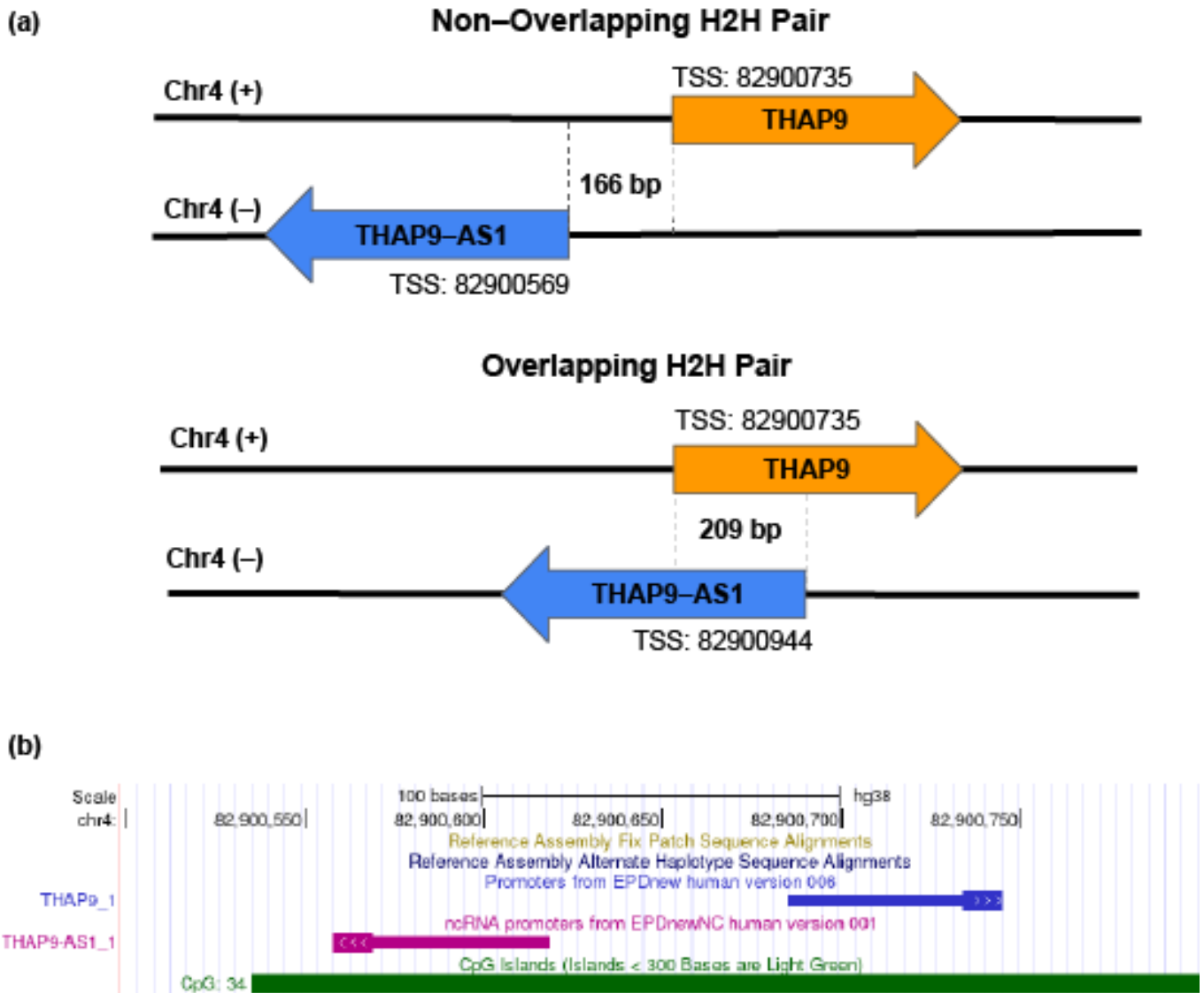


Figure 1. Identification of putative bidirectional THAP9/THAP9-AS1 promoter. (a) Schematic representation of the bidirectional genomic organization of THAP9 and THAP9-AS1 genes along with the TSS predicted by EPDnew. (b) UCSC genome browser showing THAP9 and THAP9-AS1 genes transcribed divergently based on the human GRCh38 assembly. CpG islands overlapping with the bidirectional promoter region are also indicated.

Before investigating the independent and combined role of THAP9/THAP9-AS1 in tumorigenesis, we explored the possibility of a putative bidirectional promoter for the gene pair that may be involved in the regulation of their expression. Thus, we performed an *in silico* analysis of the genomic region spanning their predicted transcriptional start sites (TSS).

According to EPDnew [38], the TSS of THAP9 is located at position 82900735 (Supplementary Figure S1a) on the sense strand of chromosome 4, while two TSSs were predicted for THAP9-AS1 at positions 82900569 and 82900944 (Supplementary Figure S1b), both located on the antisense strand of chromosome 4. Thus, the predicted intergenic

region between THAP9 and THAP9-AS1 is 166 bp (non-overlapping if THAP9-AS1 TSS is at position 82900569) or 209 bp (overlapping if THAP9-AS1 TSS is at position 82900944). In both cases, the THAP9-THAP9-AS1 sense antisense pair follows the head-to-head architecture (Figure 1a). We assumed that the predicted bidirectional promoter region was located in the region spanning -250 to $+250$ relative to the predicted THAP9 TSS and -400 to $+100$ assuming position 82900569 as the TSS for THAP9-AS1 (the selected sequence includes both TSSs for THAP9-AS1), and downloaded the corresponding sequence from EPDnew for further characterization (Supplementary Tables S1 and S2).

Most bidirectional promoters are characterized by the presence of CpG islands, which are regions that are usually devoid of DNA methylation and have a higher G+C content [39]. Thus, we examined the presence of CpG islands (CGIs) in the THAP9/THAP9-AS1 predicted bidirectional promoter region. The analysis and visualization of the selected promoter sequences for each gene, i.e., THAP9 and THAP9-AS1 using EMBOSS CPGplot (https://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/ accessed on 15 September 2021), established that they individually fulfilled the criteria for CGIs [40] and showed a GC content (Percent C + Percent G) > 50.00 and observed/expected CpG ratio > 0.60 (Supplementary Figure S1c,d). We then looked for already annotated CGIs around the THAP9/THAP9-AS1 predicted bidirectional promoter region using the UCSC Genome Browser (<http://genome.ucsc.edu/> accessed on 21 September 2021). We observed a CGI located between positions 82900535 and 82900912 on chromosome 4 overlapping with the putative promoter (Figure 1b). It is tempting to speculate that differential DNA methylation of this CGI may influence the bidirectional gene expression of THAP9/THAP9-AS1.

In addition to CGI, bidirectional promoters are often enriched with specific histone marks. Thus, we decided to look for the already annotated histone mark profile in the THAP9/THAP9-AS1 bidirectional promoter region using ENCODE (Encyclopedia of DNA Elements) [41] (datasets used—EH38E3592191, EH38E3592192), which revealed the presence of bimodal peaks of transcriptionally active histone modifications namely H3K4Me1 (associated with enhancers and the downstream region of TSS), H3K4Me3 (associated with promoters that are active or poised to be activated), and H3K27Ac (associated with enhanced transcription by blocking the spread of the repressive histone mark H3K27Me3). (Figure 2). Moreover, the THAP9/THAP9-AS1 bidirectional promoter was also characterized by DNase I hypersensitivity, which suggests bidirectional transcriptional activity [42]. The 7 cell lines included in this track are GM12878 (lymphoblastoid cell line), H1-hESC (embryonic stem cell line derived from human blastocysts), HSMM (human skeletal muscle myoblasts), HUVEC (human umbilical vein endothelial cells), K562 (human immortalized myelogenous leukemia cell line), NHEK (primary human keratinocytes), and NHLF (human lung fibroblasts).

While there is no consensus on a computational method to predict whether a promoter is bidirectional or not, certain core promoter elements are known to be essential structural features of bidirectional promoters. These elements include the TATA box, CCAAT box, B recognition element (BRE), initiator element (INR), and downstream promoter element (DPE) [44]. The TATA box exists in both unidirectional and bidirectional promoters; however, they are less frequent amongst bidirectional promoters. Bidirectional promoters generally have a higher enrichment of CCAAT boxes and the BRE element compared to unidirectional promoters, while the ratio of DPE to INR remain largely unchanged [44].

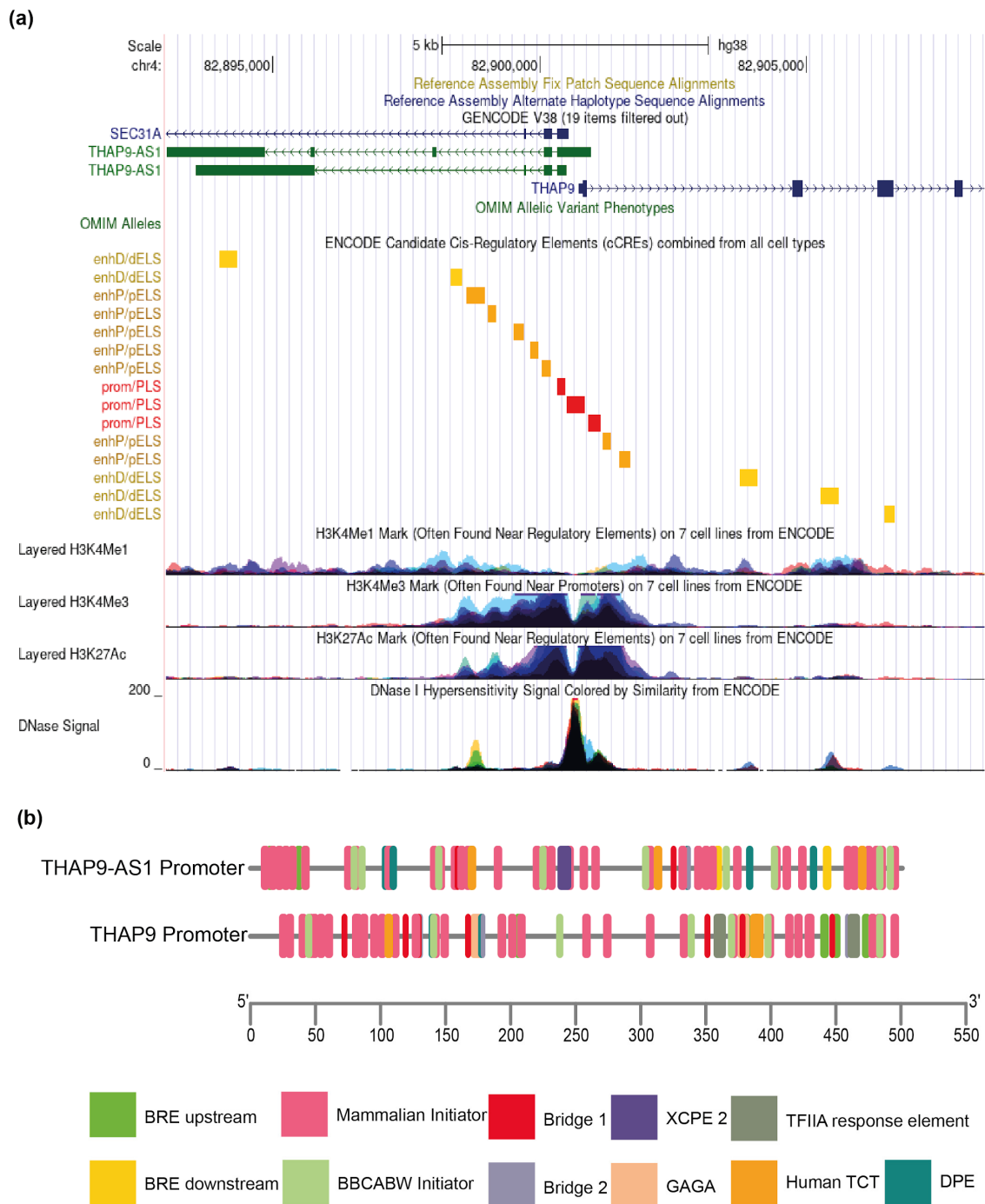


Figure 2. Characterization of THAP9/THAP9-AS1 putative bidirectional promoter region. **(a)** UCSC genome browser representing ENCODE data for THAP9/THAP9-AS1 bidirectional promoter region. The genomic region contains the putative bidirectional promoter region of the THAP9/THAP9-AS1 gene pair. The GENCODE genes track shows transcript variants for both genes. Below that is the ENCODE candidate cis-regulatory elements (cCREs) track, which shows the presence of several regulatory elements in the promoter region. The next three tracks are from ENCODE showing the H3K4Me1, H3K4Me3 and H3K27Ac marks followed by the DNase I hypersensitivity signal shown in the last track **(b)** Schematic representation of the core promoter elements predicted by ElementNT. The core promoter sequence used was -250 to $+250$ relative to the TSS of THAP9 and -400 to $+100$ for THAP9-AS1 (considering 82900569 as TSS) from EPDnew. The diagram is roughly to scale and was constructed using TBtools [43].

Therefore, we submitted the THAP9 and THAP9-AS1 promoter sequences to ElementNT (<https://www.juven-gershonlab.org/resources/element/run/> accessed on 29 September 2021) [45] to predict putative core promoter elements (Figure 2b). The elements identified within the predicted promoter region included BRE, GAGA, mammalian initiator (INR), bridge 1, DPE, BBCABW initiator, human TCT, and TFIIA response element, but not the TATA box (Figure 2b, Supplementary Tables S1 and S2). It is known that in TATA-less promoters, the INR element is the point of transcription initiation [46]. The DPE acts together with INR and is required for binding TFIID [47] and TBP-associated factors (TAFs), particularly TAF6 and TAF9 [48]. Moreover, in the TATA-less promoters, the TFIIB–BRE interaction plays a vital role in assembling the preinitiation complex and transcription initiation [49]. Thus, we note that the selected THAP9/THAP9-AS1 promoter region displays several features (BRE, DPE, INR) characteristic of bidirectional promoters and may be regulated in a coordinated fashion. Thus, the absence of TATA box and presence of other core promoter elements, CGI and histone mark signatures in the region between the TSS of THAP9 and THAP9-AS1, strongly suggest that this region contains a bidirectional promoter.

2.2. Analysis of THAP9/THAP9-AS1 Mutations in Various Tumors

The possible roles of THAP9 in disease has not been investigated. No disease-specific mutation has been reported for the THAP9 gene till date. The only reported missense SNP in THAP9 with a global MAF > 0.05 (minor allele frequency) was rs897945 (MAF = 0.3253), making it a candidate for a derived allele (DA). Derived alleles are new alleles, having a population frequency of at least 5%, which are formed by the mutation of ancestral alleles. It has been reported that disease-associated alleles are more likely to be low-frequency derived alleles [50]. In the *hTHAP9* gene, rs897945 yields a G → T nucleotide substitution, leading to a leucine-to-phenylalanine mutation at position 299 on the Tnp_P_element (Pfam ID: PF12017) domain (Figure 3). This SNP may have a role in atopy and allergic rhinitis in a Singaporean Chinese population [51].

