Competing endogenous RNA regulatory network in papillary thyroid carcinoma

SHOUHUA CHEN[1], XIAOBIN FAN[2], HE GU[1], LILI ZHANG[1] and WENHUA ZHAO[3]

Departments of [1]Breast and Thyroid Surgery, [2]Operation Room and [3]Oncology,
Shandong Provincial Qianfoshan Hospital, Shandong University, Jinan, Shandong 250014, P.R. China

Abstract. The present study aimed to screen all types of RNAs involved in the development of papillary thyroid carcinoma (PTC). RNA-sequencing data of PTC and normal samples were used for screening differentially expressed (DE) microRNAs (DE-miRNAs), long non-coding RNAs (DE-lncRNAs) and genes (DEGs). Subsequently, lncRNA-miRNA, miRNA-gene (that is, miRNA-mRNA) and gene-gene interaction pairs were extracted and used to construct regulatory networks. Feature genes in the miRNA-mRNA network were identified by topological analysis and recursive feature elimination analysis. A support vector machine (SVM) classifier was built using 15 feature genes, and its classification effect was validated using two microarray data sets that were downloaded from the Gene Expression Omnibus (GEO) database. In addition, Gene Ontology function and Kyoto Encyclopedia Genes and Genomes pathway enrichment analyses were conducted for genes identified in the ceRNA network. A total of 506 samples, including 447 tumor samples and 59 normal samples, were obtained from The Cancer Genome Atlas (TCGA); 16 DE-lncRNAs, 917 DEGs and 30 DE-miRNAs were screened. The miRNA-mRNA regulatory network comprised 353 nodes and 577 interactions. From these data, 15 feature genes with high predictive precision (>95%) were extracted from the network and were used to form an SVM classifier with an accuracy of 96.05% (486/506) for PTC samples downloaded from TCGA, and accuracies of 96.81 and 98.46% for GEO downloaded data sets. The ceRNA regulatory network comprised 596 lines (or interactions) and 365 nodes. Genes in the ceRNA network were significantly enriched in 'neuron development', 'differentiation', 'neuroactive ligand-receptor interaction', 'metabolism of xenobiotics by cytochrome P450', 'drug metabolism' and 'cytokine-cytokine receptor interaction' pathways. Hox transcript antisense RNA, miRNA-206 and kallikrein-related peptidase 10 were nodes in the ceRNA regulatory network of the selected feature gene, and they may serve import roles in the development of PTC.

Introduction

Thyroid carcinomas originate from follicular or parafollicular thyroid cells (1). Papillary thyroid carcinomas (PTC) are the most frequent histologic type of thyroid carcinoma, and account for 75-85% of reported cases (2). The incidence rate of PTC has increased rapidly in recent years, and 5-10% of patients with PTC experience a particularly aggressive development of the disease (3). The underlying molecular mechanisms of the aggressiveness and mortality in such patients remain unknown. A stratification system that may accurately and effectively identify patients with high mortality risk may have great value.

Bioinformatics-based approaches offer considerable promise for evaluating patients with high risk to specific factors (4). Long non-coding RNAs (lncRNAs) serve crucial roles in cancer biology and cancer pathways at transcriptional, post-transcriptional and epigenetic levels (5). lncRNAs generally lack conserved motifs that may be used for detection and identification; however, current highly developed sequencing technologies have made it easier to identify lncRNAs. By combining lncRNAs with known miRNAs, chromatin interactions and RNA-coding proteins, bioinformatics tools are now able to recognize and functionally annotate a large number of lncRNAs (6). Micro (mi)RNAs are post-transcriptional gene regulators that lead to translational repression, gene silencing and mRNA degradation (7). It has been reported that lncRNAs may interact with miRNAs and modulate their regulatory roles by acting as decoys (8). Computational prediction is a useful tool for the construction of miRNA-lncRNA interaction networks, and is a widespread method (9). Multiple types of non-coding RNAs (including lncRNA, circRNAs and pseudogenes) along with protein-coding mRNAs function as key competing endogenous (ce)RNAs to regulate the expression levels of other mRNAs in mammalian cells (10,11). A regulatory role of ceRNA crosstalk with cancer-associated genes has been reported (12). For example, the lncRNA hepatocellular carcinoma upregulated lncRNA was demonstrated to bind to miRNA (miR)-372 and serve a regulatory function (13).

Although the crucial roles of miRNAs, lncRNAs and ceRNAs in cell-fate determination and in various human diseases have been implicated, the regulatory interaction networks among them remain unknown. In the present study, a ceRNA regulatory network comprising miRNAs, lncRNAs and mRNAs in PTC samples was constructed, and the important nodes in the networks were extracted to establish a support vector machine (SVM) classifier.

