

Received: 2020.11.08

Accepted: 2021.03.10

Available online: 2021.03.29

Published: 2021.07.05

Identification and Characterization of Non-Coding RNAs in Thymoma

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

AE 1 **Guanglei Ji**
BCDF 2 **Rongrong Ren**
BCDF 3 **Xichao Fang**

1 First Department of Thoracic Surgery, Linyi Cancer Hospital, Linyi, Shandong, PR China
2 Department of Anesthesiology and Surgical Intensive Care Unit, Xinhua Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, PR China
3 Second Department of Thoracic Surgery, Linyi Cancer Hospital, Linyi, Shandong, PR China

Corresponding Author: Xichao Fang, e-mail: xcfang1101@126.com

Source of support: Departmental sources

Background: Thymoma is the most common tumor of the anterior mediastinum, and can be caused by infrequent malignancies arising from the epithelial cells of the thymus. Unfortunately, blood-based diagnostic markers are not currently available. High-throughput sequencing technologies, such as RNA-seq with next-generation sequencing, have facilitated the detection and characterization of both coding and non-coding RNAs (ncRNAs), which play significant roles in genomic regulation, transcriptional and post-transcriptional regulation, and imprinting and epigenetic modification. The knowledge about fusion genes and ncRNAs in thymomas is scarce.

Material/Methods: For this study, we gathered large-scale RNA-seq data belonging to samples from 25 thymomas and 25 healthy thymus specimens and analyzed them to identify fusion genes, lncRNAs, and miRNAs.

Results: We found 21 fusion genes, including KMT2A-MAML2, HADHB-REEP1, COQ3-CGA, MCM4-SNTB1, and IFT140-ACTN4, as the most frequent and significant in thymomas. We also detected 65 differentially-expressed lncRNAs in thymomas, including AFAP1-AS1, LINC00324, ADAMTS9-AS1, VLDLR-AS1, LINC00968, and NEAT1, that have been validated with the TCGA database. Moreover, we identified 1695 miRNAs from small RNA-seq data that were overexpressed in thymomas. Our network analysis of the lncRNA-mRNA-miRNA regulation axes identified a cluster of miRNAs upregulated in thymomas, that can trigger the expression of target protein-coding genes, and lead to the disruption of several biological pathways, including the PI3K-Akt signaling pathway, FoxO signaling pathway, and HIF-1 signaling pathway.

Conclusions: Our results show that overexpression of this miRNA cluster activates PI3K-Akt, FoxO, HIF-1, and Rap-1 signaling pathways, suggesting pathway inhibitors may be therapeutic candidates against thymoma.

Keywords: **MicroRNAs • RNA, Long Noncoding • Sequence Analysis, RNA • Thymus Neoplasms**

Full-text PDF: <https://www.medscimonit.com/abstract/index/idArt/929727>

 4478

 5

 7

 82



Background

The thymus, a small organ located behind the breast bone and just in front of and above the heart, is the primary site of T cell development and plays an essential role during the development of the immune system, including T cell differentiation and maturation [1]. Common signs and symptoms of tumors in the thymus include shortness of breath, cough with or without hemoptysis, chest pain, trouble swallowing, loss of appetite, and weight loss. Some thymoma diagnostic methods are chest X-rays, CT scans, MRIs, PET scans, biopsy, and the Chamberlain procedure [2]. Blood-based diagnostic markers are not currently available, but they would help to detect myasthenia gravis (MG) or other associated autoimmune disorders. The standard treatments available against thymoma and thymic carcinoma are surgery, radiation therapy, chemotherapy, hormone therapy, targeted therapy, and immunotherapy (under clinical trial) [3,4].

Thymic epithelial tumors (TETs) are infrequent malignancies arising from the epithelial cells of the thymus of middle-age patients; thymoma is the most predominant and clinically aggressive TET [5]. TETs are the most common tumors in the anterior mediastinum, but are infrequent in comparison to other thoracic tumors [6,7]. According to the American Society of Clinical Oncology (ASCO), thymomas account for 20% of all anterior mediastinal tumors, with an overall incident rate lower than 1 per 1.5 million, and a 5-year survival rate of 71% [50]. However, the survival rate differs depending on various factors such as cancer classification and stage. Thymomas are characterized by their unique association with autoimmune diseases and a thymic carcinoma [5]. According to the WHO, thymomas are histologically classified into 6 types: A, AB, B1, B2, B3, and C. Types A and AB are considered benign, with low mortality and rate of growth, whereas types B1, B2, and B3 are more aggressive and display a high propensity for spreading intrathoracically [8].

Thymomas have been associated with several paraneoplastic conditions, and a variety of immune and non-immune and endocrine disorders [9]. The most common syndromes associated with thymoma are myasthenia gravis (MG) (in ~30% of patients), pure red cell aplasia (PRCA) (in 5-10% of patients), and hypogammaglobulinemia (in 3-6% of patients). MG is characterized by development of autoimmune antibodies to the acetylcholine receptor in the postsynaptic neuromuscular junction [10], PRCA occurs due to immune-mediated suppression of erythropoiesis [11], and hypogammaglobulinemia occurs due to a collective deficiency of B and T cells that lead to an increased risk of infection [12]. In addition, thymomas are associated with other paraneoplastic conditions like polymyositis, systemic lupus erythematosus, limbic encephalitis, sensory-motor neuropathy, rheumatoid arthritis, thyroiditis,

ulcerative colitis, acute pericarditis, myocarditis, hemolytic anemia, Addison's disease, and Cushing's syndrome [10]. The thymus plays an important role during development and is very sensitive to both acute and chronic injury. The thymus goes through rapid degeneration as a result of toxic insults and due to aging at a faster rate than other tissues. However, it is capable of regenerating and restoring its function to a certain degree by a mechanism involving keratinocyte growth factor (KGF) signaling [1].

Recent advances in high-throughput technologies, such as next-generation sequencing (NGS), have revealed that a significant portion of the genome (~75%) is transcribed to non-coding RNAs (ncRNAs). Among the ncRNAs, the long ncRNAs (lncRNAs) are usually at least 200 nucleotides long, while small RNAs (such as miRNAs and siRNAs) are <200 nucleotides long. According to the ENCODE (ENCyclopedia of DNA Elements) Project Consortium (GENCODE release 34), the human genome contains around 19 959 protein-coding genes, 17 960 lncRNAs genes, and 7578 small ncRNAs [13]. lncRNAs are generally mRNA-like transcripts that have conserved secondary structures and lack long open reading frames (ORFs) and have low sequence conservation. Hence, they are difficult to detect computationally from genome sequences [14]. The roles of lncRNAs in genomic regulation are multiple, involving transcriptional and post-transcriptional regulation, and imprinting and epigenetic modifications. Studies have documented that lncRNAs play key roles during tumorigenesis and tumor progression [15-17]; in particular, aberrant lncRNA expression or lncRNA dysfunction may lead to DNA damage, immune escape, and cellular metabolic disorders in cancer cells. lncRNAs can regulate cell proliferation, apoptosis, migration, invasion, and maintenance of stemness during the development of cancer. Moreover, many lncRNAs are transcriptionally regulated by key tumor suppressors or oncogenes [15]. The inherent heterogeneity and diversity of the lncRNA world make the complex tumorigenesis process more interesting than it would otherwise be [17]. Hence, identification and characterization of lncRNAs could open the doors to new clinical approaches to understand complex tumorigenesis and treat cancers.

Gene fusions occur due to recombination of 2 unrelated genes through chromosomal rearrangements or translocations. Gene fusions are key drivers of human cancer development, and these appear to play significant roles in the aggressive behavior of cancers [18].

In this study, we attempted to identify and characterize fusion genes, lncRNAs, and miRNAs from RNA-seq data of patients with thymoma, identifying aberrant lncRNAs and miRNAs in thymoma, constructing a lncRNA-mRNA-miRNA regulatory network in thymoma, and performing a functional enrichment analysis.

Table 1. Basic characteristics of considered small RNA-seq data.

Run ID	Average spot length	Bases (Mbase)	Data size (MB)	Data file format	Sample type	WHO sub-type
SRR3341111	46	705.93	292.9	bam,sra	Thymic neoplasm	A
SRR3341114	45	497.06	230.71	bam,sra	Thymic neoplasm	A
SRR3341117	44	472.92	231.16	bam,sra	Thymic neoplasm	A
SRR3341120	43	380.61	191.63	bam,sra	Thymic neoplasm	A/B
SRR3341123	44	438.40	207.93	bam,sra	Thymic neoplasm	B3
SRR3341126	44	368.09	177.25	bam,sra	Thymic neoplasm	B3
SRR3341129	46	696.30	309.79	bam,sra	Thymic neoplasm	A
SRR3341132	46	661.74	279.82	bam,sra	Thymicneoplasm	B3
SRR3341135	46	626.43	266.13	bam,sra	Thymic neoplasm	A/B
SRR3341144	41	246.58	143.25	bam,sra	Thymic neoplasm	C
SRR3341146	79	3156.80	1669.12	bam,sra	Thymic neoplasm	C
SRR3341150	46	776.60	349.51	bam,sra	Normal thymus	–
SRR3341151	46	689.80	315.14	bam,sra	Normal thymus	–
SRR3341152	80	2253.16	1285.14	bam,sra	Normal thymus	–
SRR3341153	45	700.13	334.67	bam,sra	Normal thymus	–
SRR3341154	44	630.27	307.39	bam,sra	Normal thymus	–
SRR3341155	79	2352.09	1394.3	bam,sra	Normal thymus	–
SRR3341156	45	711.40	337.3	bam,sra	Normal thymus	–
SRR3341157	44	636.42	307.0	bam,sra	Normal thymus	–
SRR3341158	78	2289.89	1370.73	bam,sra	Normal thymus	–

Material and Methods

Data Collection and Read Quality Check

We used 25 samples of Illumina paired-end RNA-Seq data of cancerous thymomas for this study; the same is available on the NCBI-SRA website (<https://www.ncbi.nlm.nih.gov/sra>) with a SRA run series from SRR1296011 to SRR1296035. We also used healthy thymus RNA-Seq data available at a bearing run series, SRR4175275, SRR4175279, SRR4175280, SRR6753058 to SRR6753061, SRR6753146 to SRR6753153, and SRR6753211 to SRR6753220 (25 samples). Moreover, we downloaded small RNA-Seq datasets of patients with thymic malignancy and normal thymus tissues from the BioProject Accession PRJNA317556 website (<https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA317556>). We collected data from 11 thymic malignancy samples belonging to different cancer stages and 9 normal thymus gland specimens (Table 1). The raw sequence reads output from the sequencer were subjected to

quality checks. We checked the quality of the RNA-Seq data in FASTQ format using the FastQC software tool (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) [19]. The quality check results showed that all the sequence reads belonging to various sample types were of high quality (Q-score >20) and did not contain adapters.

We developed a methodological pipeline to process RNA-Seq data for the detection of fusion genes, lncRNAs, and small ncRNAs (miRNAs). Figure 1 depicts the downstream analysis along with the set of tools used. The detailed methodology is described below.