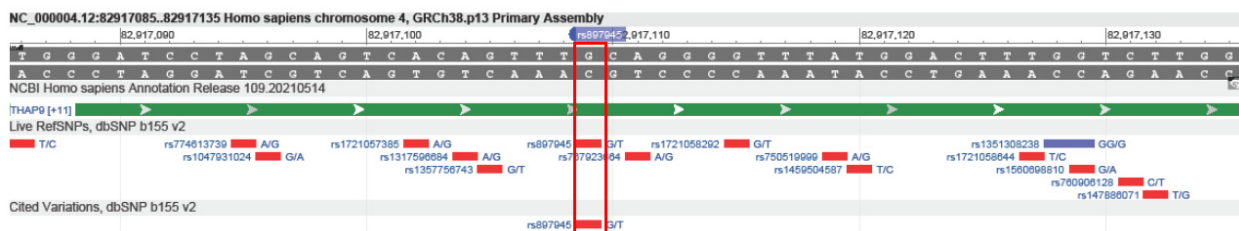


Figure 3. Genome Variation Viewer view of rs897945, which yields a G → T nucleotide substitution that leads to a Leu-to-Phe amino acid change at position 299 located on the Tnp_P_element (Pfam ID: PF12017) domain in hTHAP9 protein.

To understand the disease association of THAP9 and THAP9-AS1 in more detail, we next studied the prevalence of their genetic alterations across various human cancers, using the cBioPortal tool [34] in the “pan-cancer analysis of whole genomes (ICGC/TCGA, Nature 2020)” dataset available at <http://www.cbioportal.org> accessed on 3 October 2021. As shown in Figure 4a,d, pancreatic cancer patients had the highest alteration frequencies in both THAP9 and THAP9-AS1 (>6%), with “amplification” (i.e., more copies, often focal) being the primary alteration type. Notably, both THAP9 and THAP9-AS1 genes underwent amplification in breast cancer, non-small cell lung cancer, melanoma, embryonal tumor, and bone cancer but underwent “deep deletion” (indicates a deep loss, possibly a homozygous deletion) in uterine endometrioid carcinoma patients. Moreover, for the THAP9 gene, “mutation” appeared as the only form of alteration in all patients with colorectal cancer, mature B-cell Lymphoma, and head and neck Cancer. The sites and types of THAP9 mutations are presented in Figure 4c (Supplementary Table S3). The median months survival time was 37.73 for the altered THAP9 group (Figure 4b, 49.61 for reference group). However, it is to be noted that in Figure 4b,e, the *p*-values for both THAP9 and THAP9-AS1

are not less than 0.05, rather the value is 0.2 for THAP9 and 0.5 for THAP9-AS1, which implies that the alterations may not reflect a significant impact on the poor overall survival and prognosis in cancer. This also correlates with the alteration frequency displayed in Figure 4a,d; none of the alterations show more than a 6% frequency.

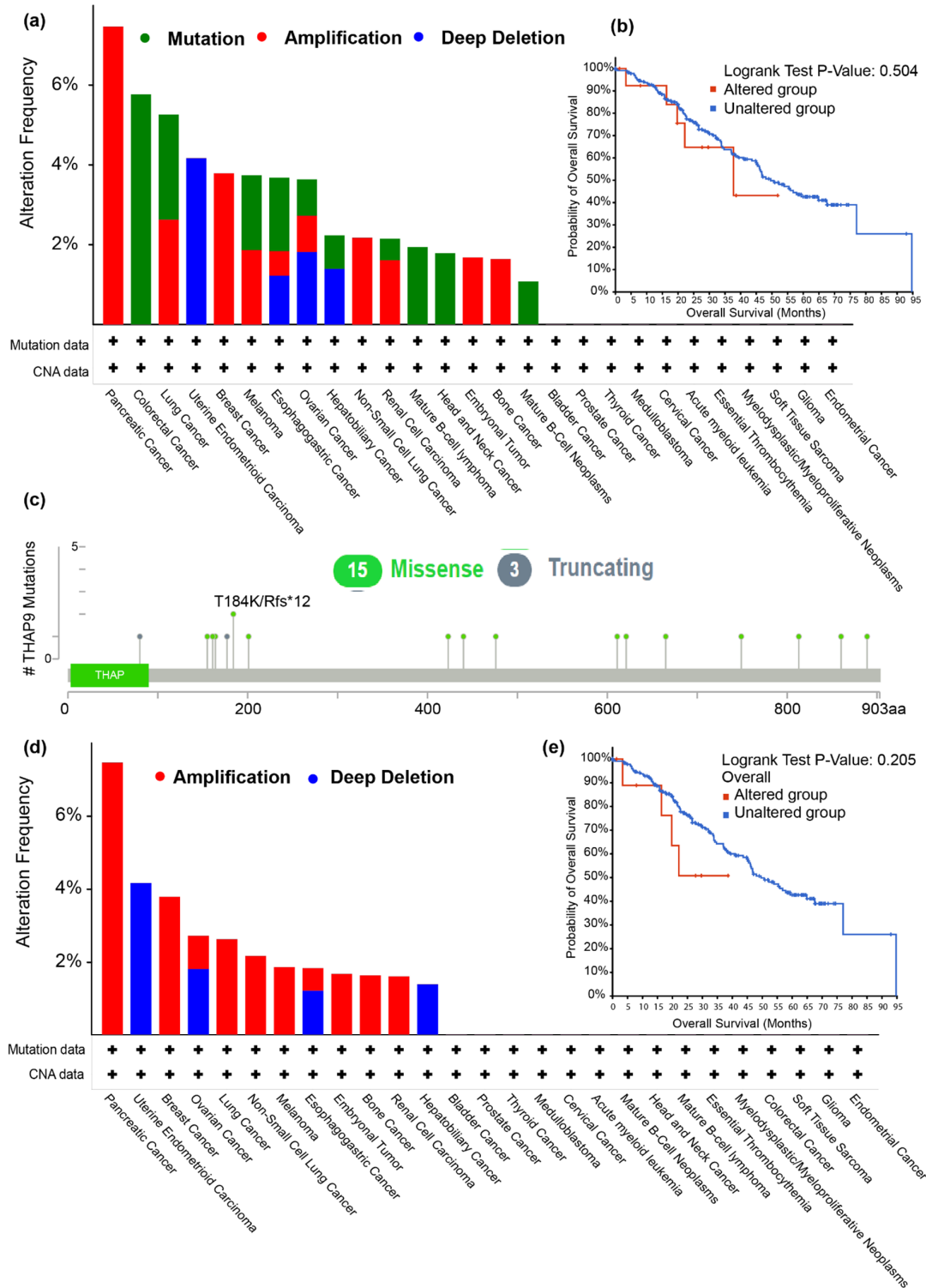


Figure 4. Mutations of THAP9 and THAP9-AS1 in different cancers in TCGA. The alteration frequencies with mutation type for (a) THAP9 and (d) THAP9-AS1, where the X-axis represents the type of alteration (red—amplification; blue—deep deletion; green—mutation) and the Y-axis represents the

frequency of the alteration in different cancers. Correlation between mutation status and overall survival of cancer patients in (b) THAP9 and (e) THAP9-AS1. The red line shows the overall survival estimates for patients with an alteration in the gene as compared to patients with no alteration (blue line). Survival analysis significance was based on the log-rank test. Note: $p < 0.05$ was considered significant. (c) Mutation sites in THAP9 (refer to Supplementary Table S3 for details).

2.3. Difference between THAP9/THAP9-AS1 Expression in Several Cancers

To compare the expression levels of THAP9 and THAP9-AS1 genes between tumor and normal samples, we analyzed their expression levels across various cancer types using TCGA [30] and GTEx [31] datasets via TIMER2.0 [32] and GEPIA2 [33].

TIMER2: We first used TIMER2.0 to evaluate the expression levels of THAP9 between primary tumor and normal samples using TCGA database. We found that THAP9 expression levels in tumor tissues of CHOL ($p < 0.001$), COAD ($p < 0.001$), ESCA ($p < 0.01$), LIHC ($p < 0.001$), LUSC ($p < 0.001$), LUAD ($p < 0.05$), and STAD ($p < 0.001$) were higher than the corresponding normal tissue (Figure 5). On the contrary, THAP9 expression levels in tumor tissues of KIRC ($p < 0.001$), KIRP ($p < 0.001$), PRAD ($p < 0.05$), THCA ($p < 0.001$), and UCEC ($p < 0.001$) were lower than the corresponding normal tissue (Figure 5). We could not perform a similar comparison in ACC, DLBC, LAML, LGG, MESO, OV, SARC, SKCM, TGCT, UCS, or UVM because they lack normal samples in TCGA database. A similar analysis could not be conducted for THAP9-AS1, since TIMER2.0 does not give any gene expression profile for THAP9-AS1.

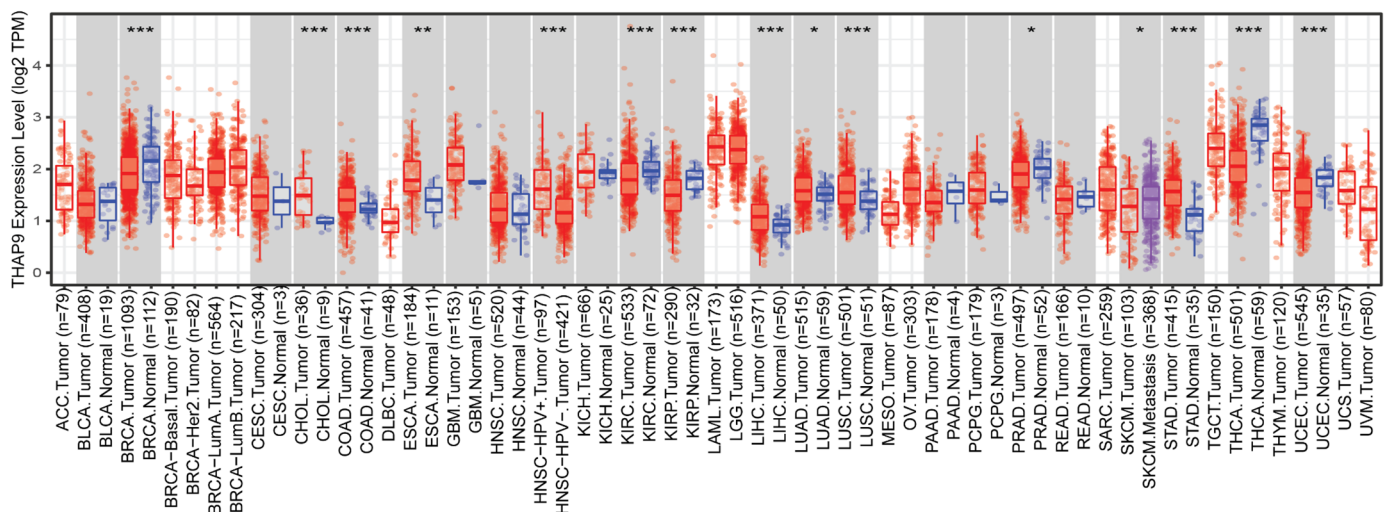


Figure 5. THAP9 gene expression levels in different tumors. THAP9 expression levels in human tumors (red) and corresponding normal tissues (blue) were obtained through TIMER2. The statistical significance computed by the Wilcoxon test is annotated by the number of stars (*: p -value < 0.05 ; **: p -value < 0.01 ; ***: p -value < 0.001).

GEPIA2: Since the TIMER2.0 database did not have the gene expression profile for THAP9-AS1, to get a broader understanding of the gene expression profiles of the two genes we used GEPIA2. It analyzes genes using both TCGA and GTEx datasets. It helped us obtain the differential gene expression profiles of THAP9 and THAP9-AS1 across 31 cancers (pan-cancer gene expression profiles of THAP9 and THAP9-AS1 in Supplementary Figures S2 and S3, respectively). We observed that THAP9 expression was downregulated in TGCT ($p < 0.01$) and was upregulated in THYM ($p < 0.01$), (Figure 6a,b). However, THAP9-AS1 showed downregulation in OV ($p < 0.01$), SKCM ($p < 0.01$), and THCA ($p < 0.01$) (Figure 6c–e) and was upregulated in THYM ($p < 0.01$), PAAD ($p < 0.01$), DLBC ($p < 0.01$), and CHOL ($p < 0.01$) (Figure 6f–i).

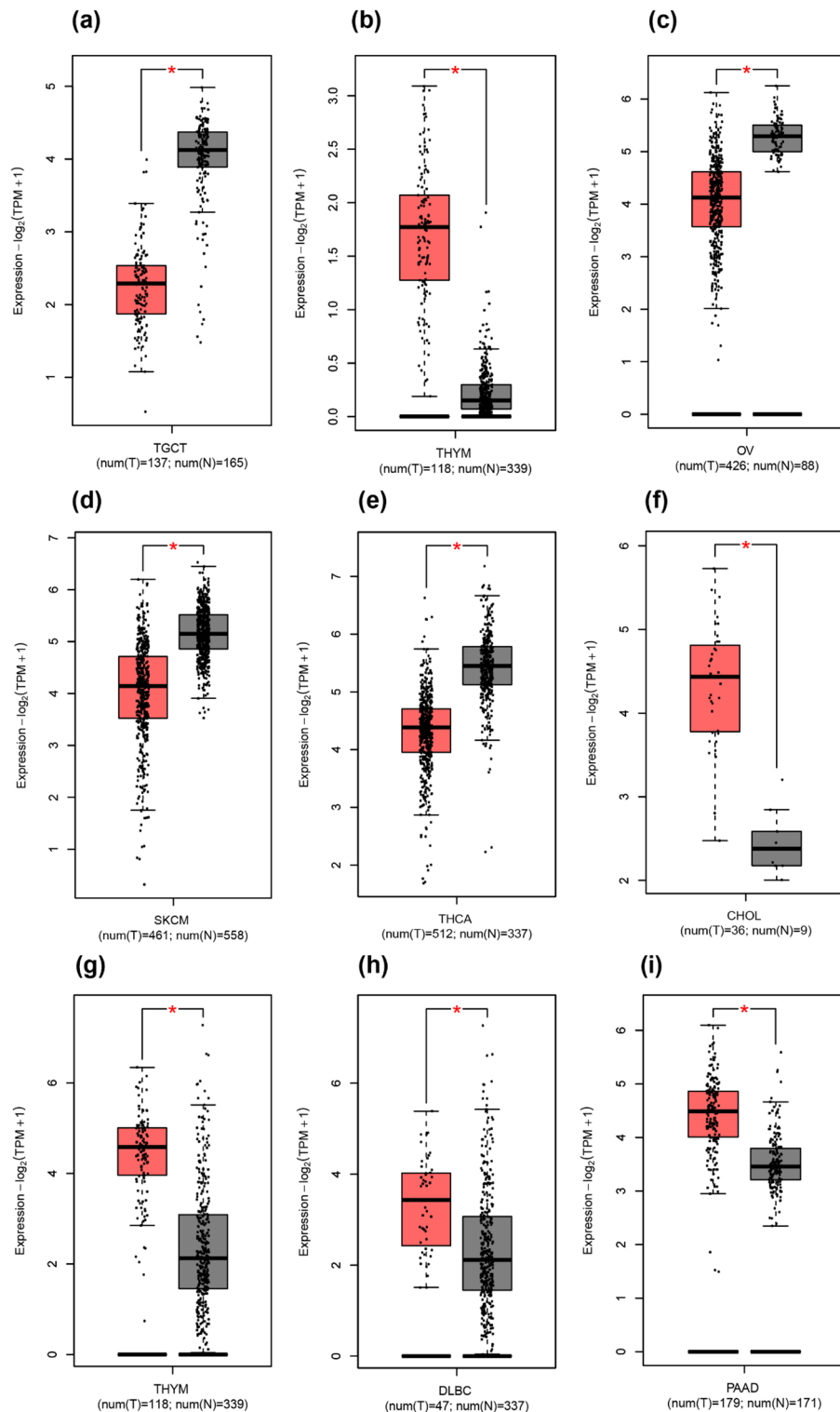


Figure 6. Box plot representation of the comparative expression levels of THAP9 (a,b) and THAP9-AS1 (c–i) in different tumor samples (red) vs. normal tissue samples (grey) from TCGA and GTEx generated using GEPIA2. Note: * $p < 0.01$. GEPIA2 uses one-way ANOVA, taking the pathological stage (X-axis) as the variable for performing differential expression of the input gene. The expression data used for the analysis was $\log_2(\text{TPM} + 1)$ (Y-axis)-transformed. (a) THAP9 is downregulated in TGCT and (b) upregulated in THYM. (c–e) THAP9-AS1 is downregulated in OV, SKCM, and THCA and (f–i) upregulated in CHOL, THYM, DLBC, and PAAD (* $p < 0.05$).