Materials and methods

Microarray data sets and data preprocessing. miRNA, mRNA and lncRNA sequencing data of PTC samples were downloaded from The Cancer Genome Atlas (TCGA) through the National Cancer Institute Genomic Data Commons data portal (https://gdc-portal.nci.nih.gov) on 20 June, 2016. The TCGA is a data-rich resource for biological discovery that collects information from ~10,000 patient samples together with clinicopathological information across >30 types of human cancer. The present study reviewed a number of previous reports that have used data from TCGA (14,15) and, thus, have selected sequencing data based on TCGA, with the RNASeqV2 platform used for mRNA and lncRNA sequencing data, and with HiSeq2000 used for miRNA sequencing data. Following barcode matching, the mRNA, miRNA and lncRNAs were annotated using the information recorded by the Human Genome Organization Gene Nomenclature Committee (http://www.genenames.org). In addition, two gene expression data sets, GSE33630 and GSE60542, were downloaded from the Gene Expression Omnibus (GEO) database. These microarray data sets were based on the GPL570 (HG-U133_Plus_2) Affymetrix Human Genome U133 Plus 2.0 Array platform. GSE33630 contained 49 PTC samples and 45 normal sample, while GSE60542 contained 33 PTC samples and 32 normal samples. The 'oligo' package in R was used to conduct the background correction and normalization as previously described (16,17).

Differentially expressed (DE)-lncRNA, -mRNA and -miRNA. DE-lncRNAs, DE-mRNAs (or DE-genes; DEGs) and DE-miRNAs between the PTC samples and the normal samples were screened using the 'edgeR' version 2.4 and 'multtest' version 1.6 packages in R (18,19). DE-lncRNAs, DE-miRNAs and DEGs were selected based on the same cut-off value of false discovery rate (FDR) adjusted by Benjamini-Hochberg method, as previously described (14,15); |log fold change (FC)| >1 and FDR <0.05 were used as cut-off criteria. Heatmaps were generated by hierarchical clustering analysis and used to identify the differences in gene expression levels between the PTC samples and the normal samples.

miRNA-regulated lncRNAs and genes. The regulatory pairs between miRNAs and lncRNAs were obtained using miRcode version 11 (http://www.mircode.org) and the starBase version 2.0 (http://starbase.sysu.edu.cn) database; regulatory pairs of miRNAs-target genes were collected from miRTarBase version 7.0 (http://mirtarbase.mbc.nctu.edu.tw). Gene-gene interaction pairs were obtained from the Biological General Repository for Interaction Datasets version 3.4 (http://thebiogrid.org), The Human Protein Reference Database

version 9.0 (http://www.hprd.org) and the Database of Interacting Proteins version 2.5 (http://dip.doe-mbi.ucla.edu). All interaction or regulatory pairs among genes, miRNAs and lncRNAs were first collected from the related databases and then the DEGs, DE-lncRNAs and DE-miRNAs were matched with them to obtain their connections, an lncRNA-miRNA regulatory network and a miRNA-mRNA network. These two networks were integrated into a comprehensive ceRNA regulatory network to demonstrate the interactions of DE-miRNAs with DE-lncRNAs or DEGs. These networks were visualized using Cytoscape software version 3.6 (http://www.cytoscape.org/).

Feature genes in the miRNA-mRNA regulatory network. Topological structure analysis was conducted to identify the feature genes in the miRNA-mRNA regulatory network. There are four common topological properties, including degree, closeness, betweenness and PageRank, which may reflect the centrality. 'Betweenness centrality' (BC) is a classical measure that quantifies the importance of a vertex within a graph, because it is a ratio of the shortest paths between vertex pairs that pass through the vertex of interest (20). In a number of previous studies, gene signatures have been identified largely based on the BC (21,22); therefore, the present study used BC as the representative topological parameter to indicate node centrality. The BC of genes in the network was calculated using the following formula:

$$BC = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Where s, t and v indicate different nodes in the network, in which s and t are vertex pairs. σ_{st} is the shortest path numbers (the total number of the shortest paths) from node s to node t, and $\sigma_{st}(v)$ is the path numbers passes node v from s to t. BC values ranged between 0 and 1, and values that were closer to 1 indicated a higher node (gene) degree in the network.

The number of DEGs in the miRNA-mRNA regulatory network is larger than that of the ceRNA network. Therefore, the feature genes were preliminarily selected in the miRNA-mRNA regulatory network to avoid missing important feature genes. The nodes in the miRNA-mRNA regulatory network with the top 100 BC values were preliminarily selected as the feature genes that may display significantly different expressions between PTC and normal tissue samples. Clustering analysis was conducted using these feature genes via clusterStab package version 3.4 in R.

SVM classifier construction for different samples. The selected feature genes were used to conduct recursive feature elimination (RFE) analysis (23), which could distinguish different samples in the training dataset. The feature genes were subsequently utilized to construct the SVM classifier. The robustness and portability of the classifier were determined using the GSE33630 and GSE60542 microarray data. The classification effect was evaluated using correct rate, sensitivity, specificity, positive predictive value, negative predictive value and area under receiver operating characteristic curve.

Function and pathway enrichment. Gene ontology (GO) function enrichment and Kyoto Encyclopedia of Genes

and Genomes (KEGG) pathway enrichment analyses were performed for the integrated ceRNA regulatory network using the Database for Annotation, Visualization and Integrated Discovery version 6.8 (https://david.ncifcrf.gov). Fisher exact test was applied for the enrichment using the following formula:

$$p = \sum_{i=0}^{x-1} \frac{\binom{M}{i}\binom{N-M}{K-i}}{\binom{N}{K}}$$

Where N is the total number of all human genes, M is the number of genes in enriched pathways and K is number of DEGs.