Fusion Gene Detection

Fusion genes, formed by chromosomal breakage and rejoining events, are the most common mutation class in cancers. They result from chromosomal rearrangements including translocation, inversions, deletions, and duplications [20-22]. Fusion genes lead

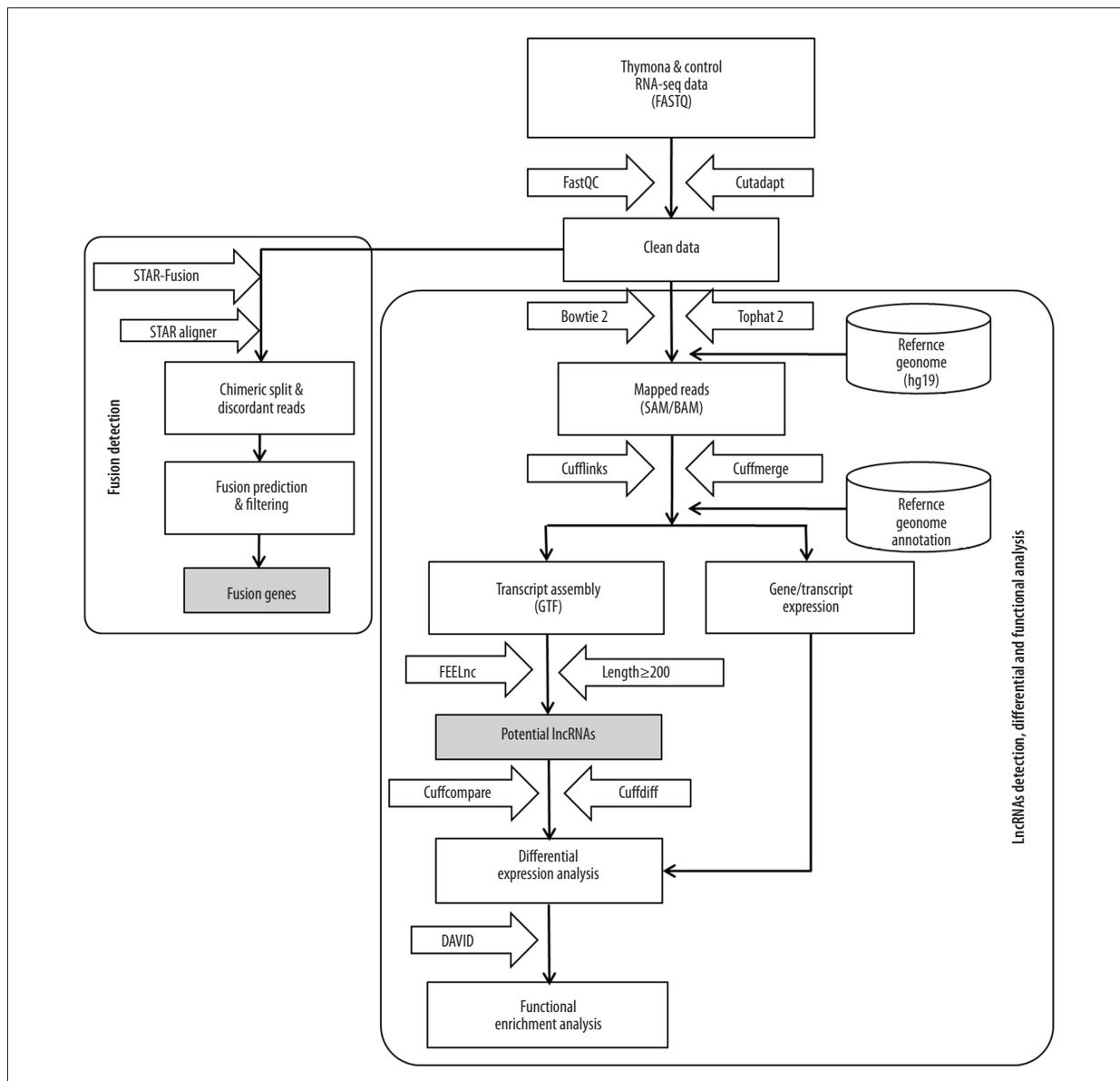


Figure 1. Methodological pipeline to process RNA-Seq data and analysis used in this study.

to chimeric transcripts that deregulate genes through juxtaposition of novel promoter or enhancer regions [21]. Common methods for fusion gene detection are fluorescence in situ hybridization (FISH), RT-PCR, and next-generation sequencing (NGS). The FISH and RT-PCR methods are used for a single gene fusion and cannot detect fusion partners or complex structural rearrangements [22]. Moreover, these 2 techniques are poorly sensitive to intra-chromosomal fusion gene detection [23,24]. The third technique (NGS) is now a popularly used method for fusion detection at the whole-genome and transcriptome scale; it employs RNA-Seq data from NGS [25]. RNA-seq are mostly used because they only require a sequence region (targeted region) of the genome to be transcribed and split into mature mRNAs, making it quick and cost-effective, while it can detect multiple fusion genes.

A plethora of fusion detection tools has been developed during the last decade, including deFuse [26], EricScript [27], JAFFA [28], STAR-Fusion [29], and SQUID [30] tools. These tools are conceptually classified into 2 groups [31]: i) mapping-first approaches, in which RNA-seq reads are first aligned to the reference genome to detect discordant mapping reads suggestive of rearrangements, and ii) assembly-first approaches, in which assembled reads are followed by identification of chimeric transcripts. In this study, we used STAR-Fusion release v1.8.0 available at <https://github.com/STAR-Fusion/STAR-Fusion/> to investigate therapeutic implications using RNA-seq data to detect fusion genes in both thymoma and healthy thymus datasets. We selected STAR-Fusion due to its outstanding performance published in one of the benchmarking studies among

23 fusion detection tools for cancer transcriptomes [31]. STAR-Fusion belongs to the Trinity Cancer Transcriptome Analysis Toolkit (CTAT) project, which takes Illumina RNA-seq reads as input and generates a list of candidate fusion genes as output. STAR-Fusion, a mapping-first approach, uses the STAR Aligner [32] that generates all the chimeric splits and discordant reads used as the input for the STAR-Fusion. STAR-Fusion maps the reads to exons of reference gene annotations and co-ordinates overlaps (Figure 1).

Read Mapping and Transcript Assembly

The clean RNA-seq data of both 25 patients with thymoma and 25 healthy individuals were first mapped to the *Homo sapiens* hg19 reference genome (downloaded from UCSC Genome Browser <http://hgdownload.soe.ucsc.edu/downloads.html#human>) using a recently updated version 2.4.0 of the Bowtie2 [33] and Tophat 2 version 2.1.1 [34] read mappers (Figure 1). The results of the read mapping were obtained in BAM/SAM format, with 1 file for each sample. To obtain a transcript assembly, we used Cufflinks, which is a widely used tool [35]. The Cufflinks tool assembles individual transcripts aligned to the genome, estimates abundance, and quantifies the expression level of genes in the form of FPKM (fragments per kilobase of exon model per million reads mapped). Multiple RNA-Seq samples were pooled and the assembled transcripts were merged into a comprehensive set of transcripts before the downstream analysis. For this purpose, we used Cuffmerge, which is a meta-assembler [35,36] that merges all the assemblies parsimoniously.

Long Non-coding RNA Detection

After the quality check, read mapping, and transcript assembly, we filtered all the transcripts having a length of 200 bp or more to detect lncRNAs. We used the FEELnc (FEExible Extraction of lncRNAs) [37] tool to detect and annotate lncRNAs from RNA-seq assembled transcripts (Figure 1). FEELnc is a Random Forest-based model trained using general features such as k-mer frequencies and relaxed open reading frames (ORFs). The FEELnc is an all-in-one package that allows filtering of non-lncRNA-like transcripts and computation of coding potential scores. Figure 1 depicts the complete pipeline to detect lncRNAs.

miRNA Detection and Expression Profiling

MicroRNAs (miRNAs) are short non-coding RNAs (~22 nucleotides in length) which inhibit gene regulation through base-pairing at 3' or 5'-untranslated regions of mRNAs [38]. These regulate more than 30% of the genes expressed in animal genomes [39]. A single miRNA may have multiple targets, and a single mRNA can be regulated by several miRNAs. The regulatory potential of miRNAs affects nearly all physiological processes,

including cell growth, cell differentiation, and apoptosis. Many human cancers have been associated with the dysregulation of certain miRNAs [40]. The expression profile of miRNAs differs significantly between cancerous and non-cancerous tissues, and between aggressive and asymptomatic cases. miRNAs have been implicated as tumor suppressors of some cancers. Moreover, miRNAs may be good diagnostic markers; ie, miRNA expression profiles can be used to classify cancers [41].

For this study, we used the miRDeep* [42] software tool to detect miRNAs and their expression profiles in thymoma and healthy thymus tissues. MiRDeep* is a powerful tool for the detection of known and novel miRNAs from small RNA-seq data. Both known and novel miRNA expression profiles portray RNA-seq reads relative to pre-miRNA hairpins. Table 1 present the small RNA-seq data we used.

lncRNA-miRNA-mRNA Network and Functional Enrichment Analyses

To construct a lncRNA-mRNA-miRNA regulatory network, we considered gene co-expression association among the top-ranked lncRNAs and mRNAs using Pearson's correlation coefficient ($R^2 \geq 0.7$). We searched for genes with similar expression patterns to those of lncRNAs in thymoma tissues using GEPIA 2.0 (<http://gepia2.cancer-pku.cn/>), based on TCGA tumor, TCGA normal, and/or GTEx expression data for the detection of similar genes. Next, we retrieved the target genes of selected miRNAs with target-scores ≥ 75 from the miRDB database [43]. We used the Cytoscape v3.8 software (<https://cytoscape.org/>) to visualize the lncRNA-miRNA-mRNA networks constructed based on intersections of co-expressed lncRNAs, miRNAs, and mRNAs. In addition, we performed functional and pathway enrichment analyses of target genes of DELs and DEMs using the DAVID 6.8 tool (<https://david.ncifcrf.gov/>), which identifies gene classes that are over- or under-represented with the GO terms, and that have associations with disease phenotypes [44,45].

Results

In this study, we extracted data from 25 samples of thymoma patients and 25 of healthy individuals, Table 2 lists the basic characteristics of the individuals who provided samples.

Fusion Gene Detection

Out of 25 thymoma samples, we observed 21 fusion events in 9 samples (36%, 9/25), and 1 sample had multiple types of fusion genes. The identified fusion genes in thymus cancer and normal samples are shown in Table 3. Our results show that 5 out of 21 identified fusion genes are significant and unique to

Table 2. Characteristics of study samples of thymus tumor/cancer (n=25) and healthy thymus. (n=25) with the number of missing values.

Parameter	Category	Number (%) or Median (IQR)	Missing values (%)
Thymoma (thymus tumor/cancer)			
Age (years)	–	55.5 (37-76)	3 (12)
Sex	Male	12 (54.54)	3 (12)
	Female	10 (45.45)	
WHO Histo-type	A-B3	16 (64)	0 (0)
	C type	9 (36)	
Disease stage	I-II	12 (48)	0 (0)
	III-IV	13 (52)	
Healthy thymus			
Age (years)	–	2.39 (5 months-22 years)	0 (0)
Sex	Male	9 (2.18)	0 (0)
	Female	16 (2.51)	

Table 3. Summary of detected fusion genes and their statistics (arranged by significance level).