Combining the results from the two methods, we observed that THAP9 and THAP9-AS1 expression levels were coordinately upregulated in CHOL and THYM and were coordinately downregulated in THCA compared with the corresponding normal samples.

Head-to-head genes are often coregulated by bidirectional promoters, although there have been reports of conditional regulation of bidirectional gene pairs as well. For example, some gene pairs such as murine RanBP1/Htf9-c are coregulated only in a common window of the cell cycle [7,52]. On the other hand, human HSP60/HSP10 displays coordinated expression in response to induction signals [8]. We have previously reported that THAP9 and THAP9-AS1 exhibit different gene expression patterns under diverse stress conditions in the S-phase of the cell cycle. THAP9-AS1 is consistently upregulated under stress, whereas THAP9 exhibits both downregulation and upregulation [29].

Thus, given the above, it is possible that THAP9 and THAP9-AS1 show diverse expression patterns in different cancers. The differential expression of THAP9 and THAP9-AS1 in different tumor types suggests that the two genes may have tumor-specific regulatory mechanisms.

2.4. Prognostic Analysis of THAP9 and THAP9-AS1

We used the datasets from TCGA and GTEx via GEPIA2 to investigate the correlation of THAP9 and THAP9-AS1 expression with patients' prognoses across different tumor types. The survival heat map of hazard ratio (HR) values for overall and disease-free survival (Figures 7a and 8a) shows the prognostic impacts of THAP9 and THAP9-AS1 in multiple cancer types. A poor prognosis and poor overall survival were linked to the upregulation of THAP9 expression in LGG and STAD and its downregulation in HNSC and KIRC (Figure 7b). Moreover, poor DFS (disease-free survival) was linked with upregulated THAP9 expression in BLCA and CESC and its downregulated expression in KIRC and THYM (Figure 8b). Similarly, in the case of THAP9-AS1, its upregulation was linked to a poor prognosis and poor overall survival in ACC, LGG, PRAD, SARC, and THCA (Figure 7c) and poor DFS in ACC, KICH, and MESO, while its downregulated expression was linked to poor DFS in KIRC (Figure 8c).

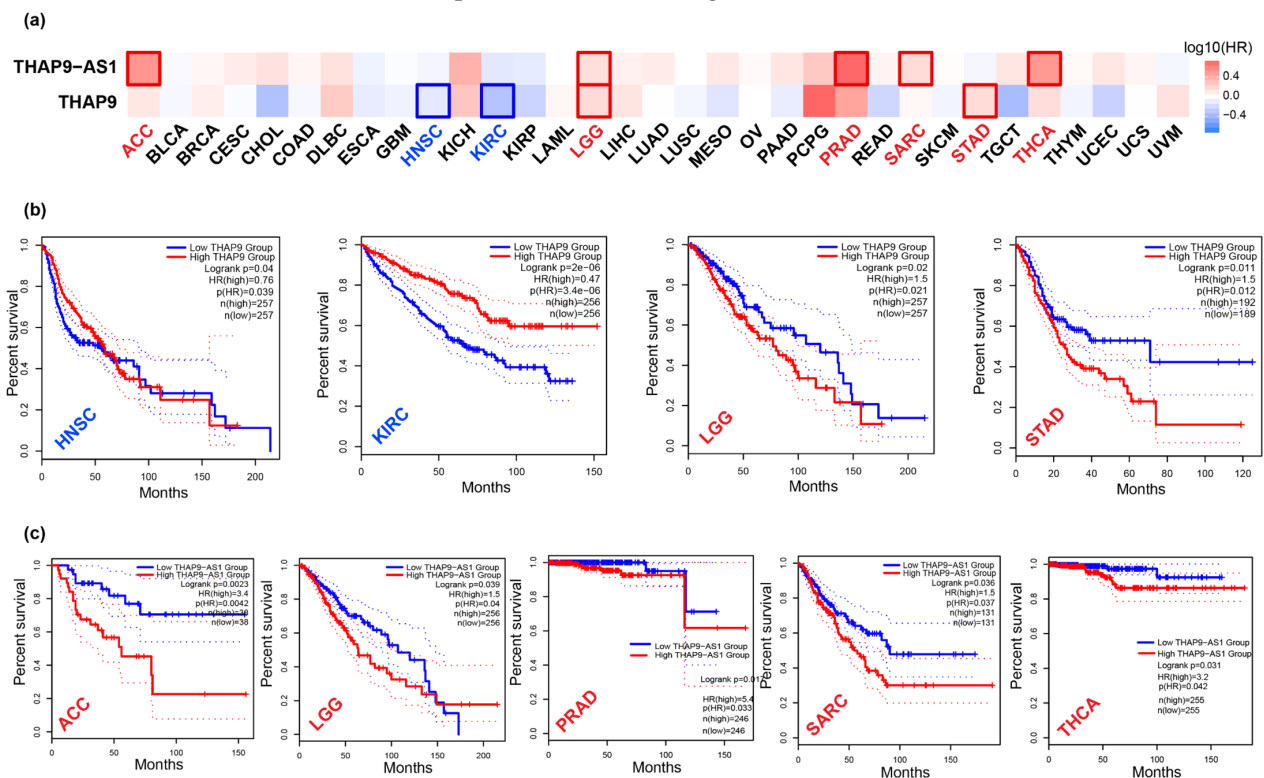


Figure 7. Overall patient survival analysis using GEPIA2. (a) The relationship between THAP9 and THAP9-AS1 gene expression and the overall survival prognosis of cancers in TCGA. Median was

selected as a threshold for separating high-expression and low-expression cohorts. The red and blue blocks represent higher and lower risks, respectively, with an increase in gene expression. The bounding boxes depict the significant ($p < 0.05$) unfavorable and favorable results, respectively. The overall survival and gene expression rates (from TCGA) of (b) THAP9 in HNSC, KIRC, LGG, and STAD; and of (c) THAP9-AS1 in ACC, LGG, PRAD, SARC, and THCA.

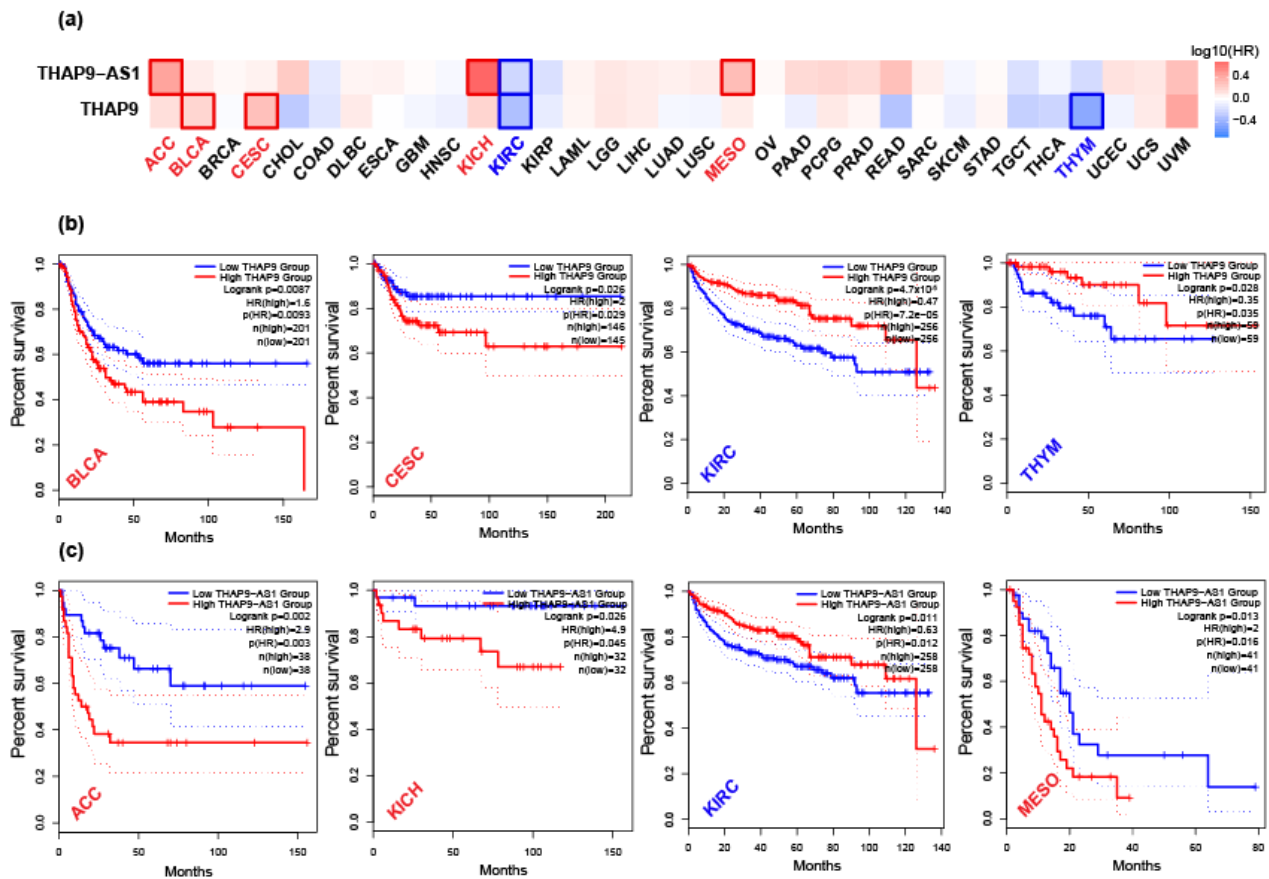


Figure 8. Disease-free survival analysis using GEPIA2. (a) The relationship between THAP9 and THAP9-AS1 gene expression and the disease-free survival prognosis of cancers in TCGA. The median was selected as a threshold for separating high-expression and low-expression cohorts. The red and blue blocks represent higher and lower risks, respectively, with an increase in the gene expression. The bounding boxes depict the significant ($p < 0.05$) unfavorable and favorable results, respectively. The disease-free survival and gene expression rates of (b) THAP9 in BLCA, CESC, KIRC, and THYM; and of (c) THAP9-AS1 in ACC, KICH, KIRC, and MESO.

These findings suggest that *THAP9*- and *THAP9-AS1* related cancer prognoses differ with different cancer types and show much less correlation with the cancer types in which the two genes are differentially expressed. Therefore, the potential of using these genes as a pan-cancer survival indicator is limited.

2.5. Understanding the Role of THAP9 and THAP9-AS1 Using Guilt-By-Association Analysis

A GBA (guilt by association) [53] analysis is often used to predict an unknown gene's function by grouping it with known genes that share its transcriptional behavior. Genes turned on or turned off together under various conditions may be part of the same cellular processes [54]. The exact cellular functions of THAP9 and THAP9-AS1 are unknown. Thus, we decided to compare the expression of THAP9 with THAP9-AS1 and 34125 other genes in 9571 tumor and associated normal samples from 22 human cancers fetched from TCGA [using HTSeq count datasets from TCGA (Supplementary Table S4), excluding cancer types with less than 3 normal samples (rows highlighted in red)]. The gene co-

expression network for THAP9 and THAP9-AS1 was constructed using WGCNA followed by a differential gene correlation analysis using DGCA. We also investigated Gene Ontology and KEGG pathways (Supplementary Tables S5 and S6) for the genes co-expressed with the two genes and the genes differentially correlated across normal vs. tumor samples (Supplementary Tables S7–S9). An analysis of the functions of genes co-expressed with THAP9 and THAP9-AS1 may provide insights into their possible physiological roles.

2.5.1. Gene Co-Expression Analysis

To identify the genes that are co-expressed with THAP9 and THAP9-AS1 in normal vs. tumor samples in each cancer type, we utilized the WGCNA R package [35] to build a weighted co-expression network for the two genes. The samples of each tumor and normal pair were clustered to identify the gene modules representing genes co-expressed with THAP9 and THAP9-AS1. To represent the most frequently co-expressed genes, we selected the top 20 genes co-expressed with THAP9 (Supplementary Table S8) and THAP9-AS1 (Supplementary Table S9), combining all tumors and normal samples across the cancer types (Figure 9).