Results

Microarray data sets and preprocessing. A total of 506 samples comprising 447 PTC samples and 59 normal samples were obtained from TCGA, from which a total of 877 lncRNAs, 1,046 miRNAs and 1,8348 mRNAs were annotated from the sequencing data. The microarray data sets GSE33630 and GSE60542 were downloaded from the GEO database, which included 45 and 32 normal samples, as well as 49 and 33 PTC samples, respectively.

DE-lncRNAs, DEGs, and DE-miRNAs. Using thresholds of FDR <0.05 and |logFC| >1, 16 DE-lncRNAs were identified (Table I), as well as 917 DEGs and 30 DE-miRNAs. Clustering analysis of DE-lncRNAs, DE-miRNAs and DEGs suggested that tumor samples may be distinguished from normal samples based on the expression levels of DE-lncRNAs, DE-miRNAs and DEGs (Fig. 1).

lncRNA-miRNA, miRNA-mRNA and gene-gene regulatory networks. A total of 246 lncRNA-miRNA regulatory pairs were screened from miRcode and starBase, including 19 DE-miRNAs (Table II). The constructed miRNA-lncRNA regulatory network comprised 8 DE-lncRNAs and 7 DE-miRNAs (Fig. 2). A total of 13,159 target genes of the 30 DE-miRNAs were obtained from miRTarBase database; 406 DEGs were among these target genes. The miRNA-mRNA regulatory pairs were integrated with gene-gene interaction pairs to construct a comprehensive miRNA-mRNA regulatory network (Fig. 3). The network comprised 353 nodes (miRNAs or genes) and 577 lines (interactions).

Feature genes extraction. Degree distribution of the nodes demonstrated that the constructed miRNA-regulatory network followed the characteristics of scale-free networks. A total of 100 feature genes were screened according to the BC values. The clustering analysis indicated that tumor samples may be distinguished from the normal samples based on the expression level of the top 100 feature genes ranked by BC value (Fig. 4).

SVM classifier. A total of 15 feature genes with high predictive precision (>95%) were obtained following refinement by the RFE algorithm, as demonstrated in Fig. 5A, the accuracy was >95% when there were 15 feature genes. Basic information

Table I. Differentially expressed lncRNAs between papillary thyroid carcinomas samples and normal tissue samples.

lncRNA	logFC	P-value	FDR
NHEG1	-2.63948	2.87×10^{-02}	9.36×10^{-02}
FAM27B	-2.49586	2.48×10^{-05}	2.57×10^{-04}
C12orf77	-2.04917	2.05×10^{-07}	3.01×10^{-06}
RNU11	-1.85277	1.24×10^{-06}	1.45×10^{-05}
TERC	-1.66684	9.66×10^{-22}	1.70×10^{-19}
PWRN2	-1.58396	9.41×10^{-04}	5.52×10^{-03}
DIO3OS	-1.45712	7.29×10^{-08}	1.28×10^{-06}
TCL6	-1.19462	1.84×10^{-11}	8.10×10^{-10}
SMCR5	-1.08884	4.66×10^{-10}	1.37×10^{-08}
MYCNOS	-1.0739	1.65×10^{-10}	5.81×10^{-09}
HYMAI	-1.05663	2.20×10^{-08}	4.84×10^{-07}
C9orf170	-1.03293	1.43×10^{-07}	2.29×10^{-06}
SFTA1P	1.325098	4.13×10^{-18}	3.63×10^{-16}
C16orf82	1.437648	6.26×10^{-03}	2.75×10^{-02}
UCA1	2.279467	3.41×10^{-14}	2.00×10^{-12}
HOTAIR	2.707475	2.77×10^{-03}	1.43×10^{-02}

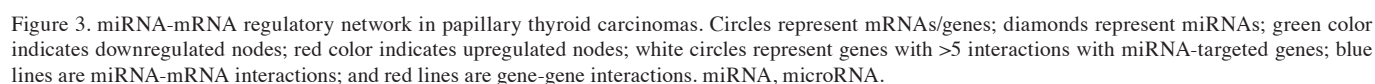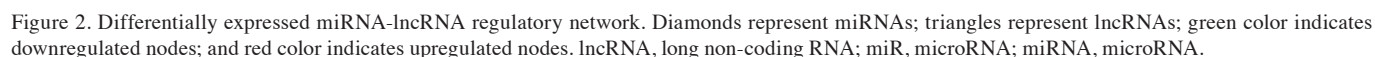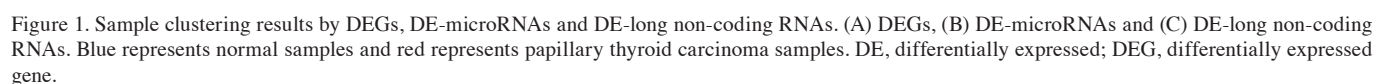FC, fold change; FDR, false discovery rate; lncRNA, long non-coding RNA.

Table II. Differentially expressed lncRNAs that are regulated by differentially expressed microRNAs.

lncRNA	microRNA
C16orf82	hsa-miR-18b and hsa-miR-876
DIO3OS	hsa-miR-7, hsa-miR-206 and hsa-miR-18b
HOTAIR	hsa-miR-9-3 and hsa-miR-206
MYCNOS	hsa-miR-551b, hsa-miR-18b and hsa-miR-7
PWRN2	hsa-miR-9-3 and hsa-miR-206
SFTA1P	hsa-miR-206 and hsa-miR-9-3
TCL6	hsa-miR-9-3, hsa-miR-18b, hsa-miR-490 and hsa-miR-7
UCA1	hsa-miR-206

lncRNA, long non-coding RNA; miR, microRNA.

on these 15 genes is provided in Table III. The SVM classifier constructed by the 15 genes demonstrated an accuracy of 96.05% (486/506; Fig. 5B).