S. No.	Detected fusion genes	chr1	Breakpoint1	chr2	Breakpoint2	Strands (1/2)	Thymoma samples	Healthy samples
1	KMT2A-MAML2	11	118436492	11	95976698	(+/-)	9/25	1/25
2	HADHB-REEP1	2	26502983	2	86491164	(+/-)	6/25	0/25
3	COQ3-CGA	6	99842080	6	87804824	(+/+)	6/25	1/25
4	MCM4-SNTB1	8	48878682	8	48878526	(+/-)	5/25	0/25
5	IFT140-ACTN4	16	1616292	19	39138547	(+/-)	4/25	0/25
6	ZAP70-GLI2	2	98349346	2	121555044	(-/-)	6/25	2/25
7	UBR1-FLNB	15	43340689	3	58104716	(+/-)	6/25	3/25
8	PREPL-P4HTM	2	44569556	3	49038871	(-/+)	4/25	3/25
9	R3HDM1-KATNA1	2	136418963	6	149959696	(+/-)	4/25	3/25
10	ANXA6-SLC36A3	5	150496688	5	150675829	(-/-)	4/25	3/25
11	ADCY3-KDM3A	2	25065253	2	86684194	(+/-)	3/25	3/25
12	EP300-PRR5	22	41560134	22	45110471	(+/+)	3/25	3/25
13	DTNB-PSIP1	2	25600422	9	15510115	(+/+)	3/25	2/25
14	PDS5B-IL17RB	13	33233417	3	53890871	(+/+)	3/25	2/25
15	CLIP4-C5	2	29358532	9	123732529	(+/-)	3/25	1/25
16	EML4-TMEM163	2	42396776	2	135309662	(+/-)	3/25	1/25
17	MFS2B-DNMT3A	2	24236155	2	25458576	(-/+)	3/25	1/25
18	R3HDM1-WDR6	2	136482840	3	49053386	(-/-)	3/25	1/25
19	PIGS-SLC46A1	17	26898546	17	26733228	(-/-)	3/25	1/25
20	ATP9B-SYNPR	18	76903860	3	63466508	(+/+)	3/25	1/25
21	ARHGEF4-USP45	2	131785518	6	99936076	(-/+)	2/25	1/25

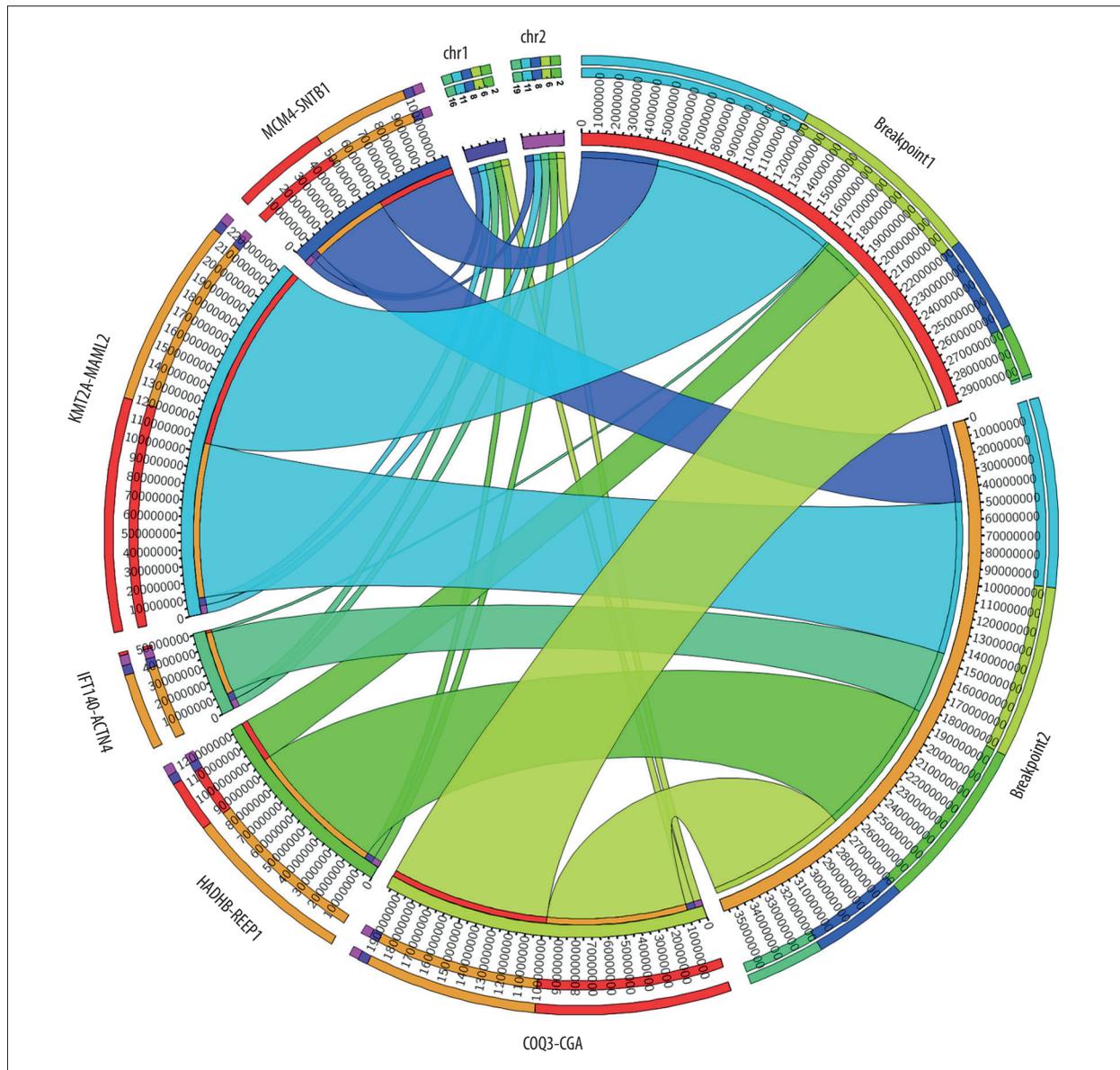


Figure 2. Circos plot of fusion genes mapped to chromosome and breakpoints on genome.

thymomas, namely KMT2A-MAML2 (detected in 9 samples of thymoma and 1 sample of the healthy thymus), HADHB-REEP1 (detected in 6 samples of thymoma), COQ3-CGA (detected in 6 samples of thymoma, and 1 sample of healthy thymus), MCM4-SNTB1 (detected in 5 samples of thymoma), and IFT140-ACTN4 (detected in 4 samples of thymoma). **Table 3** presents the fusion chromosomes and breakpoints. To further analyze the results, we created a Circos plot [46] of these 5 significant fusion genes (**Figure 2**) illustrating genomic rearrangements and the associations between 2 genomic positions. To validate the identified fusion genes computationally, we searched the junction sequence (identified by the STAR-Fusion) in the sequence reads of the samples and confirmed their presence. We used a methodology similar to that described by Singh and Li [47].

Identification of Aberrant lncRNAs

To detect the aberrantly expressed lncRNAs in thymoma, we performed a genome-wide lncRNA differential expression study between thymoma and healthy thymus tissues. We obtained a list of 65 differentially-expressed lncRNAs over the 2 sample types (P value <0.05 and log of fold-change (logFC) >2.0) (**Table 4**), out of which 6 lncRNAs had a $|\logFC| >5.0$. To validate the significance of the differentially-expressed lncRNAs (DElncRNAs) in thymoma using an independent tertiary dataset, we used the TCGA (The Cancer Genome Atlas) database. We compared the expression levels of 6 lncRNAs (P value=0, FDR=0, logFC >5.0), namely AFAP1-AS1, LINC00324, ADAMTS9-AS1, VLDLR-AS1, LINC00968, and NEAT1, between thymoma (THYM) and

Table 4. lncRNAs dysregulated in thymoma discovered by RNA-seq and validated by TCGA.

Transcript ID	Gene ID	Gene symbol	logFC	logCPM	p-Value	lncRNA type	Regulation type	FDR
ENST00000608442	ENSG00000272620	AFAP1-AS1	5.9488	3.5952	0.0000	Antisense	Up	0.0000
ENST00000315707	ENSG00000178977	LINC00324/ C17ORF44	5.8974	3.6420	0.0000	lincRNA	Up	0.0000
ENST00000601022	ENSG00000241158	ADAMTS9-AS1	-5.7239	3.9160	0.0000	antisense	Down	0.0000
ENST00000453601	ENSG00000236404	VLDLR-AS1	5.7215	3.6212	0.0000	antisense	Up	0.0000
ENST00000523664	ENSG00000246430	LINC00968	-5.3190	3.2842	0.0000	lincRNA	Down	0.0000
ENST00000616315	ENSG00000245532	NEAT1	-5.1874	3.0155	0.0000	lincRNA	Down	0.0000
ENST00000425914	ENSG00000224609	HSD52	-4.4465	2.2834	0.0000	lincRNA	Down	0.0000
ENST00000423808	ENSG00000232079	LINC01697	-4.1166	3.4826	0.0000	lincRNA	Down	0.0000
ENST00000330148	ENSG00000182366	FAM87A	4.0398	3.1122	0.0000	lincRNA	Down	0.0000
ENST00000553575	ENSG00000258498	DIO3OS	3.9456	5.7375	0.0000	lincRNA	Up	0.0001
ENST00000609696	ENSG00000225670	CADM3-AS1	-3.9167	4.9436	0.0000	antisense	Down	0.0001
ENST00000484765	ENSG00000242268	LINC02082	3.9025	2.1784	0.0000	lincRNA	Up	0.0003
ENST00000577000	ENSG00000185168	LINC00482	-3.7068	2.9948	0.0000	lincRNA	Down	0.0001
ENST00000500381	ENSG00000246363	LINC02458	3.4937	5.3952	0.0000	lincRNA	Up	0.0001
ENST00000530198	ENSG00000255176	AP000941.1	3.4311	2.5965	0.0001	antisense	Up	0.0102
ENST00000603439	ENSG00000271579	AC078880.3	-3.3799	2.5707	0.0001	lincRNA	Down	0.0158
ENST00000242109	ENSG00000122548	KIAA0087	-3.3738	6.4171	0.0000	lincRNA	Down	0.0088
ENST00000523301	ENSG00000245812	LINC02202	3.3481	5.4172	0.0000	lincRNA	Up	0.0001
ENST00000575424	ENSG00000262097	LINC02185	-3.2869	2.3678	0.0000	lincRNA	Down	0.0019
ENST00000605631	ENSG00000270412	AL136084.2	-3.2391	3.6568	0.0000	antisense	Down	0.0002
ENST00000626206	ENSG00000238018	AC093110.1	-3.1506	4.6319	0.0000	antisense	Down	0.0087
ENST00000603514	ENSG00000271334	LINC02104	3.1179	2.7402	0.0000	lincRNA	Up	0.0000
ENST00000418372	ENSG00000233117	LINC00702	3.0428	4.4037	0.0001	lincRNA	Up	0.0158
ENST00000572479	ENSG00000263257	AC040173.1	-3.0091	2.0691	0.0003	lincRNA	Down	0.0408
ENST00000519795	ENSG00000253978	CTB-113P19.1	2.9814	2.1697	0.0001	antisense	Up	0.0158
ENST00000609941	ENSG00000272823	AL445423.1	2.9430	2.9938	0.0000	lincRNA	Up	0.0049
ENST00000505254	ENSG00000249669	CARMN	2.9374	5.3167	0.0000	lincRNA	Up	0.0009
ENST00000417193	ENSG00000213373	LINC00671	-2.8571	3.3597	0.0000	lincRNA	Down	0.0033
ENST00000574724	ENSG00000228157	AC007952.2	2.8459	3.0662	0.0004	retained_intron	Up	0.0463
ENST00000556053	ENSG00000259134	LINC00924	2.8254	4.6589	0.0001	lincRNA	Up	0.0158
ENST00000418029	ENSG00000223901	AP001469.1	-2.8033	2.8406	0.0000	antisense	Down	0.0004
ENST00000505807	ENSG00000250208	FZD10-AS1	2.7905	5.3926	0.0001	lincRNA	Up	0.0191
ENST00000620326	ENSG00000277631	PGM5P3-AS1	2.7716	3.3111	0.0000	lincRNA	Up	0.0050

Table 4 continued. lncRNAs dysregulated in thymoma discovered by RNA-seq and validated by TCGA.