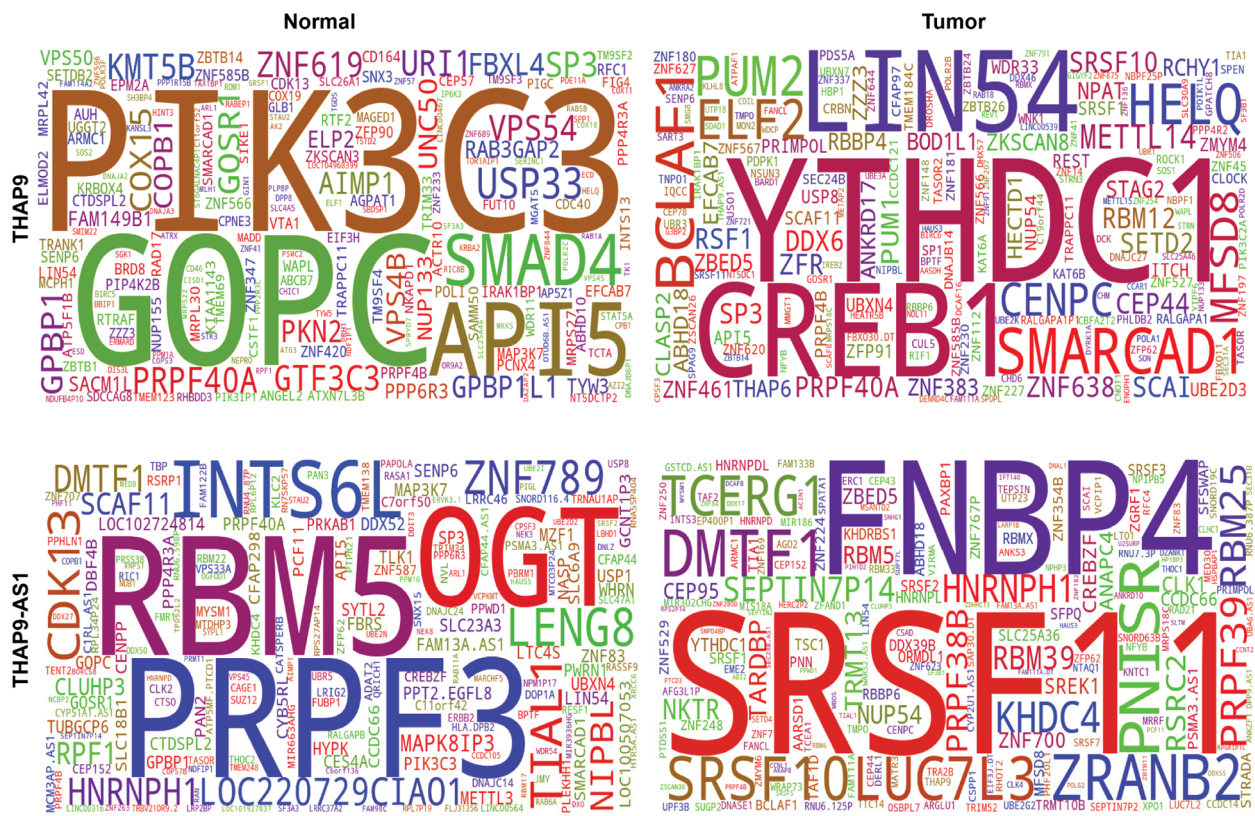


Figure 9. Consensus of genes co-expressed with THAP9 and THAP9-AS1. Word cloud of top 20 genes frequently co-expressed with: (1st row) THAP9 in all combined normal (**top-left**) vs. tumor (**top-right**) samples; (2nd row) THAP9-AS1 in combined normal (**bottom-left**) vs. tumor (**bottom-right**) samples. The co-expressing genes were identified using the WGCNA Bioconductor package and plotted using the Wordcloud python package. The height of a word is directly proportional to the frequency of co-expression with THAP9. The top 20 co-expressed genes are plotted only for representation purposes; details of the full gene cluster associated with THAP9 and THAP9-AS1 in each cancer type (normal and tumor tissues separately) are available in Supplementary Tables S8 and S9.

In normal samples (within BRCA, KIRC, STAD, and THCA datasets), THAP9 and THAP9-AS1 belong to the same gene cluster, suggesting their coregulation (Supplementary Table S7). Some genes that are overrepresented in these clusters are GPBP1, API5, PIK2C3, GOSR1 and PRPF40A. GPBP1, also known as Vasculin, is a promoter binding protein that

is reported to have roles in atherosclerosis [55], hypertension, hypercholesterolemia [56], and Alzheimer's disease [57,58]. AP15 prevents apoptosis in the absence of growth factor [59–62], while the BECN1-PIK3C3 complex plays a crucial role in autophagy [63]. The GOSR1 protein is responsible for cellular trafficking [64] and is frequently upregulated in esophageal squamous cell carcinoma tissues [65]. PRPF40A is associated with pre-mRNA splicing [66,67], genetic diseases such as Rett syndrome [68], Huntington's disease [69], and cancers [70] such as lung cancer [71] and pancreatic ductal adenocarcinoma [72].

Similarly, in tumor samples, the two genes are part of the same gene cluster in HNSC, LUAD, STAD, and UCEC. Genes overrepresented in these clusters include YTHDC1, SRSF10, BLCAF1, MFSD8, and ABHD18. YTHDC1 is a nuclear protein involved in splicing cancer-causing transcripts [73] and plays a regulatory role in several cancers such as breast and prostate cancer [74–77]. SRSF10 is an SR protein and splicing regulator that activates splicing when phosphorylated and inhibits splicing when dephosphorylated [78,79]. Bclaf1 is a tumor suppressor gene [80] involved in T-cell activation [81], repairing DNA damage [82,83], and pre-mRNA splicing [84], with a regulatory role in colon cancer [79]. Alterations in MFSD8 have been associated with a neurodegenerative disorder called vLINCL, which causes seizures, progressive mental and motor deterioration, myoclonus, visual failure, and premature death [85–89]. The ABHD18 protein is a genetic marker for hepatocellular carcinoma (HCC) in Asian populations [90].

Next, we set to identify the functional association of the THAP9 and THAP9-AS1 co-expression modules. We used 'ShinyGO', which performs an in-depth analysis of gene lists that includes a graphical visualization of enrichment, pathway, gene characteristic, and protein interactions [91]. The noteworthy pathways from GO analysis for THAP9 and THAP9-AS1 are visualized in Figures 10 and 11, respectively (Supplementary Tables S5 and S6).

THAP9: As presented in Figure 10 (Supplementary Table S5), in normal individuals, genes co-expressed with THAP9 were involved in RNA biosynthesis and organelle organization (Figure 10a). The co-expressed genes in both normal and tumor samples were markedly enriched in the nucleoplasm and nuclear lumen (Figure 10b) and were significantly involved in binding nucleic acids (especially DNA) (Figure 10c); this is interesting as THAP9 also appears to localize in the nucleus [92] and possibly binds DNA via an amino terminal DNA-binding THAP domain [17]. We also observed the enrichment of several KEGG pathways associated with (Figure 10d) Herpes simplex virus 1 infection (normal and tumor), as well as neurodegenerative disorders such as Alzheimer's disease, Parkinson's disease, and Huntington's disease (normal samples).

THAP9-AS1: Similarly, Figure 11 (Supplementary Table S6) shows that genes co-expressed with THAP9-AS1 in both normal and tumor samples were markedly enriched in the nucleoplasm and nuclear lumen (Figure 11b) and associated with herpes simplex virus 1 infection (Figure 11d). In contrast, it appeared that the co-expressed genes were significantly enriched in cilia only in the tumor samples and involved in RNA binding in normal samples.

2.5.2. Differential Gene Correlation Analysis

A differential gene correlation or co-expression analysis can identify biologically important differentially correlated genes that cannot be detected using a regular gene co-expression analysis or differential gene expression analyses. It is suggested that if there is a change in the correlation between the expression of two genes under certain conditions (i.e., they are differentially correlated), they possibly regulate or are regulated by the condition [93–95]. Many studies have used differential correlation analyses to identify genes underlying differences between healthy and diseased samples or between different tissues, cell types, or species [96–99]. Genes that are functionally related tend to have similar expression profiles; therefore, a differential gene correlation analysis that can compare the expression correlation of THAP9 and THAP9-AS1 with other genes in normal vs. tumor samples can give us insight into biological processes and molecular pathways that distinctly involve the two genes in the two conditions.

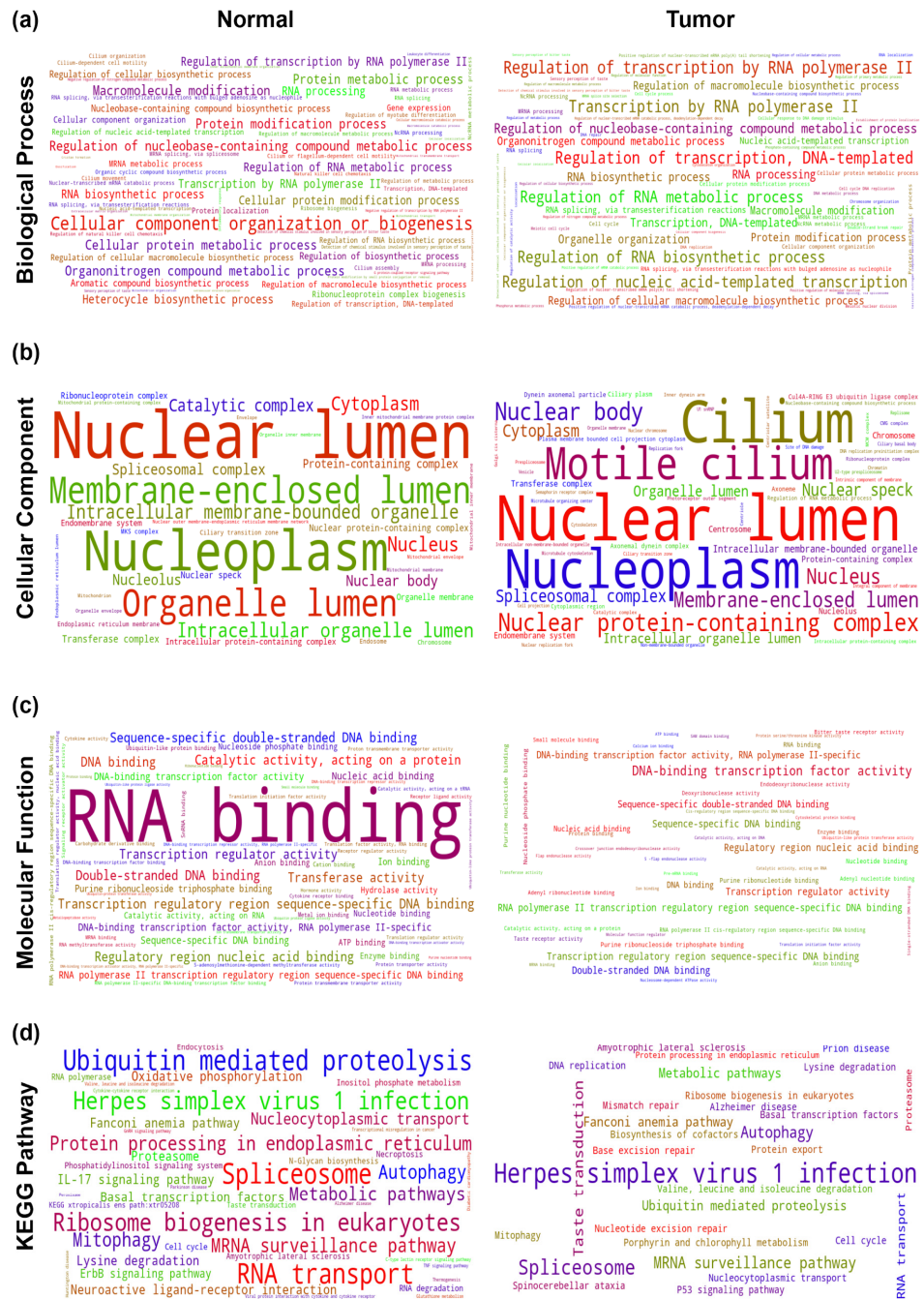


Figure 11. Gene ontology (GO) and KEGG pathway analyses of genes co-expressed with THAP9-AS1 in normal vs. tumor samples. The enrichment test was performed for normal vs. tumor samples (for each cancer) using ShinyGO for the top 10 enriched terms, with the significance cutoff for adjusted *p*-values being set at 0.05. The font sizes in the word cloud are proportional to their frequency after the enrichment rates were merged for all cancers (left side for normal samples and right side for tumor samples). Word clouds of enriched GO terms in (a) the biological process category, (b) cellular component category, (c) molecular functions category, and (d) KEGG pathways.

In this study, we used a DGCA to identify the genes differentially correlated with THAP9 (Supplementary Table S10). and THAP9-AS1(Supplementary Table S11) under various tumor vs. paired normal conditions. We used the RNA-seq HTSeq count dataset of 22 cancer and paired normal samples from TCGA. A pan-cancer analysis of head-to-head gene pairs [100] reported that these gene pairs show significantly stronger positive

correlations in tumor compared to normal samples, regardless of tumor types. Moreover, bidirectional promoters are known to regulate the expression of many cancer-related genes such as BRCA1 and TP53 [101,102]. Thus, we calculated the correlation between the THAP9 and THAP9-AS1 H2H genes.

Interestingly, when we compared the expression patterns of THAP9 and THAP9-AS1 in normal and tumor samples, they always showed a positive correlation in each cancer type, suggesting their coordinated regulation (Supplementary Figure S4). It is noteworthy that we have also predicted a bidirectional promoter region between the two genes (Figure 1).

We then calculated the differences in Spearman correlations for all genes with THAP9 and THAP9-AS1 to identify the genes differentially correlated with THAP9 and THAP9-AS1 between the tumor and the paired normal samples in each cancer type. Further, we measured the Gene Ontology (GO) enrichment of the genes differentially correlated with THAP9 and THAP9-AS1 with a gain and loss of correlation in tumor vs. normal samples.

THAP9: Looking at the combined results from all the cancers (Figure 12, results for each cancer separately in Supplementary Figure S5), genes that lost correlation with THAP9 were enriched in nuclear chromatin (Figure 12b) and often involved in processes such as DNA-mediated transposition, negative regulation of gene expression, and ion channel activity (Figure 12a,c).

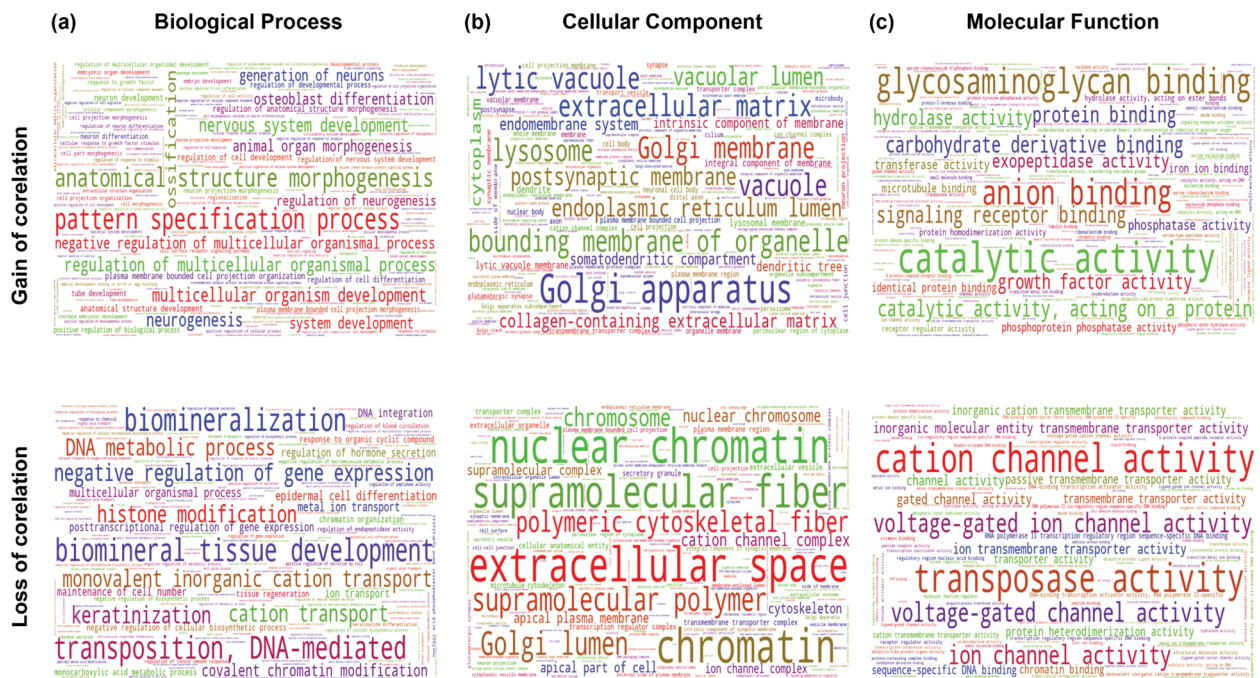


Figure 12. Gene Ontology analysis of genes differentially correlated with THAP9 in normal vs. tumor samples (top—genes that gained a correlation with THAP9; bottom—genes that lost a correlation with THAP9). Word clouds of enriched (a) GO biological process, (b) GO cellular component, and (c) GO molecular functions. The differential gene correlation analysis, followed by the gene ontology analysis for the differentially correlated genes, was performed using the DGCA package in R.