In the validation test, the SVM classifier was able to correctly distinguish 44 out of 45 normal samples in the GSE33630 data set, and 47 out of 49 PTC samples, with an accuracy of 96.81% (Fig. 6A). For the GSE60542 data set, SVM classifier could correctly distinguish all the 32 normal samples and 32 out of 33 PTC samples, with the accuracy of 98.46% (Fig. 6B). As demonstrated in Table IV, the SVM classifier demonstrated a satisfied classifying characteristics (Table IV).

Figure 1. Sample clustering results by DEGs, DE-microRNAs and DE-long non-coding RNAs. (A) DEGs, (B) DE-microRNAs and (C) DE-long non-coding RNAs. Blue represents normal samples and red represents papillary thyroid carcinoma samples. DE, differentially expressed; DEG, differentially expressed gene.

Figure 2. Differentially expressed miRNA-lncRNA regulatory network. Diamonds represent miRNAs; triangles represent lncRNAs; green color indicates downregulated nodes; and red color indicates upregulated nodes. lncRNA, long non-coding RNA; miR, microRNA; miRNA, microRNA.

Figure 3. miRNA-mRNA regulatory network in papillary thyroid carcinomas. Circles represent mRNAs/genes; diamonds represent miRNAs; green color indicates downregulated nodes; red color indicates upregulated nodes; white circles represent genes with >5 interactions with miRNA-targeted genes; blue lines are miRNA-mRNA interactions; and red lines are gene-gene interactions. miRNA, microRNA.
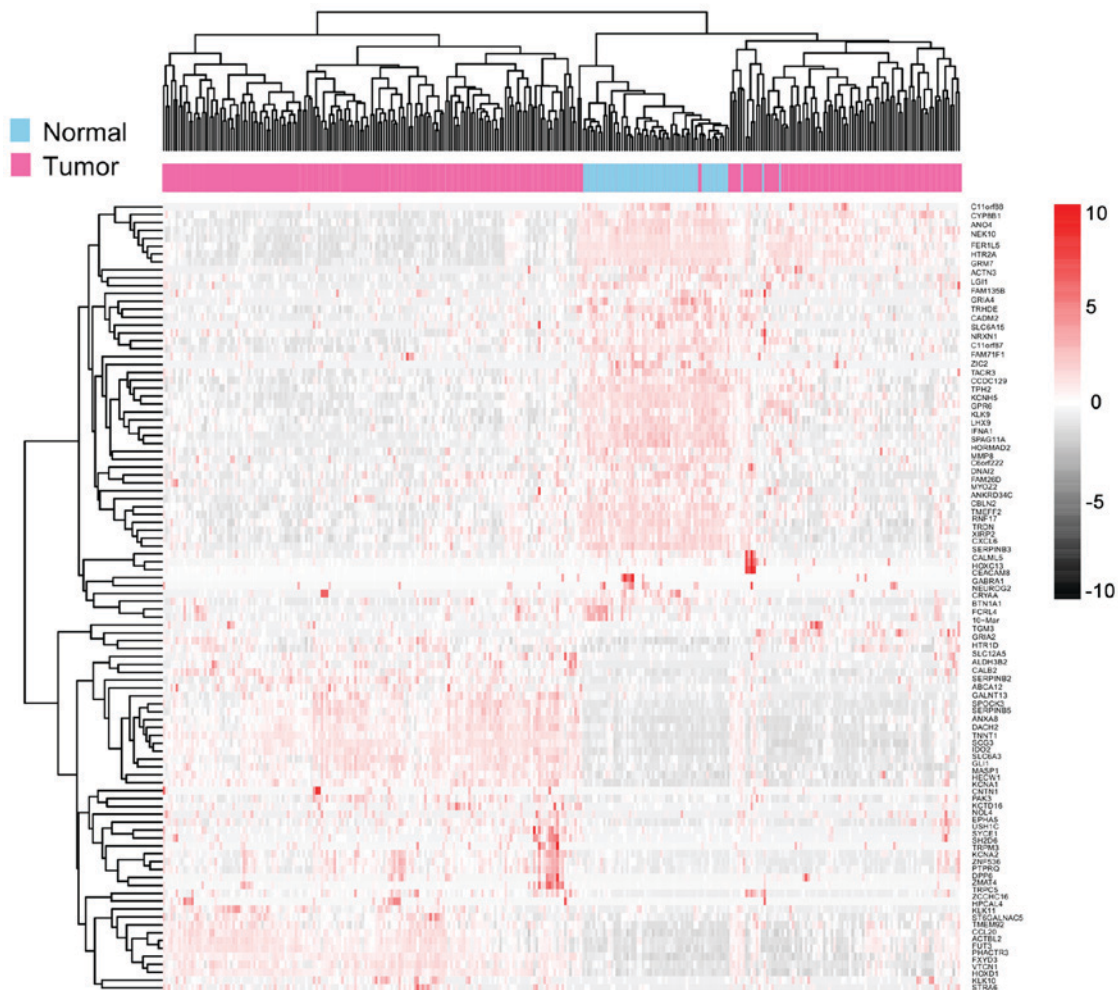
Figure 4. Clustering results of the feature genes with top 100 betweenness centrality values of different samples in the TCGA downloaded dataset.
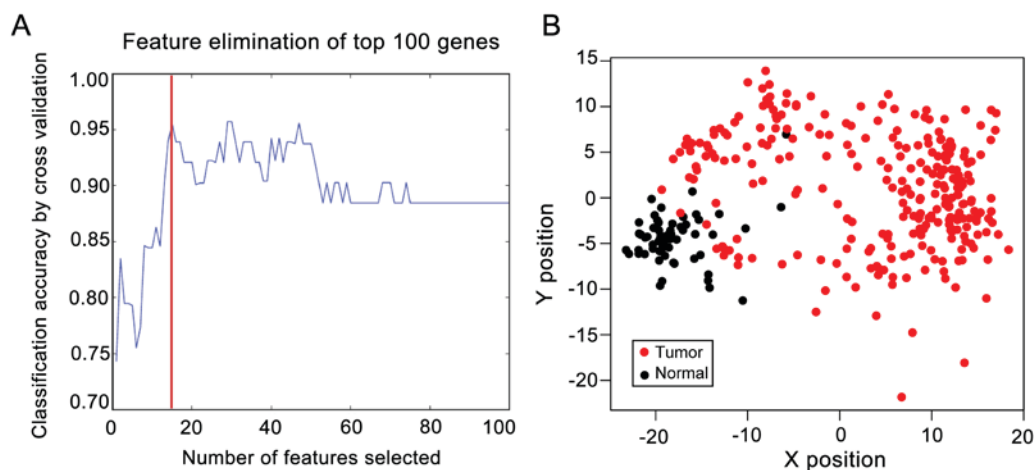