Transcript ID	Gene ID	Gene symbol	logFC	logCPM	p-Value	lncRNA type	Regulation type	FDR
ENST00000417354	ENSG00000230630	DNM3OS	2.7346	5.5551	0.0001	antisense	Up	0.0098
ENST00000619486	ENSG00000274080	AC005089.1	2.7190	3.0662	0.0000	sense_intronic	Up	0.0087
ENST00000427825	ENSG00000230623	AC104461.1	-2.6769	2.4333	0.0004	lincRNA	Down	0.0454
ENST00000577295	ENSG00000263586	HID1-AS1	2.6724	3.0135	0.0000	antisense	Up	0.0029
ENST00000430895	ENSG00000234484	AL032821.1	-2.6391	2.1961	0.0002	antisense	Down	0.0265
ENST00000446355	ENSG00000237813	AC002066.1	2.6230	2.1554	0.0003	antisense	Up	0.0398
ENST00000563759	ENSG00000260868	LINC01960	-2.6040	2.7845	0.0001	lincRNA	Down	0.0182
ENST00000624364	ENSG00000280339	AP001528.2	2.5928	5.5979	0.0000	TEC	Up	0.0022
ENST00000538077	ENSG00000255874	LINC00346	2.5205	5.8892	0.0004	lincRNA	Up	0.0478
ENST00000434306	ENSG00000234235	BOK-AS1	-2.5061	2.1082	0.0000	antisense	Down	0.0003
ENST00000513055	ENSG00000237187	NR2F1-AS1	2.4916	6.0892	0.0002	antisense	Up	0.0265
ENST00000573168	ENSG00000262319	AC007952.6	2.4867	2.9364	0.0002	antisense	Up	0.0272
ENST00000600473	ENSG00000269688	AC008982.2	-2.4289	3.8376	0.0001	sense_intronic	Down	0.0158
ENST00000450804	ENSG00000227467	LINC01537	2.4021	4.2479	0.0000	lincRNA	Up	0.0050
ENST00000435312	ENSG00000230148	HOXB-AS1	2.2070	6.9580	0.0001	antisense	Up	0.0120
ENST00000567984	ENSG00000260953	AC009093.4	2.1585	2.9114	0.0001	lincRNA	Up	0.0158
ENST00000437930	ENSG00000180139	ACTA2-AS1	2.0767	5.4572	0.0002	antisense	Up	0.0299
ENST00000508887	ENSG00000237125	HAND2-AS1	2.0078	3.8768	0.0003	antisense	Down	0.0021

normal tissues (Figure 3), and found them to be significantly dysregulated in thymomas. To further extend our analysis, we assessed the disease-free survival (DFS) of patients with high- and low-expression groups using the Mantel-Cox test. We found that both low and high expressions of these lncRNAs directly affect patient survival. Among these 6 lncRNAs, the high expression of AFAP1-AS1 and low expression of LINC00324 greatly affected the survival of thymoma patients. Figure 4 presents a DFS map showing the survival contribution of the top 25 dysregulated lncRNAs in various cancers, including thymoma (THYM), using TCGA datasets with a significance level of 0.05 estimated using the Mantel-Cox test. The DFS map in Figure 4 shows that aberrant expression of these 6 lncRNAs can be used to characterize thymomas from other cancers.

miRNA Detection and Analysis

To detect miRNAs and analyze differentially-expressed miRNAs (DEMs), we processed the datasets listed in Table 1. A significant number of miRNAs were overexpressed in the thymomas and their expressions were absent in the healthy thymus

tissues. We carried out 1695 miRNA expression profiles. We detected 294 DEMs ($|\logFC| \geq 1.0$); 198 miRNAs were upregulated and 96 were downregulated in thymomas (Supplementary Table 1 lists all the DEMs we found). Among the 294 DEMs, 10 miRNAs had $|\logFC| > 10.0$; Figure 5 shows their boxplot and a significant difference in the expression between thymomas and normal tissues. Figure 6 shows the logFC parameters of the top 50 DEMs with most DEMs upregulated with a maximal logFC of 15, and a few downregulated DEMs with a maximal logFC of 10.

Functional Annotation of the Top 10 Ranked lncRNAs

We examined the functional annotation of the top 10 ranked lncRNAs using the DAVID 5.8 tool. Each lncRNA was found to be associated with specific functions. For example, LINC00324 and FAM87A were associated with the sequence feature of putative uncharacterized protein C17orf44, and its transmembrane region, respectively. Also, LINC00324 was tissue-enriched with GO terms Brain and Lymph; FAM87A was tissue-enriched with the Go term Stomach; and HSD52 was

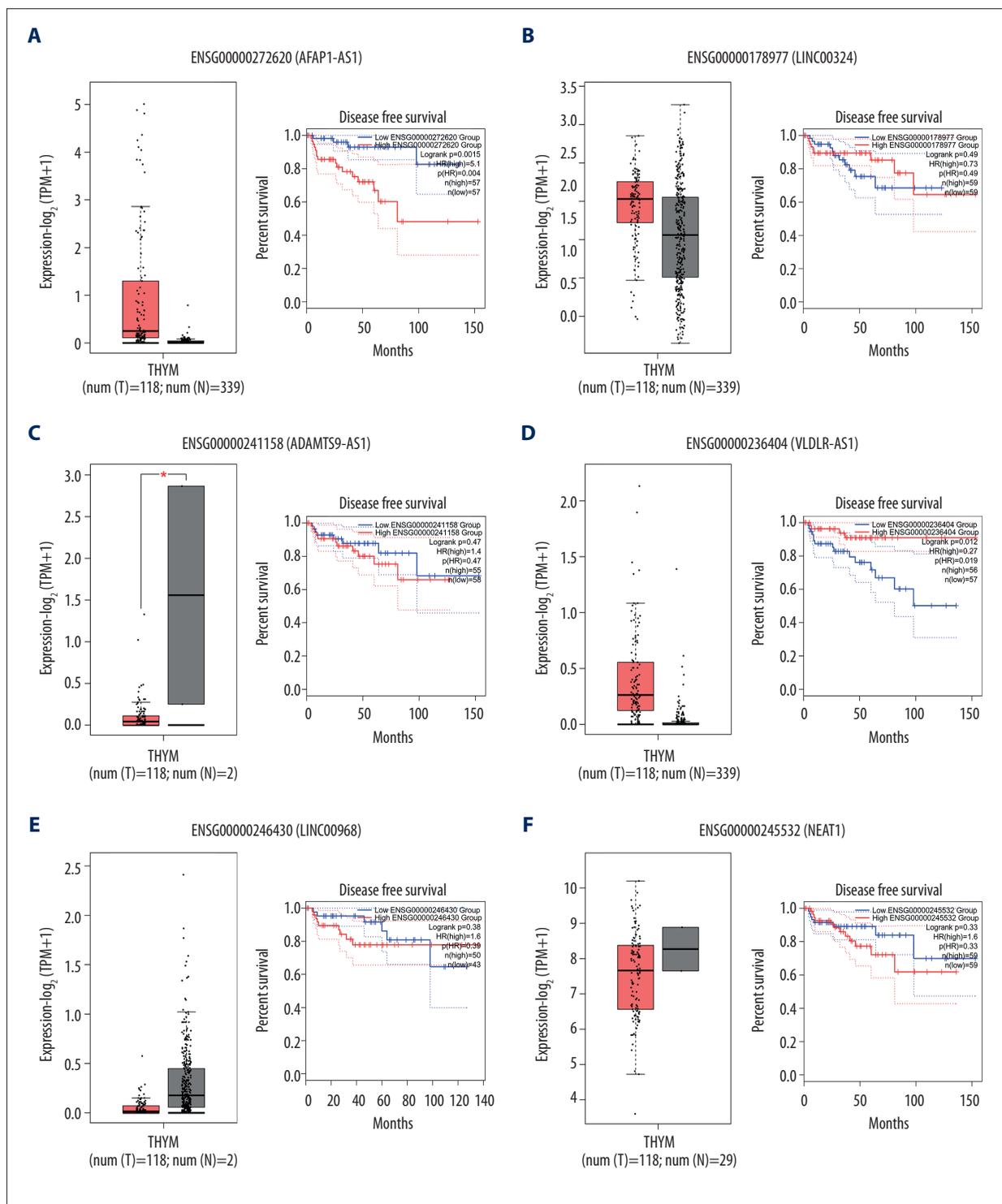


Figure 3. (A-F) Boxplot of normalized expression [\log_2 (TPM+1)] of thymoma samples (red color) and normal samples (gray color), and disease-free survival (DFS) analysis of patients in high- (red line) and low-expression (blue line) groups for the top 6 dysregulated lncRNAs.