THAP9-AS1: In all cancer samples (Figure 13, results for each cancer separately in Supplementary Figure S6), genes that gained in correlation with THAP-AS1 were enriched in immune system processes (Figure 13a,c). Moreover, genes that showed a correlated expression (gain or loss) with THAP9-AS1 were enriched in the plasma membrane and cytoplasm.

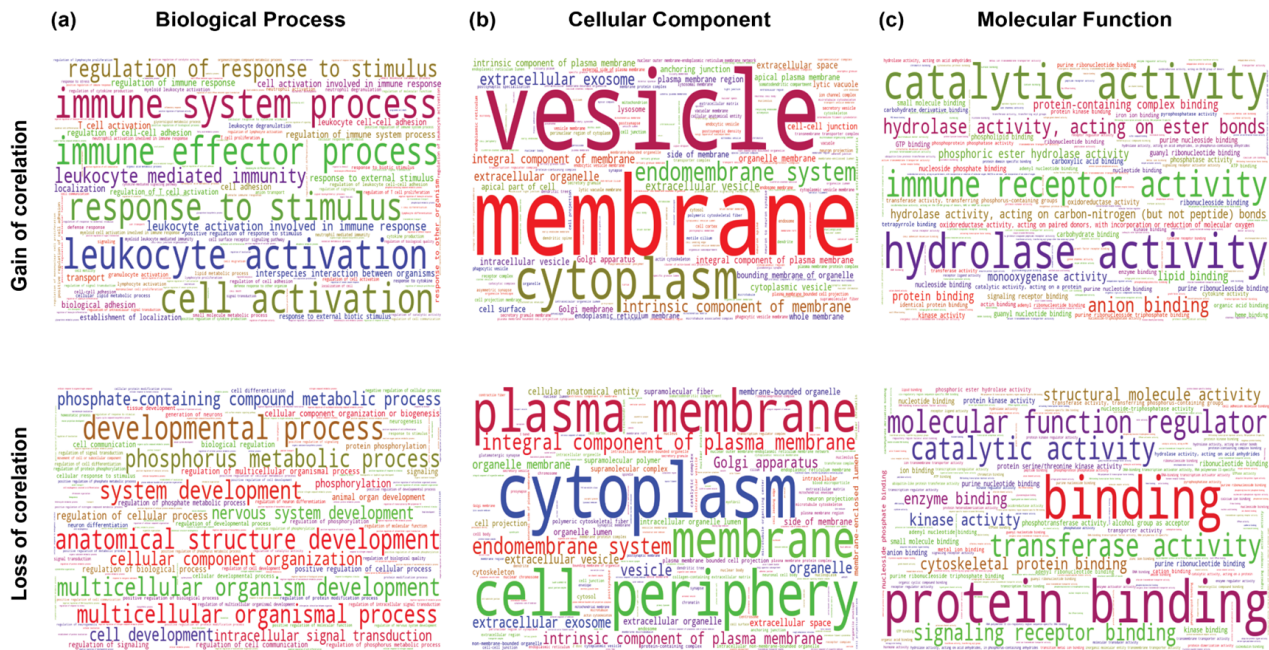


Figure 13. Gene Ontology analysis of genes differentially correlated with THAP9-AS1 in normal vs. tumor samples (top—genes that gained correlation with THAP9-AS1; bottom—genes that lost correlation with THAP9). Word clouds of enriched (a) GO biological process, (b) GO cellular component, and (c) GO molecular functions. The differential gene correlation analysis, followed by the gene ontology analysis for the differentially correlated genes, was performed using the DGCA package in R.

3. Discussion

This study investigated the pan-cancer expression patterns of THAP9 and THAP9-AS1, which are a pair of sense–antisense genes that occur in a “head-to-head” orientation on chromosome 4q21. The human genome contains numerous pairs of genes with similar “head-to-head” orientations with transcription start sites separated by less than 1 kb [2]. Several of these gene pairs are regulated by a single bidirectional promoter [103]. Bidirectional promoters typically have high GC contents, frequently lack TATA boxes, and are often conserved among mouse orthologs [44,104]. Interestingly, these structural features are also present in the predicted bidirectional promoter region of the THAP9/THAP9-AS1 gene pair.

Bidirectional promoters may regulate the coordinated expression of two genes (within a gene pair) that have complementary roles [105] and help maintain stoichiometric quantities of each gene’s expression [103]. They are also responsible for driving the transcription of genes involved in the same cellular pathway or genes that need to be sequentially activated [52,105]. Thus, we decided to investigate whether the expression levels of THAP9 and THAP9-AS1 were correlated and if the gene pair was possibly regulated by their putative bidirectional promoter.

It has been reported that THAP9 is a highly conserved gene, which has been identified in 178 organisms [106]. The human THAP9 gene has 6 isoforms, out of which only one is known to encode for a protein that is homologous to the *Drosophila* P-element transposase. hTHAP9 belongs to the THAP (Thanatos-associated protein) protein family in humans, containing twelve proteins (hTHAP0–hTHAP11) [17]. Many THAP family proteins are known to be involved in human diseases. THAP1 has been associated with DYT6 dystonia [18], THAP5 and THAP1 have been linked to apoptosis [19,20], the LRRC49/THAP10 bidirectional gene pair is involved in breast cancer [21], and THAP11 has been implicated in colon and gastric cancers [22,107]. THAP9-AS1 (THAP9 antisense) is a newly annotated (by Ensembl) lncRNA coding gene that encodes 12 long non-coding RNAs. Recent reports have

suggested that the THAP9-AS1 lncRNA is involved in pancreatic cancer, septic shock, and neutrophil apoptosis [25,26]. However, the roles of THAP9 and THAP9-AS1 across human cancers are not well understood. This study investigated the relationship between THAP9 and THAP9-AS1 expression and their possible roles in tumorigenesis via a pan-cancer analysis of TCGA and GTEx databases.

The gene expression analysis using TIMER2 and GEPIA2 suggested that both over- and under-expression of THAP9 and THAP9-AS1 frequently occurred in various cancers. We observed that THAP9 was upregulated in CHOL, COAD, ESCA, LIHC, LUSC, LUAD, STAD, and THYM and downregulated in KIRC, KIRP, PRAD, THCA, TGCT, and UCEC. On the other hand, THAP9-AS1 was upregulated in CHOL, THYM, DLBC, and PAAD, while it was downregulated in OV, SKCM, and THCA. We also observed that compared with the corresponding normal tissues, THAP9 and THAP9-AS1 expression levels were coordinately upregulated in CHOL and THYM but coordinately downregulated in THCA. Therefore, the independent and coordinated alteration in THAP9 and THAP9-AS1 expression in various cancers indicates that they may have different biological functions in different cancers. Regardless, the aberrant expression levels of the THAP9 and THAP9-AS1 gene pair were associated with a poor prognosis in many types of cancer, which suggested their role as a potential prognostic cancer biomarker. Moreover, we observed that both the THAP9 and THAP9-AS1 genes were mutated (often amplified) in several cancers (TCGA dataset).

Comprehensive gene expression studies can help in predicting gene function. Genes that share an expression pattern, i.e., are turned on or off together under various conditions, may encode proteins that constitute the same multiprotein machine or are involved in a complex, coordinated activity. Characterizing an unknown gene's function by grouping it with known genes that share its transcriptional behavior is called a "guilt by association analysis (GBA)" [108], which can be simply explained as "a man is known by the company he keeps". A GBA involves a gene co-expression analysis followed by gene interaction network construction by clustering associations from gene co-expression data. Genes that belong to the same cluster may be involved in common cellular pathways or processes.

Our GO analysis suggested that many of the genes that were co-expressed with THAP9 were also involved in DNA and metal binding, much like THAP9 homologs like *Drosophila* P-element transposase, which binds DNA via a characteristic zinc-finger-type THAP domain. KEGG pathway analysis of genes co-expressed with THAP9 and THAP9-AS1 demonstrated the enrichment of pathways related to Herpes simplex virus 1 infection as well as several neurodegenerative disorders like Alzheimer's disease, Parkinson's disease, and Huntington's disease. It has been suggested that herpes simplex virus 1 infection may be a causative agent of Alzheimer's disease [109]. THAP9 has previously been reported to be upregulated (5-fold) in tuberculous meningitis (TBM) patients co-infected with HIV compared to patients with TBM alone [110]. It will be interesting to investigate the possible role of THAP9 in neurological disorders.

Our analysis demonstrates that although THAP9 and THAP9-AS1 show diverse expression patterns in different cancers, their expression in normal and tumor samples was positively correlated in each cancer type. This suggests the coordinated regulation of the two genes (Supplementary Figure S4). It is tempting to speculate that this coordinated regulation is mediated by the predicted bidirectional promoter region between the two genes (Figure 1). Moreover, the differential expression of THAP9 and THAP9-AS1 in different tumor types suggests that the two genes may have tumor-specific regulatory mechanisms.

4. Materials and Methods

4.1. Analysis of Promoter Sequence

The promoter sequences were downloaded using EPDnew [38], which is a new section under the well-known Eukaryotic Promoter Database (EPD) (<https://epd.epfl.ch> accessed on 13 September 2021) [38]. The EPD is an annotated non-redundant collection of eukaryotic POL II promoters where the transcription start site (TSS) has been determined

experimentally. The core promoter elements in the region were identified using the Elements Navigation Tool (ElementNT) [45], a user-friendly, web-based interactive tool for predicting and displaying putative core promoter elements and their biologically relevant combinations. ElementNT's predictions are based on biologically functional core promoter elements and can infer core promoter compositions. ElementNT does not assume prior knowledge of the actual TSS position and can annotate any given sequence.

The bidirectional promoter region identified the CpG islands and other epigenetic marks using the UCSC Genome Browser (<http://genome.ucsc.edu/> accessed on 21 September 2021). The CpG islands were plotted using EMBOSS Cpgplot (https://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot accessed on 15 September 2021) [1,2,111,112].

4.2. Mutation Analysis in Different Types of Tumors

The cBioPortal web server (<http://www.cbioportal.org> accessed 30 November 2021) [34] is a comprehensive website that explores, visualizes, and analyzes multidimensional cancer genomics data. We used the “cancer types summary” module on the “TCGA PanCanAtlas” dataset available on the cBioPortal web server. Furthermore, the correlation between the genetic alteration of the two genes and their overall survival prognosis was explored in the “comparison” module. The prognostic values are presented with log-rank p -values.

4.3. Gene Expression Analysis

A differential gene expression analysis, to investigate the changes in expression of specific genes in tumor vs. normal samples, was performed using TIMER2.0 [32] and GEPIA2 [33] online tools.

4.3.1. TIMER2.0

TIMER2.0 (Tumor Immune Estimation Resource, version 2; <http://timer.cistrome.org/> accessed on 15 October 2021) is a comprehensive resource for systematically analyzing differential gene expression levels between tumor and adjacent normal tissues. We used “THAP9” and “THAP9-AS1” as the inputs in the “Gene_DE” module to evaluate the expression levels of the two genes in tumor tissue and adjacent normal tissues from 32 cancer types from TCGA [30]. “THAP9-AS1” was not available in TIMER2.0.

4.3.2. GEPIA2

GEPIA2 (Gene Expression Profiling Interactive Analysis, version 2; <http://gepia2.cancer-pku.cn/#analysis> accessed on 16 October 2021) is an interactive web server for analyzing mRNA expression data from tumors and normal samples from TCGA (the Cancer Genome Atlas) and GTEx (Genotype–Tissue Expression) projects. We used the “expression analysis box plots” module of GEPIA2 to obtain box plots of THAP9 and THAP9-AS1 expression levels between tumor and normal tissues. We set the p -value cutoff as 0.01, the log₂FC (fold change) cutoff as 1, and used the “match TCGA normal and GTEx data” option.

4.4. Prognostic Analysis of THAP9 and THAP9-AS1

The GEPIA2 webserver was used to explore the prognostic values of THAP9 and THAP9-AS1 in different types of tumors in TCGA. The “survival map” module of GEPIA2 was used to obtain the overall survival (OS) and disease-free survival (DFS) significance map data with cutoff-high (50%) and cutoff-low (50%) values to split the high-expression and low-expression cohorts. The survival data were visualized with hazard ratio, 95% confidence interval, and log-rank p -values.

4.5. Guilt by Association Analysis

4.5.1. Construction of Weighted Gene Co-Expression Network

GBA (guilt by association) analysis [53] was used to identify co-expressing genes in each tumor and associated normal samples.

Sample Collection: We downloaded HTSeq-counts RNA-Seq data of 33 tumor types from TCGA (the Cancer Genome Atlas) database (<https://portal.gdc.cancer.gov/> accessed on 20 October 2021) using the GDC Data Transfer Tool (<https://gdc.cancer.gov/access-data/gdc-data-transfer-tool> accessed on 20 October 2021). The row names of the downloaded HTSeq-count matrix were Ensembl gene identifiers, and the column names represented TCGA sample IDs. The dataset included the expression profiles of 60,483 genes from 11,094 patients. Constraints used to generate the manifest file to be used with GDC Data Transfer Tool are as follows: data category: transcriptome profiling; data type: gene expression quantification; experimental strategy: RNA-Seq; workflow type: HTSeq-counts; data format: txt; access: open; program: TCGA. Details about the samples can be found in Supplementary Table S4: Cancer types with less than 3 normal samples (rows highlighted in red)] were excluded; thus we finally analyzed 22 cancer types.

Sample Preprocessing: The Ensembl gene IDs link to gene information in the Ensembl database [37]. The “org.Hs.eg.db” R Bioconductor package [111] was used to convert the Ensembl gene IDs to the gene symbols. Ensembl IDs that did not have an official gene symbol were dropped from the analysis. After filtering the samples, we were left with gene expression values of 34,125 genes across 11,094 samples.