Figure 5. Classification effects of the extracted feature genes on PTC samples. (A) Feature elimination of all 100 feature genes. The vertical red line shows that when there were 15 genes, the accuracy could reach >95%. (B) Sample distribution distinguished by the support vector machine classifier. PTC, papillary thyroid carcinomas.

Function and pathway enrichment. The ceRNA regulatory network comprised 365 nodes and 596 lines (Fig. 7). Genes in the ceRNA network were significantly enriched in 20 GO function terms, most of which related to neuron development

and differentiation (Table V). The present study focused on the Biological Process (BP) ontologies in GO as the purpose of this current study is the RNA regulatory network of PTC. In addition, genes in the ceRNA regulatory network were

Table III. 15 feature genes and related DE-microRNAs.

Gene	logFC	P-value	FDR	BC	Related DE-microRNA
DACH2	1.1311	3.99×10^{-13}	1.00×10^{-11}	0.0247	hsa-miR-18b and hsa-miR-206
DPP6	-1.0414	1.09×10^{-44}	2.17×10^{-42}	0.0879	hsa-miR-187 and hsa-miR-506
GALNT13	1.0542	3.72×10^{-08}	5.42×10^{-07}	0.1683	hsa-miR-187, hsa-miR-18b and hsa-miR-873
GRIA4	-1.1509	1.06×10^{-10}	2.07×10^{-09}	0.3746	hsa-miR-187, hsa-miR-18b, hsa-miR-506 and hsa-miR-873
GRM7	-1.1511	1.84×10^{-09}	3.10×10^{-08}	0.0856	hsa-miR-206 and hsa-miR-506
HECW1	-1.3015	8.89×10^{-31}	8.19×10^{-29}	0.0603	hsa-miR-206 and hsa-miR-506
HTR1D	1.2698	2.92×10^{-14}	8.30×10^{-13}	0.1310	hsa-miR-18b, hsa-miR-206 and hsa-miR-873
KLK10	1.2011	9.98×10^{-42}	1.66×10^{-39}	0.0247	hsa-miR-18b and hsa-miR-206
SH2D6	-1.3458	3.08×10^{-35}	3.91×10^{-33}	0.0247	hsa-miR-18b and hsa-miR-206
STRA6	1.4452	5.51×10^{-48}	1.37×10^{-45}	0.0614	hsa-miR-506 and hsa-miR-873
TRHDE	-1.1062	2.25×10^{-12}	5.21×10^{-11}	0.0357	hsa-miR-187 and hsa-miR-206
TRPC5	2.0281	1.86×10^{-58}	1.03×10^{-55}	0.0942	hsa-miR-18b, hsa-miR-206 and hsa-miR-506
TRPM3	-1.3570	4.04×10^{-32}	4.07×10^{-30}	0.0345	hsa-miR-206 and hsa-miR-506
VTCN1	1.9145	3.25×10^{-38}	4.64×10^{-36}	0.0350	hsa-miR-18b and hsa-miR-506
ZCCHC16	1.7752	5.37×10^{-49}	1.39×10^{-46}	0.0247	hsa-miR-18b and hsa-miR-206