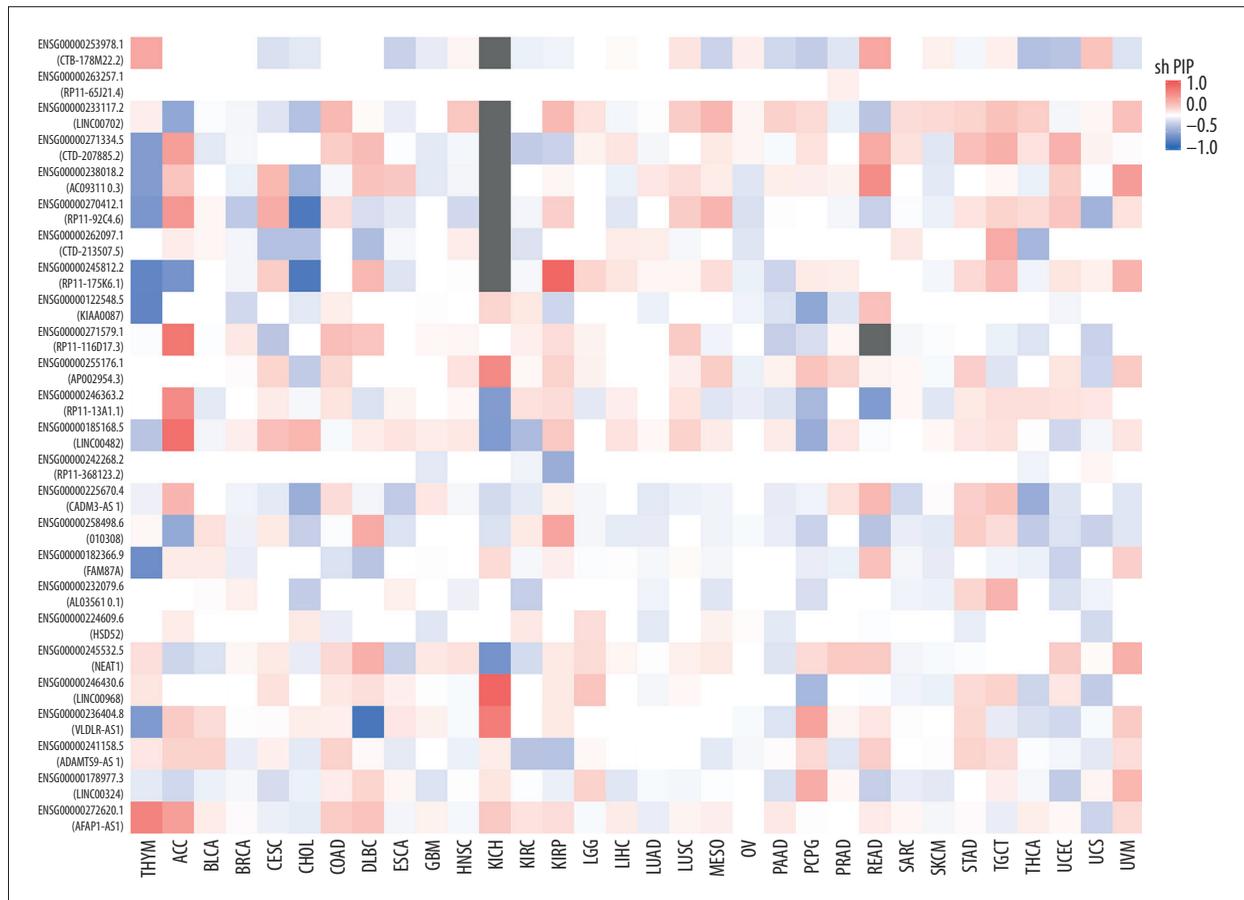


Figure 4. Disease-free survival (DFS) map showing survival contribution of top 25 dysregulated lncRNAs in various cancers including thymoma (THYM) using TCGA datasets with a significance level of 0.05 estimated using the Mantel-Cox test.

tissue-enriched with the Go term Testis. Moreover, VLDLR-AS1 and LINC00324 were associated with embryo development, thyroid tumor disease, germ cell tumor disease, and others; while NEAT1 was associated with adrenal tumor disease, thyroid tumor disease, and others.

Furthermore, we used the UCSC-TFBS algorithm of the DAVID tool to study protein interactions including transcription factors (TFs) with a set of target genes. Out of the top 10 ranked lncRNAs, 4 (AFAP1-AS1, VLDLR-AS1, LINC00324, and HSD52) are involved in protein interactions and have functions related to TFs. **Table 5** summarizes the functional annotations.

Differentially-expressed lncRNA-mRNA-miRNA Interaction Network

We constructed an lncRNA-mRNA-miRNA regulation network based on the gene co-expression correlation analysis between DELs, mRNAs, and DEMs identified in thymomas (**Figure 7**). In this network, the upregulated miRNA hsa-let-7a-3 exhibits interactions with 8 protein-coding genes (INSR, IGF1, IL10, IGF1R, ITGB3, COL5A2, ZNF322, PXDN, TGFB1) and can increase their

expressions. The majority of target genes of DELs and DEMs are enriched in several biological pathways, including the PI3K-Akt signaling pathway (P value=0.0), the FoxO signaling pathway (P value=0.0), the HIF-1 signaling pathway (P value=0.0), the proteoglycans in cancer (P value=0.01), and other cancer pathways (P value=0.01), the Supplementary Table 2 shows these details. Most target genes were associated with various GO terms, including immune response (GO: 0006955), hepatic immune response (GO: 0002384), positive regulation of DNA replication (GO: 0045740), positive regulation of cell proliferation (GO: 0008284), positive regulation of MAPK cascade (GO: 0043410), positive regulation of DNA replication (GO: 0045740), positive regulation of cell proliferation (GO: 0008284), negative regulation of extrinsic apoptotic signaling pathway (GO: 2001237), and negative regulation of apoptotic process (GO: 0043066). Supplementary Table 3 shows these details. Our network and pathway analyses show the overexpression of miRNA clusters activates the PI3K-Akt/FoxO/HIF-1/Rap-1 signaling pathway, suggesting that PI3K/Akt/HIF-1/Rap-1 inhibitors may be therapeutic targets for thymoma patients. The hsa-let-7a family of miRNAs is thought to inhibit migration, invasion, and tumor growth by targeting the AKT2

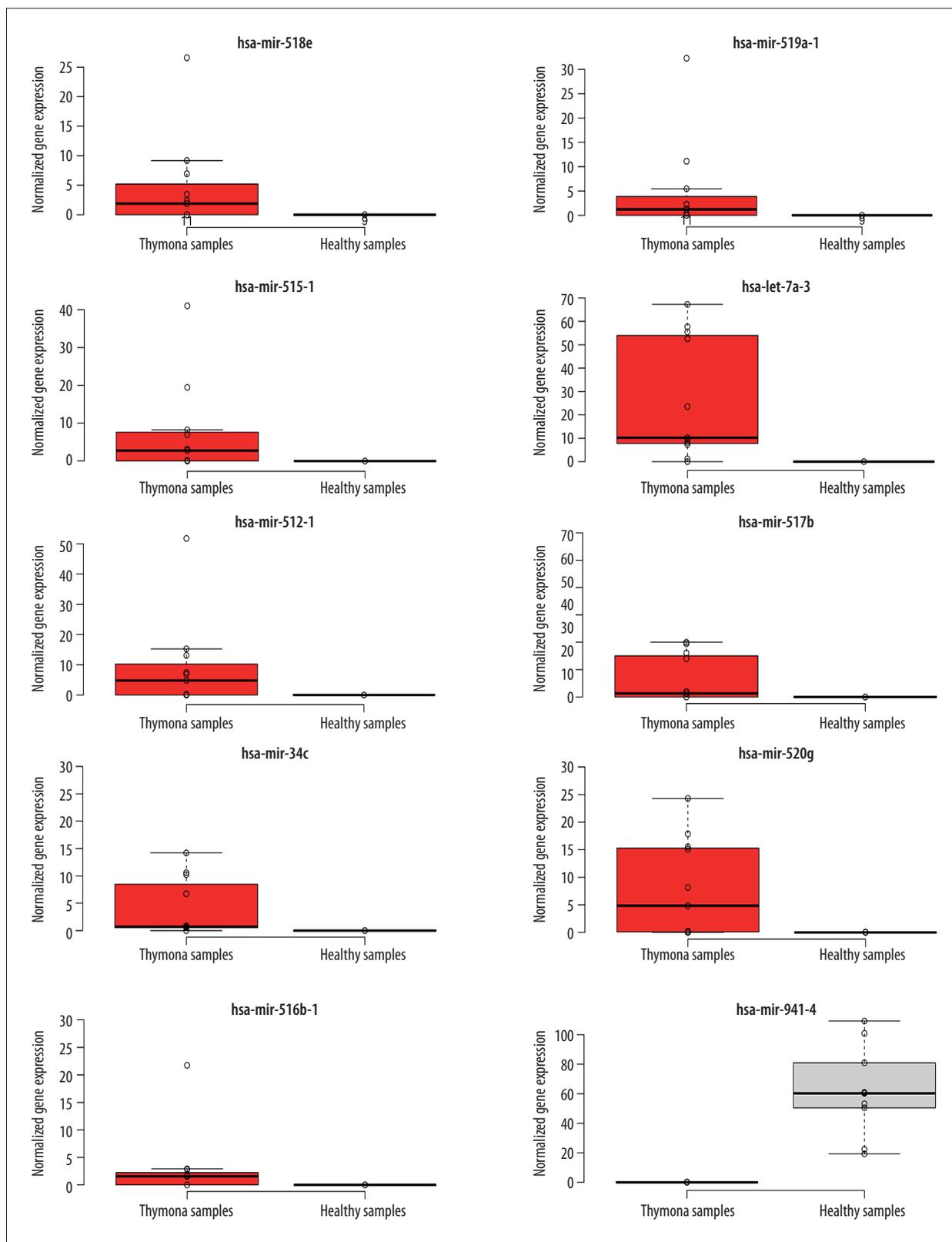


Figure 5. Boxplot of first 10 differentially-expressed miRNAs (DEMs) showing their normalized gene expression over thymoma versus those of healthy samples.

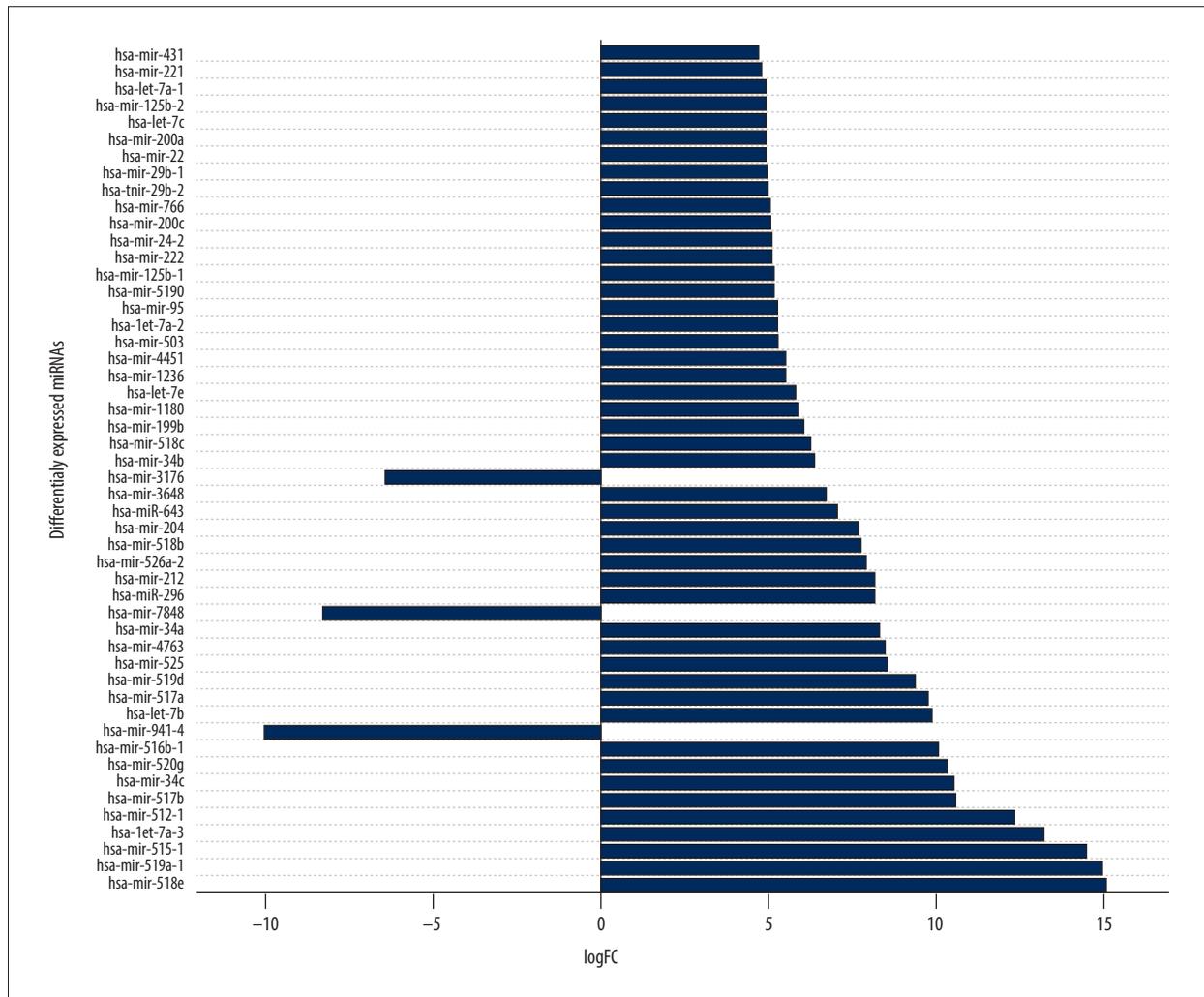


Figure 6. First 50 differentially-expressed miRNAs (DEMs) and their logFC values.