WGCNA analysis: The gene expression values of each tumor and paired normal samples were subjected to the WGCNA, R Bioconductor package [35] for the weighted co-expression network construction. We used the “blockwiseModules” function in WGCNA, which performs automatic network construction and module detection processes on large expression datasets in a block-wise manner. In summary, it calculates the similarity matrix between each pair of genes across all samples based on its Pearson’s correlation value. Then, the similarity matrix is transformed into an adjacency matrix. Subsequently, the topological overlap matrix (TOM) and the corresponding dissimilarity (1-TOM) value are computed. Finally, a dynamic tree cut (DTC) algorithm detects gene co-expression modules. WGCNA distinguishes gene clusters by color names (tan, turquoise, brown etc.). The signed modules were constructed with a cut height of 0.995 and a minimum module size of 30 genes. Then, we used the modules associated with THAP9 and THAP9-AS1 in Cytoscape [112] to visualize the top 20 genes co-expressed with THAP9 and THAP9-AS1 in each tumor vs. normal pair.

4.5.2. Gene Ontology (GO) and KEGG Pathway Enrichment Analysis

A GO analysis [113] is a helpful method for annotating genes and gene sets with biological characteristics for high-throughput genome or transcriptome data. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway [114] is a knowledge base for the systematic analysis of gene functions. A GO and KEGG pathway enrichment analysis was performed using the “ShinyGO” web server [91], a Shiny application developed based on several R/Bioconductor packages. A *p*-value cutoff (FDR) of 0.05 was set as the cut-off criterion for extracting the top 10 enriched GO terms (biological process (BP), cellular component (CC), and molecular function (MF)) and KEGG pathways. Further, we merged all of the GO-BP, GO-CC, GO-MF, and KEGG pathways enriched in all tumor and normal samples and used the “word cloud” package in python to visualize the overall enrichment GO and KEGG pathways in normal vs. tumor samples.

4.5.3. Differential Correlation Analysis

For the differential co-expression analysis between the tumor and normal samples in each cancer, R-package DGCA version 1.0.2 [36] was used. DGCA is an R package designed to detect differences in the correlations of gene pairs between distinct biological conditions. DGCA uses correlation coefficients transformed into normalized Z-scores to identify differentially correlated genes and modules while performing downstream analysis, including data visualization, GO enrichment, and network construction tools.

Firstly, we checked the correlation between THAP9 and THAP9-AS1 in each tumor vs. normal pair using the “plotCors” function in DGCA, which uses Pearson’s correlation by default.

Following this, genes that were differentially correlated with THAP9 and THAP9-AS1 were computed with the “ddCorAll” function using “corrType” as the Spearman correlation. This pipeline provided the Spearman coefficient and the corresponding *p*-values for each pair of genes across samples. Significant changes in differential correlation between the two conditions (tumor vs. normal) were then identified using a Fisher’s Z-test. The correlation between THAP9/THAP9-AS1 and other genes was classified as having a gain of correlation or loss of correlation and based upon the threshold for correlation significance; the gene pairs were grouped into nine different correlation classes (+/+; +/-; +/0; -/+; -/0; -/-; 0/+; 0/0; 0/-). The classes show the correlation as positive (+), negative (-), or not significant (0) for each gene and condition when contrasting the groups (tumor/normal). A GO term enrichment analysis of differential correlation-classified genes was performed using the DGCA function “ddcorGO.”

5. Conclusions

We conducted a pan-cancer analysis of the THAP9 and THAP9-AS1 gene pair in various cancers. We explored the association of their aberrant expression with patient survival outcomes, followed by analyzing their functional association using gene co-expression and differential gene correlation analysis. This study has several limitations. Firstly, although we used TCGA and GTEx datasets, data about particular cancer types were not available. Secondly, given the myriad individual differences among cancer patients, it was challenging to cover all possible variations. Finally, our analysis was solely computational and relied on public databases. Future studies to validate the expression and function of the two genes at the cellular and molecular levels will shed more light on their physiological and pathological relevance.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ncrna8040051/s1>: Supplementary Table S1: Predicted core promoters for THAP9 (Element results). The predicted promoter sequence was downloaded from EPDnew -250 to 250 relative to the TSS predicted by EPDnew which is located at position 82900735 on the sense strand of chromosome 4; Supplementary Table S2: Predicted core promoters for THAP9-AS1 (Element results). The predicted promoter sequence was downloaded from EPDnew -400 to 100 relative to the TSS predicted by EPDnew, assuming position 82900569 as TSS (the selected sequence includes both TSSs for THAP9-AS1) from EPDnew; Supplementary Table S3: Details of cancer-related mutations in THAP9 (downloaded from CBioPortal, Dataset: ‘Pan-cancer analysis of whole genomes (ICGC/TCGA, Nature 2020)’); Supplementary Table S4: Information for samples used for WGCNA, and DGCA analysis. Row names refer to the TCGA Cancer IDs and column names refer to the types of samples; Supplementary Table S5: Top 10 enriched GO Terms and KEGG pathways for genes coexpressing with THAP9 in various conditions calculated using ShinyGO. TCGA cancer type (1st column); GO & KEGG Pathway terms (2nd column) where BP is Biological Processes, CC is Cellular Components & MF is Molecular Functions; sample type, tumor, or normal (3rd column); top 10 enriched terms (last column); Supplementary Table S6: Top 10 enriched GO Terms and KEGG pathways for genes coexpressing with THAP9-AS1 in various conditions calculated using ShinyGO. TCGA cancer type (1st column); GO & KEGG Pathway terms (2nd column) where BP is Biological Processes, CC is Cellular Components & MF is Molecular Functions; sample type, tumor, or normal (3rd column); top 10 enriched terms (last column); Supplementary Table S7: Gene cluster (obtained by WGCNA analysis) associated with THAP9 and THAP9-AS1 in normal and tumor conditions in various cancers (TCGA dataset). Each cluster is represented by a color in WGCNA; Supplementary Table S8: Genes coexpressing (obtained by WGCNA R package) with THAP9 in various conditions. Each sheet is named as “CancerType_SampleType” and contains 5 columns THAP9 (column 1), genes coexpressing with THAP9 (column 2), correlation between THAP9 and genes coexpressing with THAP9 (column 3), WGCNA molecule colors of THAP9 (column 4) and coexpressing gene (column 5); Supplementary Table S9: Genes coexpressing (obtained by WGCNA R package) with THAP9-AS1 in various conditions. Each sheet contains 5 columns THAP9-AS1 (column 1), genes coexpressing with THAP9-AS1 (column 2), correlation between THAP9-AS1 and genes coexpressing with THAP9-AS1 (column 3), WGCNA molecule colors of THAP9-AS1 (column 4) and coexpressing gene (column 5); Supplementary Table S10_1 & Supplementary Table S10_2: Differential gene correla-

tions analysis (calculated using DGCA R package) for THAP9. Each sheet (named by TCGA cancer id) contains details of genes differentially correlated (normal to tumor) with THAP9 in individual cancers; Supplementary Table S11_1 & Supplementary Table S11_2: Differential gene correlations analysis (calculated using DGCA R package) for THAP9-AS1. Each sheet (named by TCGA cancer id) contains details of genes differentially correlated (normal to tumor) with THAP9-AS1 in individual cancers; Supplementary Figure S1: THAP9 & THAP9-AS1 Promoter analysis (predicted by EPDnew database). Details of Transcription Start Site and Promoter sequence for (a) THAP9, (b) THAP9-AS1. EPDnew predicts two TSS for THAP9-AS1. CpG Islands (predicted by EMBOSS CPGplot) (c) THAP9 (d) THAP9-AS1; Supplementary Figure S2: Boxplot (generated using GEPIA2) of differential gene expression profile of THAP9 across 31 cancer types (TCGA and GTEx datasets). Y-axis: $\log(\text{TPM}+1)$ transformed gene expression values for THAP9, x-axis: various cancer types (Red: tumor samples, Grey: normal samples); Supplementary Figure S3: Boxplot (generated using GEPIA2) of differential gene expression profile of THAP9-AS1 across 31 cancer types (TCGA and GTEx datasets). Y-axis: $\log(\text{TPM}+1)$ transformed gene expression values for THAP9-AS1, x-axis: various cancer types (Red: tumor samples, Grey: normal samples); Supplementary Figure S4: Plots (generated using DGCA) representing the correlation between THAP9 and THAP9-AS1 gene expression in various cancers in cancer (red plot) vs. normal (green plot) samples. X-axis: gene expression profile of THAP9, y-axis: gene expression profile of THAP9-AS1; Supplementary Figure S5: Gene ontology enrichment of genes associated with gain or loss of correlation with THAP9 in normal vs. tumor samples in various cancers. Each plot (generated using DGCA R package) contains one TCGA cancer type (labelled on top of the plot). Up to 5 enriched GO terms are shown; Supplementary Figure S6: Gene ontology enrichment of genes associated with gain or loss of correlation with THAP9-AS1 in normal vs. tumor samples in various cancers. Each plot (generated using DGCA R package) contains one TCGA cancer type (labelled on top of the plot). Up to 5 enriched GO terms are shown.

Author Contributions: Conceptualization, writing review and editing S.M. and R.R.; methodology, validation, formal analysis, original draft preparation, visualization R.R.; supervision, project administration, funding acquisition, S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by IIT Gandhinagar, SERB (ECR/2016/000479), DBT [Ramalingaswami Fellowship BT/RLF/Re-entry/43/2013, BT/PR16074/BID/7/569/2016), CISCO (CG# 2207376), GSBTM (RSS 2019-20).

Data Availability Statement: Not applicable.

Acknowledgments: We would like to thank Raviraj Sukhadiya for technical support and Ashutosh Srivastava and Anjali Rajwar for their feedback and suggestions.

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

Abbreviations

ACC	adrenocortical carcinoma;
BLCA	bladder urothelial carcinoma;
BRCA	breast invasive carcinoma;
CESC	cervical and endocervical cancers;
CHOL	cholangiocarcinoma;
COAD	colon adenocarcinoma;
DLBC	lymphoid neoplasm diffuse large B-cell lymphoma;
ESCA	esophageal carcinoma;
GBM	glioblastoma multiforme;
HNSC	head and neck squamous cell carcinoma;
KICH	kidney chromophobe;
KIRC	kidney renal clear cell carcinoma;
KIRP	kidney renal papillary cell carcinoma;
LAML	acute myeloid leukemia;
LGG	brain lower grade glioma;

LIHC	liver hepatocellular carcinoma;
LUAD	lung adenocarcinoma;
LUSC	lung squamous cell carcinoma;
MESO	mesothelioma;
OV	ovarian serous cystadenocarcinoma;
PAAD	pancreatic adenocarcinoma;
PCPG	pheochromocytoma and paraganglioma;
PRAD	prostate adenocarcinoma;
READ	rectum adenocarcinoma;
SARC	sarcoma;
SKCM	skin cutaneous melanoma;
STAD	stomach adenocarcinoma;
STES	stomach and esophageal carcinoma;
TGCT	testicular germ cell tumors;
THCA	thyroid carcinoma;
THYM	thymoma;
UCEC	uterine corpus endometrial carcinoma;
UCS	uterine carcinosarcoma;
UVM	uveal melanoma.

References

- Hurst, L.D.; Pál, C.; Lercher, M.J. The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* **2004**, *5*, 299–310. [[CrossRef](#)] [[PubMed](#)]
- Adachi, N.; Lieber, M.R. Bidirectional Gene Organization: A Common Architectural Feature of the Human Genome. *Cell* **2002**, *109*, 807–809. [[CrossRef](#)]
- Li, Y.-Y.; Yu, H.; Guo, Z.-M.; Guo, T.-Q.; Tu, K.; Li, Y.-X. Systematic Analysis of Head-to-Head Gene Organization: Evolutionary Conservation and Potential Biological Relevance. *PLoS Comput. Biol.* **2006**, *2*, e74. [[CrossRef](#)]
- Trinklein, N.D.; Aldred, S.F.; Hartman, S.J.; Schroeder, D.I.; Otililar, R.P.; Myers, R.M. An Abundance of Bidirectional Promoters in the Human Genome. *Genome Res.* **2004**, *14*, 62–66. [[CrossRef](#)]
- Burbelo, P.D.; Martin, G.R.; Yamada, Y. Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 9679–9682. [[CrossRef](#)]
- Heikkilä, P.; Soinen, R.; Tryggvason, K. Directional regulatory activity of cis-acting elements in the bidirectional alpha 1(IV) and alpha 2(IV) collagen gene promoter. *J. Biol. Chem.* **1993**, *268*, 24677–24682. [[CrossRef](#)]
- Schuettengruber, B.; Doetzlhofer, A.; Kroboth, K.; Wintersberger, E.; Seiser, C. Alternate activation of two divergently transcribed mouse genes from a bidirectional promoter is linked to changes in histone modification. *J. Biol. Chem.* **2003**, *278*, 1784–1793. [[CrossRef](#)]
- Hansen, J.J.; Bross, P.; Westergaard, M.; Nielsen, M.N.; Eiberg, H.; Børglum, A.D.; Mogensen, J.; Kristiansen, K.; Bolund, L.; Gregersen, N. Genomic structure of the human mitochondrial chaperonin genes: HSP60 and HSP10 are localised head to head on chromosome 2 separated by a bidirectional promoter. *Hum. Genet.* **2003**, *112*, 71–77. [[CrossRef](#)]
- Balbin, O.A.; Malik, R.; Dhanasekaran, S.M.; Prensner, J.R.; Cao, X.; Wu, Y.-M.; Robinson, D.; Wang, R.; Chen, G.; Beer, D.G.; et al. The landscape of antisense gene expression in human cancers. *Genome Res.* **2015**, *25*, 1068–1079. [[CrossRef](#)]
- Auriol, E.; Billard, L.-M.; Magdinier, F.; Dante, R. Specific binding of the methyl binding domain protein 2 at the BRCA1-NBR2 locus. *Nucleic Acids Res.* **2005**, *33*, 4243–4254. [[CrossRef](#)]
- Luo, L.; Lu, F.M.; Hart, S.; Foroni, L.; Rabbani, H.; Hammarström, L.; Yuille, M.R.; Catovsky, D.; Webster, A.D.; Vorechovský, I. Ataxia-telangiectasia and T-cell leukemias: No evidence for somatic ATM mutation in sporadic T-ALL or for hypermethylation of the ATM-NPAT/E14 bidirectional promoter in T-PLL. *Cancer Res.* **1998**, *58*, 2293–2297. [[PubMed](#)]
- Shinya, E.; Shimada, T. Identification of two initiator elements in the bidirectional promoter of the human dihydrofolate reductase and mismatch repair protein 1 genes. *Nucleic Acids Res.* **1994**, *22*, 2143–2149. [[CrossRef](#)] [[PubMed](#)]
- Chen, P.-Y.; Chang, W.-S.W.; Chou, R.-H.; Lai, Y.-K.; Lin, S.-C.; Chi, C.-Y.; Wu, C.-W. Two non-homologous brain diseases-related genes, SERPINI1 and PDCD10, are tightly linked by an asymmetric bidirectional promoter in an evolutionarily conserved manner. *BMC Mol. Biol.* **2007**, *8*, 2. [[CrossRef](#)]
- Majumdar, S.; Singh, A.; Rio, D.C. The Human THAP9 Gene Encodes an Active P-Element DNA Transposase. *Science* **2013**, *339*, 446–448. [[CrossRef](#)]
- Majumdar, S.; Rio, D.C. P transposable elements in *Drosophila* and other eukaryotic organisms. *Microbiol. Spectr.* **2015**, *3*, MDNA3-0004-2014. [[CrossRef](#)]
- Campagne, S.; Saurel, O.; Gervais, V.; Milon, A. Structural determinants of specific DNA-recognition by the THAP zinc finger. *Nucleic Acids Res.* **2010**, *38*, 3466–3476. [[CrossRef](#)] [[PubMed](#)]
- Sabogal, A.; Lyubimov, A.Y.; Corn, J.E.; Berger, J.M.; Rio, D.C. THAP proteins target specific DNA sites through bipartite recognition of adjacent major and minor grooves. *Nat. Struct. Mol. Biol.* **2010**, *17*, 117–123. [[CrossRef](#)]