BC, betweenness centrality; DE, differentially expressed; FC, fold change; FDR, false discover rate; miR, microRNA.
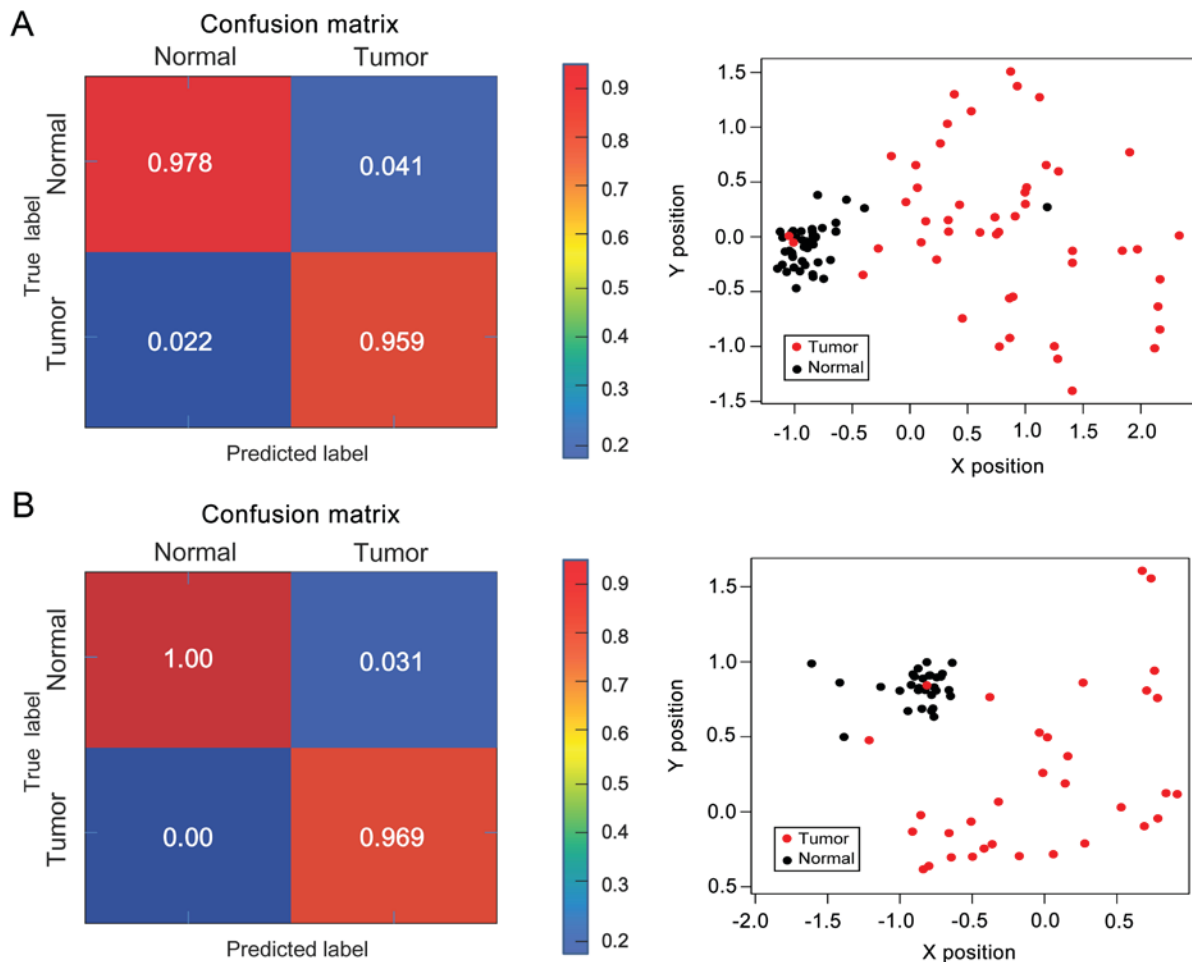

Figure 6. Classification confusion matrix and scatter plot of support vector machine classifier on GEO downloaded datasets. (A) The left fig. illustrates the accuracy of the SVM classifier to classify the normal and tumor sample in GSE33630 dataset; The right fig. showed the distribution of the samples in GSE33630 dataset. (B) The left Figure showed the accuracy of the SVM classifier to classify the normal and tumor sample in GSE60542 dataset; The right Figure showed the distribution of the samples in GSE60542 dataset.

Table IV. Classifying characteristics of the support vector machine classifier on all downloaded data sets.

Data set	n	Correct rate	Sensitivity	Specificity	PPV	NPV	AUROC
TCGA	506	0.961	0.9619	0.949	0.993	0.921	0.998
GSE33630	94	0.968	0.959	0.978	0.979	0.956	0.991
GSE60542	65	0.985	0.969	1.000	1.000	0.969	0.973