in papillary thyroid carcinoma [48]. Increased plasma levels of hsa-let-7 have been found in patients with breast, prostate, colon, renal, liver, gastric, ovarian, and thyroid cancers [49].

Discussion

Thymomas, predominant and clinically aggressive among the thymic epithelial tumors, are the most common tumors in the anterior mediastinum, but are infrequent in comparison to other thoracic tumors. The thymus is the primary site of T cell development and plays an essential role in the development of the immune system, T cell differentiation, and maturation [1]. The thymus has been shown to be associated with several paraneoplastic conditions, and a variety of immune and non-immune and endocrine disorders [9]. One-third of patients with thymus malignancy are mostly asymptomatic during their initial stages. However, as the malignancy progresses, patients start to present signs such as cough, chest-pains,

nerve palsy, and superior vena cava syndrome [50]. Engels reported some major related cancers that can form due to thymoma, including lung cancer, thyroid cancer, prostate cancer, leukemia, and sarcomas [51]. Liu et al [52] indicated in their case study that an elderly patient suffering from acute gastric volvulus (AGV) had a primary thymoma, and they hypothesized that thymoma/thymus cancers can develop either individually or underlying a secondary disease.

Genomic instability (ie, gene mutations, translocations, deletions, copy number alterations, and inversions) can lead to cancer and other diseases. These genetic events can also trigger gene fusions. Fusion genes are a class of oncogenes that can be targeted for therapeutics or as diagnostic tools because of their inherent expression in tumor tissues. Next-generation sequencing has revolutionized sequencing by creating high-throughput and low-cost methods, and fusion gene detection has become easy and cost-effective. Targeting of fusion genes has facilitated the development of several successful anti-cancer drugs

Table 5. lncRNAs and their predicted protein interactions using the UCSC-TFBS algorithm.

ID	Gene Name	Species	UCSC_TFBS
84740	AFAP1-AS1 (AFAP1 antisense RNA 1)	<i>Homo sapiens</i>	AML1, AP1FJ, AP2REP, AP4, AREB6, ARNT, ARP1, CDC5, CDPCR3HD, COMP1, CP2, CREBP1, ELK1, EN1, ER, FOXJ2, FREAC7, GCNF, GRE, HAND1E47, HFH3, HNF3B, HOXA3, HTF, LYF1, MEF2, MSX1, MYCMAX, MYOD, NFAT, NFY, NKX25, NKX61, P53, PAX4, PAX5, PPARA, PPARG, RFX1, RORA2, RP58, SREBP1, SRY, TAL1BETAITF2, TAXCREB, TBP, TCF11, USF, ZIC2, ZID
401491	VLDLR-AS1 (VLDLR antisense RNA 1)	<i>Homo sapiens</i>	AHRARNT, AML1, AP1, AP1FJ, AP2GAMMA, AP4, AREB6, ARNT, ATF, BACH1, BACH2, BRACH, CART1, CDC5, CDP, CDPCR1, CEBP, CEBPB, CHX10, COMP1, COUP, CP2, E47, E4BP4, EN1, ER, EVI1, FOXD3, FOXJ2, FOXO4, FREAC4, FREAC7, GATA, GATA1, GCNF, GF11, GR, HEN1, HFH1, HFH3, HLF, HMX1, HNF1, HNF3B, HOXA3, HSF2, HTF, ISRE, LHX3, LMO2COM, LUN1, LYF1, MAX, MEF2, MEIS1, MEIS1BHOXA9, MSX1, MYCMAX, MYOD, MYOGNF1, MZF1, NCX, NF1, NFAT, NFY, NKX22, NKX25, NKX3A, NKX61, NRSF, OCT, OCT1, OLF1, P300, P53, PAX4, PAX6, PBX1, POU3F2, POU6F1, PPARA, PPARG, RORA2, RP58, RREB1, RSRFC4, S8, SOX5, SOX9, SREBP1, SRF, SRY, STAT1, STAT3, STAT5B, TAL1BETAITF2, TATA, TAXCREB, TCF11, TCF11MAFG, TGIF, USF, XBP1, YY1, ZIC1, ZID
284029	LINC00324 (long intergenic non-protein coding RNA 324)	<i>Homo sapiens</i>	AP1, AP4, AREB6, ARP1, BRACH, CART1, CEBP, CETS1P54, CMYB, CP2, E47, FOXJ2, FREAC4, GATA3, GRE, HAND1E47, HEN1, HNF4, ISRE, MEF2, MZF1, NFY, OCT1, OLF1, PBX1, PPARG, RORA2, STAT3, STAT5A, TAL1BETA47, TAL1BETAITF2, TCF11MAFG, TST1, USF, ZIC1
729467	HSD52 (uncharacterized LOC729467)	<i>Homo sapiens</i>	AP1, ARP1, BACH1, BACH2, BRACH, CEBP, CEBPB, COUP, EGR3, EVI1, FOXD3, FOXO3, FOXO4, FREAC2, GATA1, HFH1, HNF3B, IK3, IRF2, IRF7, ISRE, LUN1, MEF2, MZF1, NFAT, NFE2, NKX22, OCT, OCT1, PAX4, POU3F2, PPARG, RP58, RSRFC4, SRF, STAT, TAXCREB, TCF11MAFG, YY1

and targeted therapies [53]. The roles of lncRNAs in genomic regulation are multiple, they involve transcriptional and post-transcriptional regulation, and imprinting and epigenetic modifications. Studies have documented lncRNAs that play roles in tumorigenesis and tumor progression [15-17]; in particular, aberrant lncRNA expression or lncRNA dysfunction can lead to DNA damage, immune escape, and cellular metabolic disorders in cancer cells. lncRNAs can regulate cell proliferation, apoptosis, migration, invasion, and maintenance of stemness during the development of cancer. In addition, many lncRNAs are transcriptionally regulated by key tumor suppressors or oncogenes [15]. The inherent heterogeneity and diversity of the lncRNA world make the complex tumorigenesis process more interesting than it would otherwise be [17]. Hence, identification and characterization of lncRNAs could lead to development of new clinical approaches to understand complex tumorigenesis and to treat cancers. In this study, we identified and characterized fusion genes, lncRNAs, and miRNAs from RNA-seq data of thymoma patients, identified aberrant lncRNAs and miRNAs in thymomas, built an lncRNA-mRNA-miRNA regulatory network in thymoma, and performed a functional enrichment analysis.

Our RNA-seq analysis results show frequent translocations of fusion gene events with KMT2A-MAML2, and others such as HADHB-REEP1, COQ3-CGA, and MCM4-SNTB1. KMT2A encodes a transcriptional co-factor that plays an important role in gene expression regulation during early development and hematopoiesis. This protein binds DNA and methylates the histone H3 at lysine-4 to regulate HOX and other genes. KMT2A has been reported to be a translocation partner with more than 80 unique fusion partners [54]. MAML2, a member of the Mastermind-like family of proteins, binds the ankyrin repeat domain Notch receptors (ICN1-4). The KMT2A-MAML2 gene fusion was reported in a patient with therapy-related acute myeloid leukemia [55], and it was also reported in aggressive histologic thymoma subtypes [56], a finding in line with our results. The KMT2A-MAML2 fusion has also been shown to suppresses promoter activation of the NOTCH1 target gene HES1 [57], and it demonstrates oncogenic activities [58]. The next most frequent fusion genes detected in our analysis were HADHB-REEP1 and COQ3-CGA. HADHB encodes a mitochondrial trifunctional protein associated with peroxisomal disease, valproic acid pathway, and mitochondrial fatty