18. Sengel, C.; Gavarini, S.; Sharma, N.; Ozelius, L.J.; Bragg, D.C. Dimerization of the DYT6 dystonia protein, THAP1, requires residues within the coiled-coil domain. *J. Neurochem.* **2011**, *118*, 1087–1100. [[CrossRef](#)]
19. Balakrishnan, M.P.; Cilenti, L.; Mashak, Z.; Popat, P.; Alnemri, E.S.; Zervos, A.S. THAP5 is a human cardiac-specific inhibitor of cell cycle that is cleaved by the proapoptotic Omi/HtrA2 protease during cell death. *Am. J. Physiol. Heart Circ. Physiol.* **2009**, *297*, H643–H653. [[CrossRef](#)]
20. Roussigne, M.; Kossida, S.; Lavigne, A.-C.; Clouaire, T.; Ecochard, V.; Glories, A.; Amalric, F.; Girard, J.-P. The THAP domain: A novel protein motif with similarity to the DNA-binding domain of P element transposase. *Trends Biochem. Sci.* **2003**, *28*, 66–69. [[CrossRef](#)]
21. Santos, E.D.; de Bessa, S.A.; Netto, M.M.; Nagai, M.A. Silencing of LRRC49 and THAP10 genes by bidirectional promoter hypermethylation is a frequent event in breast cancer. *Int. J. Oncol.* **2008**, *33*, 25–31. [[CrossRef](#)]
22. Zhang, J.; Zhang, H.; Shi, H.; Wang, F.; Du, J.; Wang, Y.; Wei, Y.; Xue, W.; Li, D.; Feng, Y.; et al. THAP11 Functions as a Tumor Suppressor in Gastric Cancer through Regulating c-Myc Signaling Pathways. *BioMed Res. Int.* **2020**, *2020*, e7838924. [[CrossRef](#)] [[PubMed](#)]
23. Li, X.-X.; Liang, X.-J.; Zhou, L.-Y.; Liu, R.-J.; Bi, W.; Zhang, S.; Li, S.-S.; Yang, W.-H.; Chen, Z.-C.; Yang, X.-M.; et al. Analysis of Differential Expressions of Long Non-coding RNAs in Nasopharyngeal Carcinoma Using Next-generation Deep Sequencing. *J. Cancer* **2018**, *9*, 1943–1950. [[CrossRef](#)] [[PubMed](#)]
24. Li, X.-X.; Wang, L.-J.; Hou, J.; Liu, H.-Y.; Wang, R.; Wang, C.; Xie, W.-H. Identification of Long Noncoding RNAs as Predictors of Survival in Triple-Negative Breast Cancer Based on Network Analysis. *Biomed. Res. Int.* **2020**, *2020*, 8970340. [[CrossRef](#)]
25. Li, N.; Yang, G.; Luo, L.; Ling, L.; Wang, X.; Shi, L.; Lan, J.; Jia, X.; Zhang, Q.; Long, Z.; et al. lncRNA THAP9-AS1 Promotes Pancreatic Ductal Adenocarcinoma Growth and Leads to a Poor Clinical Outcome via Sponging miR-484 and Interacting with YAP. *Clin. Cancer Res.* **2020**, *26*, 1736–1748. [[CrossRef](#)]
26. Jia, W.; Zhang, J.; Ma, F.; Hao, S.; Li, X.; Guo, R.; Gao, Q.; Sun, Y.; Jia, J.; Li, W. Long noncoding RNA THAP9-AS1 is induced by *Helicobacter pylori* and promotes cell growth and migration of gastric cancer. *Onco. Targets Ther.* **2019**, *12*, 6653–6663. [[CrossRef](#)]
27. Jiang, N.; Zhang, X.; He, Y.; Luo, B.; He, C.; Liang, Y.; Zeng, J.; Li, W.; Xian, Y.; Zheng, X. Identification of key protein-coding genes and lncRNAs in spontaneous neutrophil apoptosis. *Sci. Rep.* **2019**, *9*, 15106. [[CrossRef](#)]
28. Cheng, J.; Ma, H.; Yan, M.; Xing, W. THAP9-AS1/miR-133b/SOX4 positive feedback loop facilitates the progression of esophageal squamous cell carcinoma. *Cell Death Dis.* **2021**, *12*, 401. [[CrossRef](#)]
29. Sharma, V.; Thakore, P.; Krishnan, M.; Majumdar, S. Stress induced Differential Expression of THAP9 & THAP9-AS1 in the S-phase of cell cycle. *bioRxiv* **2021**. [[CrossRef](#)]
30. Tomczak, K.; Czerwińska, P.; Wiznerowicz, M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol.* **2015**, *19*, A68–A77. [[CrossRef](#)]
31. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **2013**, *45*, 580–585. [[CrossRef](#)] [[PubMed](#)]
32. Li, T.; Fu, J.; Zeng, Z.; Cohen, D.; Li, J.; Chen, Q.; Li, B.; Liu, X.S. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **2020**, *48*, W509–W514. [[CrossRef](#)] [[PubMed](#)]
33. Tang, Z.; Kang, B.; Li, C.; Chen, T.; Zhang, Z. GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* **2019**, *47*, W556–W560. [[CrossRef](#)] [[PubMed](#)]
34. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* **2012**, *2*, 401–404. [[CrossRef](#)]
35. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [[CrossRef](#)]
36. McKenzie, A.T.; Katsyv, I.; Song, W.-M.; Wang, M.; Zhang, B. DGCA: A comprehensive R package for Differential Gene Correlation Analysis. *BMC Syst. Biol.* **2016**, *10*, 106. [[CrossRef](#)]
37. Howe, K.L.; Achuthan, P.; Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Azov, A.G.; Bennett, R.; Bhai, J.; et al. Ensembl 2021. *Nucleic Acids Res.* **2021**, *49*, D884–D891. [[CrossRef](#)]
38. Dreos, R.; Ambrosini, G.; Périer, R.C.; Bucher, P. The Eukaryotic Promoter Database: Expansion of EPDnew and new promoter analysis tools. *Nucleic Acids Res.* **2015**, *43*, D92–D96. [[CrossRef](#)]
39. Antequera, F. Structure, function and evolution of CpG island promoters, CMLS. *Cell. Mol. Life Sci.* **2003**, *60*, 1647–1658. [[CrossRef](#)]
40. Gardiner-Garden, M.; Frommer, M. CpG islands in vertebrate genomes. *J. Mol. Biol.* **1987**, *196*, 261–282. [[CrossRef](#)]
41. Davis, C.A.; Hitz, B.C.; Sloan, C.A.; Chan, E.T.; Davidson, J.M.; Gabdank, I.; Hilton, J.A.; Jain, K.; Baymuradov, U.K.; Narayanan, A.K.; et al. The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res.* **2018**, *46*, D794–D801. [[CrossRef](#)] [[PubMed](#)]
42. Bornelöv, S.; Komorowski, J.; Wadelius, C. Different distribution of histone modifications in genes with unidirectional and bidirectional transcription and a role of CTCF and cohesin in directing transcription. *BMC Genom.* **2015**, *16*, 300. [[CrossRef](#)] [[PubMed](#)]
43. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant.* **2020**, *13*, 1194–1202. [[CrossRef](#)] [[PubMed](#)]
44. Yang, M.Q.; Elnitski, L.L. Diversity of core promoter elements comprising human bidirectional promoters. *BMC Genom.* **2008**, *9*, S3. [[CrossRef](#)] [[PubMed](#)]

45. Sloutskin, A.; Danino, Y.M.; Orenstein, Y.; Zehavi, Y.; Doniger, T.; Shamir, R.; Juven-Gershon, T. ElemeNT: A computational tool for detecting core promoter elements. *Transcription* **2015**, *6*, 41–50. [[CrossRef](#)]
46. Yang, C.; Bolotin, E.; Jiang, T.; Sladek, F.M.; Martinez, E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* **2007**, *389*, 52–65. [[CrossRef](#)]
47. Smale, S.T.; Kadonaga, J.T. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* **2003**, *72*, 449–479. [[CrossRef](#)]
48. Kutach, A.K.; Kadonaga, J.T. The Downstream Promoter Element DPE Appears to Be as Widely Used as the TATA Box in *Drosophila* Core Promoters. *Mol. Cell. Biol.* **2000**, *20*, 4754–4764. [[CrossRef](#)]
49. Lagrange, T.; Kapanidis, A.N.; Tang, H.; Reinberg, D.; Ebricht, R.H. New core promoter element in RNA polymerase II-dependent transcription: Sequence-specific DNA binding by transcription factor IIB. *Genes Dev.* **1998**, *12*, 34–44. [[CrossRef](#)]
50. Lachance, J. Disease-associated alleles in genome-wide association studies are enriched for derived low frequency alleles relative to HapMap and neutral expectations. *BMC Med. Genom.* **2010**, *3*, 57. [[CrossRef](#)]
51. Andiappan, A.K.; Wang, D.Y.; Anantharaman, R.; Parate, P.N.; Suri, B.K.; Low, H.Q.; Li, Y.; Zhao, W.; Castagnoli, P.; Liu, J.; et al. Genome-Wide Association Study for Atopy and Allergic Rhinitis in a Singapore Chinese Population. *PLoS ONE* **2011**, *6*, e19719. [[CrossRef](#)] [[PubMed](#)]
52. Guarguaglini, G.; Battistoni, A.; Pittoggi, C.; di Matteo, G.; di Fiore, B.; Lavia, P. Expression of the murine RanBP1 and Htf9-c genes is regulated from a shared bidirectional promoter during cell cycle progression. *Biochem. J.* **1997**, *325 Pt 1*, 277–286. [[CrossRef](#)] [[PubMed](#)]
53. Oliver, S. Guilt-by-association goes global. *Nature* **2000**, *403*, 601–602. [[CrossRef](#)]
54. Van Dam, S.; Vösa, U.; van der Graaf, A.; Franke, L.; de Magalhães, J.P. Gene co-expression analysis for functional classification and gene–disease predictions. *Brief. Bioinform.* **2017**, *19*, 575–592. [[CrossRef](#)] [[PubMed](#)]
55. Bijmens, A.P.J.J.; Gils, A.; Jutten, B.; Faber, B.C.G.; Heeneman, S.; Kitslaar, P.J.E.H.M.; Tordoir, J.H.M.; de Vries, C.J.M.; Kroon, A.A.; Daemen, M.J.A.P.; et al. Vasculin, a novel vascular protein differentially expressed in human atherogenesis. *Blood* **2003**, *102*, 2803–2810. [[CrossRef](#)] [[PubMed](#)]
56. Ong, W.-Y.; Ng, M.P.-E.; Loke, S.-Y.; Jin, S.; Wu, Y.-J.; Tanaka, K.; Wong, P.T.-H. Comprehensive Gene Expression Profiling Reveals Synergistic Functional Networks in Cerebral Vessels after Hypertension or Hypercholesterolemia. *PLoS ONE* **2013**, *8*, e68335. [[CrossRef](#)]
57. Mathys, H.; Davila-Velderrain, J.; Peng, Z.; Gao, F.; Mohammadi, S.; Young, J.Z.; Menon, M.; He, L.; Abdurrob, F.; Jiang, X.; et al. Single-cell transcriptomic analysis of Alzheimer’s disease. *Nature* **2019**, *570*, 332–337. [[CrossRef](#)]
58. Lee, T.; Lee, H. Identification of Disease-Related Genes That Are Common between Alzheimer’s and Cardiovascular Disease Using Blood Genome-Wide Transcriptome Analysis. *Biomedicines* **2021**, *9*, 1525. [[CrossRef](#)]
59. Cho, H.; Chung, J.-Y.; Song, K.-H.; Noh, K.H.; Kim, B.W.; Chung, E.J.; Ylala, K.; Kim, J.H.; Kim, T.W.; Hewitt, S.M.; et al. Apoptosis inhibitor-5 overexpression is associated with tumor progression and poor prognosis in patients with cervical cancer. *BMC Cancer* **2014**, *14*, 545. [[CrossRef](#)]
60. Krejci, P.; Pejchalova, K.; Rosenbloom, B.E.; Rosenfelt, F.P.; Tran, E.L.; Laurell, H.; Wilcox, W.R. The antiapoptotic protein Api5 and its partner, high molecular weight FGF2, are up-regulated in B cell chronic lymphoid leukemia. *J. Leukoc. Biol.* **2007**, *82*, 1363–1364. [[CrossRef](#)]
61. Mao, C.-P.; Wu, T.; Song, K.-H.; Kim, T.W. Immune-mediated tumor evolution: Nanog links the emergence of a stem like cancer cell state and immune evasion. *Oncimmunology* **2014**, *3*, e947871. [[CrossRef](#)] [[PubMed](#)]
62. Song, K.-H.; Cho, H.; Kim, S.; Lee, H.-J.; Oh, S.J.; Woo, S.R.; Hong, S.-O.; Jang, H.S.; Noh, K.H.; Choi, C.H.; et al. API5 confers cancer stem cell-like properties through the FGF2-NANOG axis. *Oncogenesis* **2017**, *6*, e285. [[CrossRef](#)] [[PubMed](#)]
63. Zhu, W.; Swaminathan, G.; Plowey, E.D. GA binding protein augments autophagy via transcriptional activation of BECN1-PIK3C3 complex genes. *Autophagy* **2014**, *10*, 1622–1636. [[CrossRef](#)]
64. Morris, S.; Geoghegan, N.D.; Sadler, J.B.A.; Koester, A.M.; Black, H.L.; Laub, M.; Miller, L.; Heffernan, L.; Simpson, J.C.; Mastick, C.C.; et al. Characterisation of GLUT4 trafficking in HeLa cells: Comparable kinetics and orthologous trafficking mechanisms to 3T3-L1 adipocytes. *PeerJ* **2020**, *8*, e8751. [[CrossRef](#)] [[PubMed](#)]
65. Kagaya, A.; Shimada, H.; Shiratori, T.; Kuboshima, M.; Nakashima-Fujita, K.; Yasuraoka, M.; Nishimori, T.; Kurei, S.; Hachiya, T.; Murakami, A.; et al. Identification of a novel SEREX antigen family, ECSA, in esophageal squamous cell carcinoma. *Proteome Sci.* **2011**, *9*, 31. [[CrossRef](#)]
66. Abovich, N.; Rosbash, M. Cross-Intron Bridging Interactions in the Yeast Commitment Complex Are Conserved in Mammals. *Cell* **1997**, *89*, 403–412. [[CrossRef](#)]
67. Kao, H.Y.; Siliciano, P.G. Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. *Mol. Cell Biol.* **1996**, *16*, 960–967. [[CrossRef](#)]
68. Buschdorf, J.P.; Strätling, W.H. A WW domain binding region in methyl-CpG-binding protein MeCP2: Impact on Rett syndrome. *J. Mol. Med.* **2004**, *82*, 135–143. [[CrossRef](#)]
69. Faber, P.W.; Barnes, G.T.; Srinidhi, J.; Chen, J.; Gusella, J.F.; MacDonald, M.E. Huntingtin Interacts with a Family of WW Domain Proteins. *Hum. Mol. Genet.* **1998**, *7*, 1463–1474. [[CrossRef](#)]
70. Huo, Z.; Zhai, S.; Weng, Y.; Qian, H.; Tang, X.; Shi, Y.; Deng, X.; Wang, Y.; Shen, B. PRPF40A as a potential diagnostic and prognostic marker is upregulated in pancreatic cancer tissues and cell lines: An integrated bioinformatics data analysis. *Oncotargets Ther.* **2019**, *12*, 5037–5051. [[CrossRef](#)]