AUROC, area under receiver operating characteristic curve; TCGA, The Cancer Genome Atlas; NPV, negative predictive value; PPV, positive predictive value.
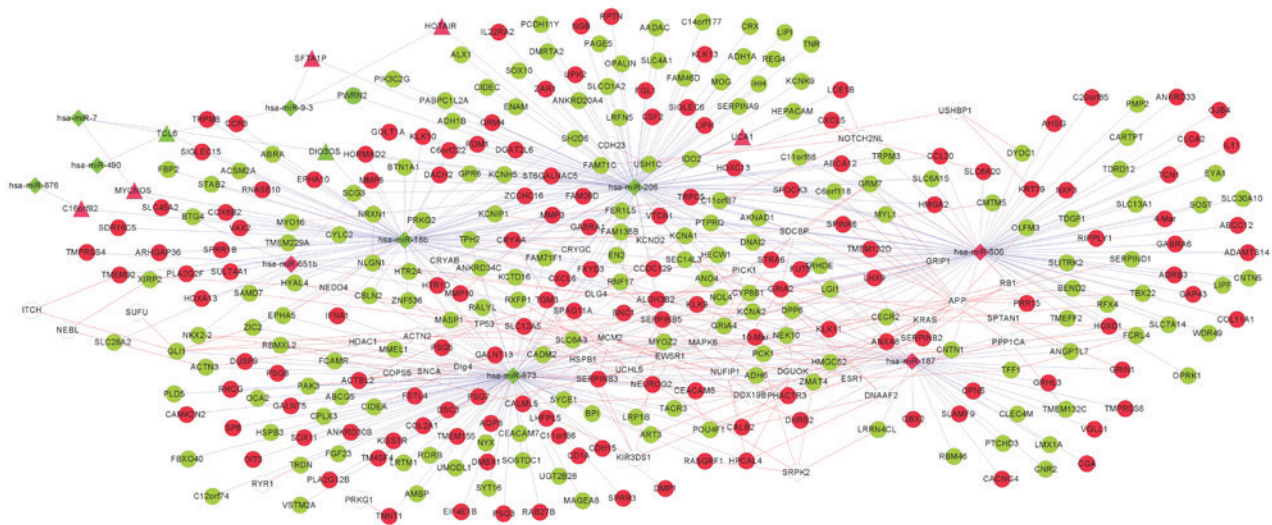

Figure 7. ceRNA regulatory network of all DE-miRNAs, DE-lncRNAs and DEGs. Circles represent mRNAs (genes); diamonds represent miRNAs; triangles represent lncRNAs. Green nodes represent downregulated expression; red nodes represent upregulated expression; white circles represent genes with >5 interactions to miRNA targeted genes. Blue lines indicate miRNA-gene and miRNA-lncRNA interactions; red lines indicate gene-gene interactions. DE, differentially expressed; DEG, differentially expressed gene; lncRNA, long non-coding RNA; miR, microRNA; miRNA, microRNA.
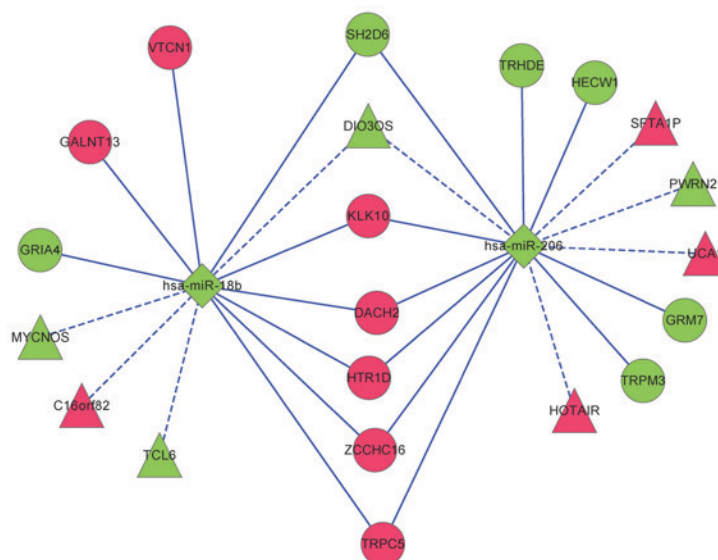

Figure 8. ceRNA regulatory network of DE-miRNAs, DE-lncRNAs and selected feature genes. The feature genes were those used for constructing the SVM classifier. Circles represent mRNAs (genes); diamonds represent miRNAs; triangles represent lncRNAs. Green nodes indicate downregulated expression; red nodes indicate upregulated expression. Solid lines represent miRNA-mRNA interactions; dotted lines represent lncRNA-miRNA interactions. DE, differentially expressed; lncRNA, long non-coding RNA; miR, microRNA; miRNA, microRNA.

significantly enriched in four KEGG pathways, including 'neuroactive ligand-receptor interaction', 'metabolism of xenobiotics by cytochrome P450', 'drug metabolism' and 'cytokine-cytokine receptor interaction pathway'.

Table V. Enriched functions for the feature genes.