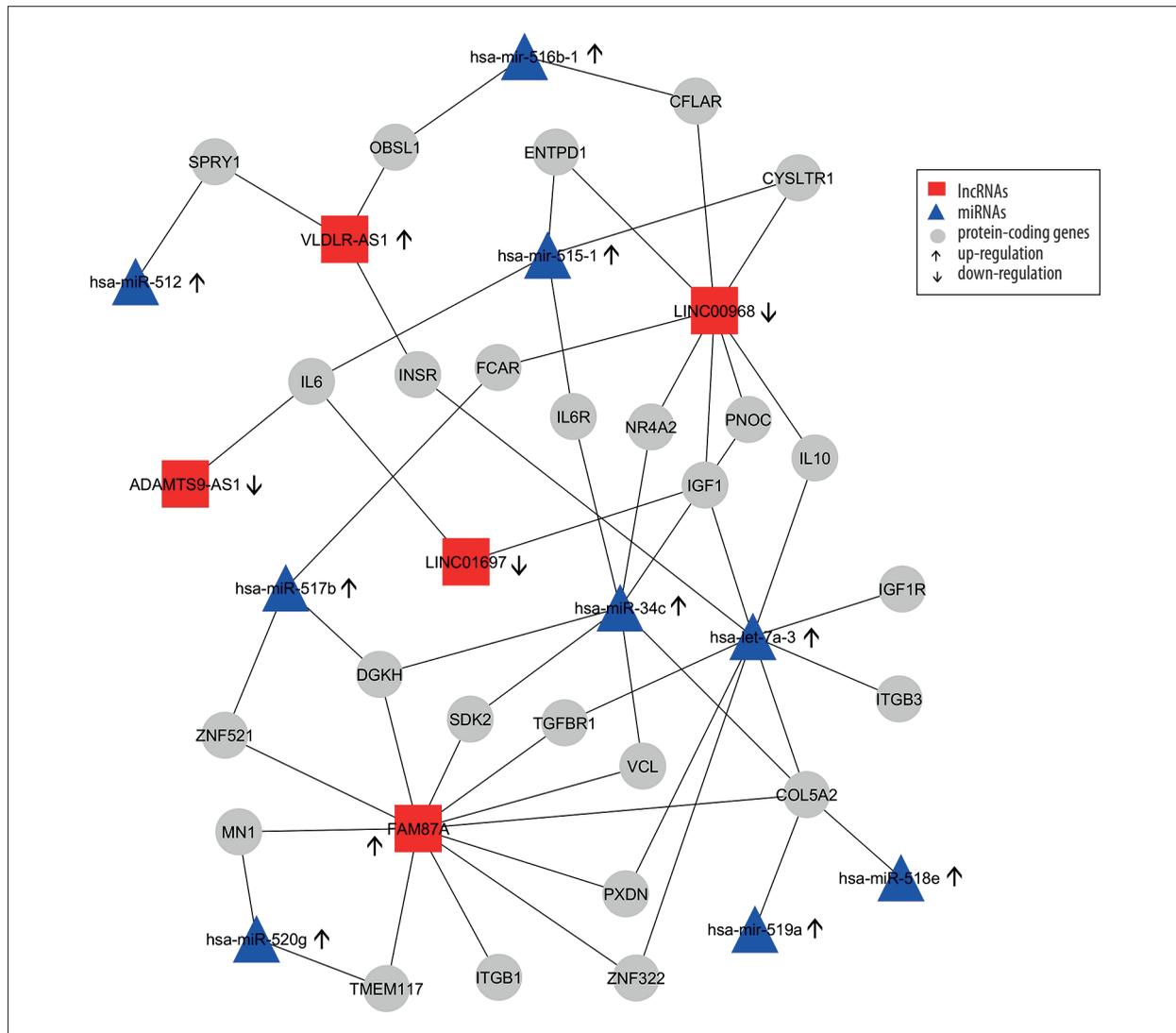


Figure 7. Differentially-expressed lncRNA-mRNA-miRNA interaction network in thymoma. Boxes represent lncRNAs, circles represent protein coding genes, and triangles represent miRNAs. Up arrows represent upregulated nodes, while down arrows represent downregulated nodes in thymomas. Lines between lncRNAs-mRNAs-miRNAs represent regulatory networks among them.

acid beta-oxidation. REEP1 encodes a mitochondrial protein whose mutation has been reported to cause spastic paraplegia, which is a neurodegenerative disorder. To the best of our knowledge, these 2 fusions are novel, and have not been reported before. The gene fusion MCM4-SNTB1 was reported in a study including a single case of metastatic thymic adenocarcinoma [59]. Other fusion transcripts reported in thymomas are FABP2-C4orf3 and CTBS-GNG5 [59], but no fusion events in TETs have been published [5]. Although, gene GTF2I is the most frequently mutated and is considered a master genetic regulator in thymomas; it has not been found to be involved in fusion events [5,60,61].

To our astonishment, we detected differentially-expressed lncRNAs (DEMs), many reported in the literature as promoting the

proliferation of various malignancies including gastric, ovarian, breast, prostate cancers, non-small cell lung cancer (NSCLC), leukemia, sarcomas, and head and neck squamous cell carcinomas [62-71]. LINC01485 has been shown to be responsible for the formation of gastric cancer by suppression of EGFR ubiquitination and triggering of the Akt signaling pathway [70]. AFAP1-AS1 has also been shown to cause severe human cancers [67]. LINC01697 is involved in the development of lung squamous cell carcinoma and can be used as a biomarker [72].

Another lncRNA, LINC00324, has been shown to cause gastric cancer by interacting with the human antigen R (HuR) [73]. ADAMTS9-AS1 has been proved to inhibit the Wnt signaling pathway in colorectal cancer, and can be considered a prognostic marker [74]. VLDLR-AS1 has been suggested to promote

fat loss in cancer cachexia [75]. On the other hand, NEAT1 has been shown to promote a variety of human cancers [68], while LINC00968 has been recognized as an oncogene that activates the cancer prevalent signaling pathways; mainly Akt, PI3K, and mTOR pathways [76]. The lncRNAs *AFAP1-AS1*, *LINC00324*, and *VLDLR-AS1* have been reported to promote the proliferation of thymomas because of their overexpression in gene expression profiles [77]. The silencing of LINC00324 has been shown to halt and suppress the G1/G0 phases of the cell cycle, thereby triggering programmed cell death and stopping the function of the Notch signaling pathway [78]. Also, the Notch1 signaling pathway of thymocyte development indicates that Notch1IC can lead to thymoma formation in humans. Abnormal expression of *AFAP1-AS1* has also been shown to promote cancer development and tumor formation, and it can be used as a biomarker for tumor identification and treatment management [79]. Moreover, *VLDLR-AS1* expression has been shown to be directly associated with thymus cancer prognosis [80]. Moreover, *ADAMTS-AS1*, *HSD52*, *LINC00968*, and *LINC01697* predict recurrence in patients with TETs, especially those in the high-risk group with shorter RFS. These lncRNAs may assist in finding TET therapeutic targets [81]. Nuclear Enriched Abundant Transcript 1 (NEAT1) on the other hand, is recognized as a multifactorial agent for tumorigenesis. NEAT1 activates the progression of T lymphocytes by triggering miRNA hsa-mir-146b-5p to overexpress NOTCH1 in the Notch signaling pathway [82]. We suggest that these lncRNAs in the Notch signaling pathway must be further evaluated in patients with thymoma to identify potential prognostic markers or therapeutic targets for treatment.

Our study detected several DEMs, most of them upregulated in thymoma tissue and associated with different biological and disease pathways. For instance, the hsa-let-7a family of miRNAs inhibit migration, invasion, and tumor growth by targeting AKT2 in papillary thyroid carcinoma [48]. Our network and pathway analysis of differentially-expressed lncRNA-mRNA-miRNA showed that overexpression of miRNA clusters activated the PI3K-Akt/FoxO/HIF-1/Rap-1 signal pathway, suggesting a possible role for PI3K/Akt/HIF-1/Rap-1 inhibitors in thymoma patients.

Supplementary Data

Supplementary Table 1. List of differentially-expressed miRNAs (DEMs) with their scores.

Supplementary Table 2. Result of KEGG pathway enrichment analysis.

Supplementary Table 3. GO biological process and molecular functional annotation.

Supplementary Tables 1, 2 and 3 available from the corresponding author on request.

Conclusions

In this study, we reported fusion genes, lncRNAs, and miRNAs detected from RNA-seq data of thymoma patients. Also, we detected aberrant lncRNAs and miRNAs, constructed a lncRNA-mRNA-miRNA regulatory network, and performed a functional enrichment analysis. We detected 21 fusion genes, out of which 5 were significantly expressed and unique in thymoma tissues (*KMT2A-MAML2*, *HADHB-REEP1*, *COQ3-CGA*, *MCM4-SNTB1*, and *IFT140-ACTN4*); some of them have been reported in the literature. Additionally, we reported 65 differentially-expressed lncRNAs in thymomas, including *AFAP1-AS1*, *LINC00324*, *ADAMTS9-AS1*, *VLDLR-AS1*, *LINC00968*, and *NEAT1*, that have been validated with the TCGA database. Our DFS analysis suggests that markers with both low or high expressions directly affect patient survival. Our DFS map demonstrated that aberrant expression of these lncRNAs can be used to differentiate thymomas from other cancers. The functional annotation of the top 10 ranked lncRNAs was associated with specific molecular functions.

Our analysis of small ncRNAs detected 1695 miRNAs, with a significant number of miRNAs overexpressed in thymomas. Network analysis of lncRNA-mRNA-miRNA regulation identified a group of miRNAs upregulated in thymomas and triggering the expression of target protein-coding genes. As a result, many biological pathways are disrupted, including the PI3K-Akt signaling pathway, the FoxO signaling pathway, the HIF-1 signaling pathway, and proteoglycans. Finally, our gene set enrichment analysis revealed that these target protein-coding genes are associated with various molecular functions, including systemic immune response, hepatic immune response, positive regulation of DNA replication, positive regulation of cell proliferation, positive regulation of MAPK cascade, negative regulation of extrinsic apoptotic signaling pathway, and negative regulation of apoptosis. Our results show that overexpression of miRNA clusters activates the PI3K-Akt, FoxO, HIF-1, and Rap-1 signaling pathways, suggesting that PI3K, Akt, HIF-1, and Rap-1 inhibitors may be therapeutic in patients with thymoma.

Conflict of Interest

None.

References:

1. Chaudhry MS, Velardi E, Dudakov JA, et al. Thymus: The next (re)generation. *Immunol Rev*. 2016;271:56-71
2. Eng TY, Fuller CD, Jagirdar J, et al. Thymic carcinoma: State of the art review. *Int J Radiat Oncol Biol Phys*. 2004;59:654-64
3. Zucali PA, De Pas T, Palmieri G, et al. Phase II study of everolimus in patients with thymoma and thymic carcinoma previously treated with cisplatin-based chemotherapy. *J Clin Oncol*. 2018;36: 342-49
4. Zhao C, Rajan A. Immune checkpoint inhibitors for treatment of thymic epithelial tumors: how to maximize benefit and optimize risk? *Mediastinum*. 2019;3:35
5. Radovich M, Pickering CR, Felau I, et al. The integrated genomic landscape of thymic epithelial tumors. *Cancer Cell*. 2018;33:244-58.e10
6. Rashid MA, Newton MAH, Hoque MdT, et al. Mixing energy models in genetic algorithms for on-lattice protein structure prediction. *Biomed Res Int*. 2013;2013:924137
7. Marx A, Chan JKC, Coindre J-M, et al. The 2015 World Health Organization classification of tumors of the thymus: Continuity and changes. *J Thorac Oncol*. 2015;10:1383-95
8. Meng F-J, Wang S, Zhang J, et al. Alteration in gene expression profiles of thymoma: Genetic differences and potential novel targets. *Thorac Cancer*. 2019;10:1129-35
9. Detterbeck FC, Parsons AM. Thymic tumors. *Ann Thorac Surg*. 2004;77:1860-69
10. Srirajaskanthan R, Toubanakis C, Dusmet M, et al. A review of thymic tumours. *Lung Cancer*. 2008;60:4-13
11. Mangan KF, Volkin R, Winkelstein A. Autoreactive erythroid progenitor-T suppressor cells in the pure red cell aplasia associated with thymoma and panhypogammaglobulinemia. *Am J Hematol*. 1986;23:167-73
12. Fried AJ, Bonilla FA. Pathogenesis, diagnosis, and management of primary antibody deficiencies and infections. *Clin Microbiol Rev*. 2009;22:396-414
13. ENCODE Project Consortium: The ENCODE (ENCyclopedia Of DNA Elements) project. *Science*. 2004;306:636-40
14. Da Sacco L, Baldassarre A, Masotti A. Bioinformatics tools and novel challenges in long non-coding RNAs (lncRNAs) functional analysis. *Int J Mol Sci*. 2011;13:97-114
15. Huarte M. The emerging role of lncRNAs in cancer. *Nat Med*. 2015;21:1253-61
16. Schmitt AM, Chang HY. Long noncoding RNAs in cancer pathways. *Cancer Cell*. 2016;29:452-63
17. Jiang M-C, Ni J-J, Cui W-Y, et al. Emerging roles of lncRNA in cancer and therapeutic opportunities. *Am J Cancer Res*. 2019;9:1354-66
18. Yu Y-P, Liu P, Nelson J, et al. Identification of recurrent fusion genes across multiple cancer types. *Sci Rep*. 2019;9:1074
19. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet:journal*. 2011;17:10-12
20. Robinson DR, Wu Y-M, Kalyana-Sundaram S, et al. Identification of recurrent NAB2-STAT6 gene fusions in solitary fibrous tumor by integrative sequencing. *Nat Genet*. 2013;45:180-85
21. Tate JG, Bamford S, Jubb HC, et al. COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res*. 2019;47:D941-47
22. Lee SJ, Hong JY, Kim K, et al. Detection of fusion genes using a targeted RNA sequencing panel in gastrointestinal and rare cancers. *J Oncol*. 2020;2020:4659062
23. Fernandez-Cuesta L, Sun R, Menon R, et al. Identification of novel fusion genes in lung cancer using breakpoint assembly of transcriptome sequencing data. *Genome Biol*. 2015;16:7
24. Cleyne A, Szalat R, Kemal Samur M, et al. Expressed fusion gene landscape and its impact in multiple myeloma. *Nat Commun*. 2017;8:1893
25. Schram AM, Chang MT, Jonsson P, et al. Fusions in solid tumours: Diagnostic strategies, targeted therapy, and acquired resistance. *Nat Rev Clin Oncol*. 2017;14:735-48
26. McPherson A, Hormozdiari F, Zayed A, et al. deFuse: An algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput Biol*. 2011;7:e1001138
27. Benelli M, Pescucci C, Marseglia G, et al. Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. *Bioinformatics*. 2012;28:3232-39
28. Davidson NM, Majewski IJ, Oshlack A. JAFFA: High sensitivity transcriptome-focused fusion gene detection. *Genome Med*. 2015;7:43
29. Haas BJ, Dobin A, Stransky N, et al. STAR-Fusion: Fast and accurate fusion transcript detection from RNA-Seq. *bioRxiv*: 2017;2017.120295
30. Ma C, Shao M, Kingsford C. SQUID: Transcriptomic structural variation detection from RNA-seq. *Genome Biol*. 2018;19:52
31. Haas BJ, Dobin A, Li B, et al. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol*. 2019;20:213
32. Dobin A, Davis CA, Schlesinger F, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15-21
33. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357-59
34. Kim D, Pertea G, Trapnell C, Pimentel H, et al. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14:R36
35. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7:562-78
36. Trapnell C, Hendrickson DG, Sauvageau M, et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. 2013;31:46-53
37. Wucher V, Legeai F, Hédan B, et al. FEELnc: A tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res*. 2017;45:e57
38. Faiza M, Tanveer K, Fatihi S, et al. Comprehensive overview and assessment of miRNA target prediction tools in human and drosophila melanogaster. *Current Bioinformatics*. 2017;14:432-45
39. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005;120:15-20
40. Stuopelytė K, Daniūnaitė K, Jankevičius F, et al. Detection of miRNAs in urine of prostate cancer patients. *Medicina (Kaunas)*. 2016;52:116-24
41. Lu J, Getz G, Miska EA, et al. MicroRNA expression profiles classify human cancers. *Nature*. 2005;435:834-38
42. An J, Lai J, Lehman M, et al. MiRDeep*: An integrated application tool for miRNA identification from RNA sequencing data. *Nucleic Acids Res*. 2012;41:727-37
43. Chen Y, Wang X miRDB: An online database for prediction of functional microRNA targets. *Nucleic Acids Res*. 2020;48:D127-31
44. Raza K. Analysis of microarray data using artificial intelligence-based techniques. *Biotechnology*, 2019;865-88
45. Raza K. Reconstruction, topological and gene ontology enrichment analysis of cancerous gene regulatory network modules. *Current Bioinformatics*. 2016;11:1-1
46. Krzywinski M, Schein J, Birol I, et al. Circos: An information aesthetic for comparative genomics. *Genome Res*. 2009;19:1639-45
47. Singh S, Li H. Prediction, characterization, and in silico validation of chimeric RNAs. *Methods Mol Biol*. 2020;2079:3-12
48. Zhou B, Shan H, Su Y, et al. Let-7a inhibits migration, invasion and tumor growth by targeting AKT2 in papillary thyroid carcinoma. *Oncotarget*. 2017;8:69746-55
49. Chirshv E, Oberg KC, Ioffe YJ, et al. Let-7 as biomarker, prognostic indicator, and therapy for precision medicine in cancer. *Clin Transl Med*. 2019;8:24
50. ASCO: Thymoma and thymic carcinoma – symptoms and signs. *CancerNet*, 2012
51. Engels EA. Epidemiology of thymoma and associated malignancies. *J Thorac Oncol*. 2010;5:S260-65
52. Liu A, Gao X, Zhao L. Thymoma with acute gastric volvulus: A case report. *BMC Cancer*. 2017;17:801
53. Kamps R, Brandão RD, Bosch BJ van den, et al. Next-generation sequencing in oncology: Genetic diagnosis, risk prediction and cancer classification. *Int J Mol Sci*. 2017;18:308
54. Milne TA, Briggs SD, Brock HW, et al. MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol Cell*. 2002;10:1107-17

55. Obama K, Furukawa Y, Tara M, et al. Secondary monocytic leukemia with rearrangement of the MLL gene occurring during the course of adult T-cell leukemia. *Int J Hematol.* 1998;68:323-26
56. Massoth LR, Hung YP, Dias-Santagata D, et al. Pan-cancer landscape analysis reveals recurrent KMT2A-MAML2 gene fusion in aggressive histologic subtypes of thymoma. *JCO Precis Oncol.* 2020;4:PO.19.00288
57. Nemoto N, Suzukawa K, Shimizu S, et al. Identification of a novel fusion gene MLL-MAML2 in secondary acute myelogenous leukemia and myelodysplastic syndrome with inv(11)(q21q23). *Genes Chromosomes Cancer.* 2007;46:813-19
58. Wächter K, Kowarz E, Marschalek R. Functional characterisation of different MLL fusion proteins by using inducible Sleeping Beauty vectors. *Cancer Lett.* 2014;352:196-202
59. Lee Y, Park S, Lee S-H, Lee H. Characterization of genetic aberrations in a single case of metastatic thymic adenocarcinoma. *BMC Cancer.* 2017;17:330
60. Petrini I, Meltzer PS, Kim I-K, et al. A specific missense mutation in GTF2I occurs at high frequency in thymic epithelial tumors. *Nat Genet.* 2014;46:844-49
61. Oberndorfer F, Müllauer L. Genomic alterations in thymoma – molecular pathogenesis? *J Thoracic Dis.* 2020;12:7536-44
62. Guo S, Chen W, Luo Y, et al. Clinical implication of long non-coding RNA NEAT1 expression in hepatocellular carcinoma patients. *Int J Clin Exp Pathol.* 2015;8:5395-402
63. Han D, Wang J, Cheng G. LncRNA NEAT1 enhances the radio-resistance of cervical cancer via miR-193b-3p/CCND1 axis. *Oncotarget.* 2018;9:2395-409
64. Huang G, He X, Wei X-L. LncRNA NEAT1 promotes cell proliferation and invasion by regulating miR-365/RGS20 in oral squamous cell carcinoma. *Oncol Rep.* 2018;39:1948-56
65. Li X, Ren Y, Zuo T. Long noncoding RNA LINC00978 promotes cell proliferation and invasion in non-small cell lung cancer by inhibiting miR-6754-5p. *Mol Med Rep.* 2018;18:4725-32
66. Fang X-N, Yin M, Li H, et al. Comprehensive analysis of competitive endogenous RNAs network associated with head and neck squamous cell carcinoma. *Sci Rep.* 2018;8:10544
67. Ji D, Zhong X, Jiang X, et al. The role of long non-coding RNA AFAP1-AS1 in human malignant tumors. *Pathol Res Pract.* 2018;214:1524-31
68. Dong P, Xiong Y, Yue J, et al. Long non-coding RNA NEAT1: A novel target for diagnosis and therapy in human tumors. *Front Genet.* 2018;9:471
69. Xiu D-H, Liu G-F, Yu S-N, et al. Long non-coding RNA LINC00968 attenuates drug resistance of breast cancer cells through inhibiting the Wnt2/β-catenin signaling pathway by regulating WNT2. *J Exp Clin Cancer Res.* 2019;38:94
70. Zhou J, Wu L, Li W, et al. Long noncoding RNA LINC01485 promotes tumor growth and migration via inhibiting EGFR ubiquitination and activating EGFR/Akt signaling in gastric cancer. *Oncotargets Ther.* 2020;13:8413-25
71. Li Y, Cao X, Li H. Identification and validation of novel long non-coding RNA biomarkers for early diagnosis of oral squamous cell carcinoma. *Front Bioeng Biotechnol.* 2020;8:256
72. Liu J, Yao Y, Hu Z, et al. Transcriptional profiling of long-intergenic noncoding RNAs in lung squamous cell carcinoma and its value in diagnosis and prognosis. *Mol Genet Genomic Med.* 2019;7:e994
73. Zou Z, Ma T, He X, et al. Long intergenic non-coding RNA 00324 promotes gastric cancer cell proliferation via binding with HuR and stabilizing FAM83B expression. *Cell Death Dis.* 2018;9:717
74. Li N, Li J, Mi Q, et al. Long non-coding RNA ADAMTS9-AS1 suppresses colorectal cancer by inhibiting the Wnt/β-catenin signalling pathway and is a potential diagnostic biomarker. *J Cell Mol Med.* 2020;24:11318-29
75. Liu H, Zhou T, Wang B, et al. Identification and functional analysis of a potential key lncRNA involved in fat loss of cancer cachexia. *J Cell Biochem.* 2018;119:1679-88
76. Liu G, Yuan D, Sun P, et al. LINC00968 functions as an oncogene in osteosarcoma by activating the PI3K/AKT/mTOR signaling. *J Cell Physiol.* 2018;233:8639-47
77. Gong J, Jin S, Pan X, et al. Identification of long non-coding RNAs for predicting prognosis among patients with thymoma. *Clin Lab.* 2018;64:1193-98
78. Wan J-F, Wan J-Y, Dong C, et al. Linc00324 promotes the progression of papillary thyroid cancer via regulating Notch signaling pathway. *Eur Rev Med Pharmacol Sci.* 2020;24:6818-24
79. Huang EY, Gallegos AM, Richards SM, et al. Surface expression of Notch1 on thymocytes: Correlation with the double-negative to double-positive transition. *J Immunol.* 2003;171:2296-304
80. Ye L, Jin W. Identification of lncRNA-associated competing endogenous RNA networks for occurrence and prognosis of gastric carcinoma. 2020;rs-45596/v1
81. Su Y, Chen Y, Tian Z, et al. lncRNAs classifier to accurately predict the recurrence of thymic epithelial tumors. *Thorac Cancer.* 2020;11:1773-83
82. Luo Y-Y, Wang Z-H, Yu Q, et al. LncRNA-NEAT1 promotes proliferation of T-ALL cells via miR-146b-5p/NOTCH1 signaling pathway. *Pathol Res Pract.* 2020;216:153212