71. Oleksiewicz, U.; Liloglou, T.; Tasopoulou, K.-M.; Daskoulidou, N.; Gosney, J.R.; Field, J.K.; Xinarianos, G. COL1A1, PRPF40A, and UCP2 correlate with hypoxia markers in non-small cell lung cancer. *J. Cancer Res. Clin. Oncol.* **2017**, *143*, 1133–1141. [[CrossRef](#)] [[PubMed](#)]
72. Wang, J.; Xie, G.; Singh, M.; Ghanbarian, A.T.; Raskó, T.; Szvetnik, A.; Cai, H.; Besser, D.; Prigione, A.; Fuchs, N.V.; et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **2014**, *516*, 405–409. [[CrossRef](#)] [[PubMed](#)]
73. Nayler, O.; Hartmann, A.M.; Stamm, S. The ER Repeat Protein Yt521-B Localizes to a Novel Subnuclear Compartment. *J. Cell Biol.* **2000**, *150*, 949–962. [[CrossRef](#)] [[PubMed](#)]
74. Hirschfeld, M.; Zhang, B.; Jaeger, M.; Stamm, S.; Erbes, T.; Mayer, S.; Tong, X.; Stickeler, E. Hypoxia-dependent mRNA expression pattern of splicing factor YT521 and its impact on oncological important target gene expression. *Mol. Carcinog.* **2014**, *53*, 883–892. [[CrossRef](#)]
75. Luxton, H.J.; Simpson, B.S.; Mills, I.G.; Brindle, N.R.; Ahmed, Z.; Stavrinides, V.; Heavey, S.; Stamm, S.; Whitaker, H.C. The Oncogene Metadherin Interacts with the Known Splicing Proteins YTHDC1, Sam68 and T-STAR and Plays a Novel Role in Alternative mRNA Splicing. *Cancers* **2019**, *11*, 1233. [[CrossRef](#)]
76. Yoo, B.K.; Emdad, L.; Su, Z.; Villanueva, A.; Chiang, D.Y.; Mukhopadhyay, N.D.; Mills, A.S.; Waxman, S.; Fisher, R.A.; Llovet, J.M.; et al. Astrocyte elevated gene-1 regulates hepatocellular carcinoma development and progression. *J. Clin. Investig.* **2009**, *119*, 465–477. [[CrossRef](#)]
77. Zhang, B.; Shao, X.; Zhou, J.; Qiu, J.; Wu, Y.; Cheng, J. YT521 promotes metastases of endometrial cancer by differential splicing of vascular endothelial growth factor A. *Tumor Biol.* **2016**, *37*, 15543–15549. [[CrossRef](#)]
78. Feng, Y.; Chen, M.; Manley, J.L. Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. *Nat. Struct. Mol. Biol.* **2008**, *15*, 1040–1048. [[CrossRef](#)]
79. Zhou, X.; Li, X.; Cheng, Y.; Wu, W.; Xie, Z.; Xi, Q.; Han, J.; Wu, G.; Fang, J.; Feng, Y. BCLAF1 and its splicing regulator SRSF10 regulate the tumorigenic potential of colon cancer cells. *Nat. Commun.* **2014**, *5*, 4581. [[CrossRef](#)]
80. Kasof, G.M.; Goyal, L.; White, E. Btf, a Novel Death-Promoting Transcriptional Repressor That Interacts with Bcl-2-Related Proteins. *Mol. Cell. Biol.* **1999**, *19*, 4390–4404. [[CrossRef](#)]
81. McPherson, J.P.; Sarras, H.; Lemmers, B.; Tamblyn, L.; Migon, E.; Matysiak-Zablocki, E.; Hakem, A.; Azami, S.A.; Cardoso, R.; Fish, J.; et al. Essential role for Bclaf1 in lung development and immune system function. *Cell Death Differ.* **2009**, *16*, 331–339. [[CrossRef](#)] [[PubMed](#)]
82. Savage, K.I.; Gorski, J.J.; Barros, E.M.; Irwin, G.W.; Manti, L.; Powell, A.J.; Pellagatti, A.; Lukashchuk, N.; McCance, D.J.; McCluggage, W.G.; et al. Identification of a BRCA1-mRNA Splicing Complex Required for Efficient DNA Repair and Maintenance of Genomic Stability. *Mol. Cell.* **2014**, *54*, 445–459. [[CrossRef](#)] [[PubMed](#)]
83. Shao, A.-W.; Sun, H.; Geng, Y.; Peng, Q.; Wang, P.; Chen, J.; Xiong, T.; Cao, R.; Tang, J. Bclaf1 is an important NF- κ B signaling transducer and C/EBP β regulator in DNA damage-induced senescence. *Cell Death Differ.* **2016**, *23*, 865–875. [[CrossRef](#)] [[PubMed](#)]
84. Varia, S.; Potabathula, D.; Deng, Z.; Bubulya, A.; Bubulya, P.A. Btf and TRAP150 have distinct roles in regulating subcellular mRNA distribution. *Nucleus* **2013**, *4*, 229–240. [[CrossRef](#)] [[PubMed](#)]
85. Aiello, C.; Terracciano, A.; Simonati, A.; Discepoli, G.; Cannelli, N.; Claps, D.; Crow, Y.J.; Bianchi, M.; Kitzmuller, C.; Longo, D.; et al. Mutations in MFSD8/CLN7 are a frequent cause of variant-late infantile neuronal ceroid lipofuscinosis. *Hum. Mutat.* **2009**, *30*, E530–E540. [[CrossRef](#)] [[PubMed](#)]
86. Aldahmesh, M.A.; Al-Hassnan, Z.N.; Aldosari, M.; Alkuraya, F.S. Neuronal ceroid lipofuscinosis caused by MFSD8 mutations: A common theme emerging. *Neurogenetics* **2009**, *10*, 307–311. [[CrossRef](#)]
87. Kousi, M.; Siintola, E.; Dvorakova, L.; Vlaskova, H.; Turnbull, J.; Topcu, M.; Yuksel, D.; Gokben, S.; Minassian, B.A.; Elleder, M.; et al. Mutations in CLN7/MFSD8 are a common cause of variant late-infantile neuronal ceroid lipofuscinosis. *Brain* **2009**, *132*, 810–819. [[CrossRef](#)]
88. Siintola, E.; Topcu, M.; Aula, N.; Lohi, H.; Minassian, B.A.; Paterson, A.D.; Liu, X.-Q.; Wilson, C.; Lahtinen, U.; Anttonen, A.-K.; et al. The Novel Neuronal Ceroid Lipofuscinosis Gene MFSD8 Encodes a Putative Lysosomal Transporter. *Am. J. Hum. Genet.* **2007**, *81*, 136–146. [[CrossRef](#)]
89. Stogmann, E.; El Tawil, S.; Wagenstaller, J.; Gaber, A.; Edris, S.; Abdelhady, A.; Assem-Hilger, E.; Leutmezer, F.; Bonelli, S.; Baumgartner, C.; et al. A novel mutation in the MFSD8 gene in late infantile neuronal ceroid lipofuscinosis. *Neurogenetics* **2009**, *10*, 73–77. [[CrossRef](#)]
90. Clifford, R.J.; Zhang, J.; Meerzaman, D.M.; Lyu, M.-S.; Hu, Y.; Cultraro, C.M.; Finney, R.P.; Kelley, J.M.; Efroni, S.; Greenblum, S.I.; et al. Genetic Variations at Loci Involved in the Immune Response Are Risk Factors for Hepatocellular Carcinoma. *Hepatology* **2010**, *52*, 2034–2043. [[CrossRef](#)]
91. Ge, S.X.; Jung, D.; Yao, R. ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **2020**, *36*, 2628–2629. [[CrossRef](#)] [[PubMed](#)]
92. Sanghavi, H.M.; Majumdar, S. Oligomerization of THAP9 Transposase via Amino-Terminal Domains. *Biochemistry* **2021**, *60*, 1822–1835. [[CrossRef](#)] [[PubMed](#)]
93. Amar, D.; Safer, H.; Shamir, R. Dissection of regulatory networks that are altered in disease via differential co-expression. *PLoS Comput. Biol.* **2013**, *9*, e1002955. [[CrossRef](#)] [[PubMed](#)]

94. Hudson, N.J.; Reverter, A.; Dalrymple, B.P. A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS Comput. Biol.* **2009**, *5*, e1000382. [[CrossRef](#)]
95. Kostka, D.; Spang, R. Finding disease specific alterations in the co-expression of genes. *Bioinformatics* **2004**, *1* (Suppl. 20), i194–i199. [[CrossRef](#)]
96. Gao, Q.; Ho, C.; Jia, Y.; Li, J.J.; Huang, H. Biclustering of linear patterns in gene expression data. *J. Comput. Biol.* **2012**, *19*, 619–631. [[CrossRef](#)] [[PubMed](#)]
97. Monaco, G.; van Dam, S.; Ribeiro, J.L.C.N.; Larbi, A.; de Magalhães, J.P. A comparison of human and mouse gene co-expression networks reveals conservation and divergence at the tissue, pathway and disease levels. *BMC Evol. Biol.* **2015**, *15*, 259. [[CrossRef](#)]
98. Pierson, E.; Consortium, G.T.E.; Koller, D.; Battle, A.; Mostafavi, S.; Ardlie, K.G.; Getz, G.; Wright, F.A.; Kellis, M.; Volpi, S.; et al. Sharing and Specificity of Co-expression Networks across 35 Human Tissues. *PLoS Comput. Biol.* **2015**, *11*, e1004220. [[CrossRef](#)] [[PubMed](#)]
99. Zeisel, A.; Muñoz-Manchado, A.B.; Codeluppi, S.; Lönnerberg, P.; la Manno, G.; Juréus, A.; Marques, S.; Munguba, H.; He, L.; Betsholtz, C.; et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **2015**, *347*, 1138–1142. [[CrossRef](#)]
100. Chen, Y.; Li, H.; Li, Y.-Y.; Li, Y. Pan-Cancer Analysis of Head-to-Head Gene Pairs in Terms of Transcriptional Activity, Co-expression and Regulation. *Front. Genet.* **2021**, *11*, 1707. [[CrossRef](#)]
101. Shu, J.; Jelinek, J.; Chang, H.; Shen, L.; Qin, T.; Chung, W.; Oki, Y.; Issa, J.-P.J. Silencing of Bidirectional Promoters by DNA Methylation in Tumorigenesis. *Cancer Res.* **2006**, *66*, 5077–5084. [[CrossRef](#)] [[PubMed](#)]
102. Yang, M.Q.; Koehly, L.M.; Elnitski, L.L. Comprehensive Annotation of Bidirectional Promoters Identifies Co-Regulation among Breast and Ovarian Cancer Genes. *PLoS Comput. Biol.* **2007**, *3*, e72. [[CrossRef](#)] [[PubMed](#)]
103. Albig, W.; Kioschis, P.; Poustka, A.; Meergans, K.; Doenecke, D. Human histone gene organization: Nonregular arrangement within a large cluster. *Genomics* **1997**, *40*, 314–322. [[CrossRef](#)]
104. Orekhova, A.S.; Rubtsov, P.M. Bidirectional Promoters in the Transcription of Mammalian Genomes. *Biochemistry* **2013**, *78*, 7. [[CrossRef](#)]
105. Dovhey, S.E.; Ghosh, N.S.; Wright, K.L. Loss of interferon-gamma inducibility of TAP1 and LMP2 in a renal cell carcinoma cell line. *Cancer Res.* **2000**, *60*, 5789–5796.
106. Rashmi, R.; Nandi, C.; Majumdar, S. Evolutionary analysis of THAP9 transposase: Conserved regions, novel motifs. *bioRxiv* **2021**. [[CrossRef](#)]
107. Parker, J.B.; Palchaudhuri, S.; Yin, H.; Wei, J.; Chakravarti, D. A transcriptional regulatory role of the THAP11-HCF-1 complex in colon cancer cell function. *Mol. Cell Biol.* **2012**, *32*, 1654–1670. [[CrossRef](#)]
108. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Studying Gene Expression and Function, Molecular Biology of the Cell*, 4th ed. Garland Science: New York, NY, USA, 2002. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK26818/> (accessed on 12 January 2022).
109. Marcocci, M.E.; Napoletani, G.; Protto, V.; Kolesova, O.; Piacentini, R.; Puma, D.D.L.; Lomonte, P.; Grassi, C.; Palamara, A.T.; de Chiara, G. Herpes Simplex Virus-1 in the Brain: The Dark Side of a Sneaky Infection. *Trends Microbiol.* **2020**, *28*, 808–820. [[CrossRef](#)] [[PubMed](#)]
110. Kumar, G.S.S.; Venugopal, A.K.; Kashyap, M.K.; Raju, R.; Marimuthu, A.; Palapetta, S.M.; Subbanayya, Y.; Goel, R.; Chawla, A.; Dikshit, J.B.; et al. Gene Expression Profiling of Tuberculous Meningitis Co-infected with HIV. *J. Proteom. Bioinform.* **2012**, *5*, 235–244. [[CrossRef](#)] [[PubMed](#)]
111. Bioconductor. Available online: <http://bioconductor.org/packages/org.Hs.eg.db/> (accessed on 12 January 2022).
112. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
113. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)] [[PubMed](#)]
114. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)] [[PubMed](#)]