Term	Count	P-value	Genes
GO:0007267~cell-cell signaling	25	4.05×10^{-5}	*CPLX3, CGA, CXCL5, OPRK1, KCNA1, CXCL6, KCNIP1, GLI1, IL11, CCL20, HTR1D, IHH, TRHDE, GABRA6, SLC12A5, GRIN1, NLGN1, NRXN1, GRIA4, GRM4, GRIA2, SIGLEC6, GRM7, CARTPT* and *HTR2A*
GO:0007268~synaptic transmission	16	1.04×10^{-4}	*CPLX3, GABRA6, OPRK1, SLC12A5, KCNA1, GRIN1, NLGN1, NRXN1, GRIA4, KCNIP1, GRM4, GRIA2, GRM7, CARTPT, HTR1D* and *HTR2A*
GO:0019226~transmission of nerve impulse	17	1.87×10^{-4}	*CPLX3, OPRK1, GABRA6, SLC12A5, KCNA1, GRIN1, NLGN1, CACNG4, NRXN1, GRIA4, KCNIP1, GRM4, GRIA2, GRM7, CARTPT, HTR1D* and *HTR2A*
GO:0051969~regulation of transmission of nerve impulse	10	6.47×10^{-4}	*CPLX3, KISS1R, GRIA2, RASGRF1, GRIN1, NLGN1, CARTPT, GRIA4, LGI1* and *HTR2A*
GO:0006811~ion transport	26	6.79×10^{-4}	*FXYD3, KCNA2, KCNA1, KCNIP1, SLCO1A2, KCNK9, SLC4A1, ANO4, OCA2, TRPM3, CLCA2, GABRA1, TRPM8, TRPC5, GABRA6, GRIN1, SLC12A5, CACNG4, GRIA4, TCN1, RHCG, GRIA2, SLC13A1, KCTD16, SLC30A10* and *KCNH5*
GO:0031644~regulation of neurological system process	10	8.62×10^{-4}	*CPLX3, KISS1R, GRIA2, RASGRF1, GRIN1, NLGN1, CARTPT, GRIA4, LGI1* and *HTR2A*
GO:0043062~extracellular structure organization	10	1.35×10^{-3}	*ADAMTS14, RXFP1, TNR, DMP1, NLGN1, POU4F1, COL2A1, NRXN1, COL11A1* and *TMPRSS6*
GO:0007610~behavior	18	1.65×10^{-3}	*DMBX1, CXCL5, OPRK1, SLC6A3, GRIN1, CXCL6, GRM4, KISS1R, CCL20, RASGRF1, CCR3, GRM7, CNR2, CARTPT, POU4F1, SERPIND1, HTR2A* and *CMTM5*
GO:0050877~neurological system process	34	2.03×10^{-3}	*SLC45A2, CPLX3, SLC6A3, OPRK1, KCNA1, COL2A1, RORB, KCNIP1, CRX, POU4F1, HTR1D, COL11A1, NYX, TRPM8, CRYAA, CNTN5, GABRA6, GRIN1, SLC12A5, NLGN1, CACNG4, VAX2, GRIA4, NRXN1, OPN5, GRM4, EYA1, LHFPL5, GRIA2, RASGRF1, GRM7, USH1C, CARTPT* and *HTR2A*
GO:0048666~neuron development	14	3.56×10^{-3}	*NEUROG2, VAX2, RORB, NRXN1, LMX1A, EN2, SLITRK2, LHFPL5, RASGRF1, TNR, GBX2, POU4F1, OLFM3* and *GAP43*
GO:0007423~sensory organ development	11	4.17×10^{-3}	*EYA1, LHFPL5, CRYAA, HOXC13, GBX2, RORB, VAX2, COL2A1, OLFM3, COL11A1* and *PTPRQ*
GO:0007389~pattern specification process	12	4.22×10^{-3}	*EYA1, HOXC13, SOSTDC1, TDGF1, GBX2, RIPPLY1, VAX2, GRHL3, SP8, GLI1, IHH* and *ALX1*
GO:0044057~regulation of system process	13	4.55×10^{-3}	*CPLX3, TACR3, GRIN1, NLGN1, GRIA4, KISS1R, ABCG5, TNNT1, GRIA2, RASGRF1, CARTPT, LGI1* and *HTR2A*
GO:0030182~neuron differentiation	16	5.17×10^{-3}	*NEUROG2, VAX2, RORB, NRXN1, LMX1A, EN2, SLITRK2, LHFPL5, RASGRF1, TNR, BTG4, GBX2, POU4F1, OLFM3, NKX2-2* and *GAP43*
GO:0001501~skeletal system development	13	5.76×10^{-3}	*EYA1, SOST, HOXA13, ENAM, DMP1, MMP8, COL2A1, HOXD1, COL11A1, GLI1, IHH, AHSG* and *ALX1*
GO:0051240~positive regulation of multicellular organismal process	11	6.39×10^{-3}	*ADRB3, KISS1R, TACR3, GRIA2, SLC6A3, CARTPT, GRIA4, LGI1, HTR2A, AHSG* and *CRX*
GO:0000904~cell morphogenesis involved in differentiation	11	6.39×10^{-3}	*SLITRK2, LHFPL5, CRYAA, TNR, GBX2, VAX2, NEUROG2, POU4F1, NRXN1, LMX1A* and *GAP43*
GO:0048667~cell morphogenesis involved in neuron differentiatio	10	7.03×10^{-3}	*SLITRK2, LHFPL5, TNR, GBX2, VAX2, NEUROG2, POU4F1, NRXN1, LMX1A* and *GAP43*
GO:0007601~visual perception	10	8.63×10^{-3}	*OPN5, SLC45A2, CRYAA, USH1C, RORB, VAX2, COL2A1, COL11A1, NYX* and *CRX*
GO:0050953~sensory perception of light stimulus	10	8.63×10^{-3}	*OPN5, SLC45A2, CRYAA, USH1C, RORB, VAX2, COL2A1, COL11A1, NYX* and *CRX*

GO, gene ontology.

ceRNA regulatory network of selected feature genes. The ceRNA regulatory network of the selected feature genes was constructed (Fig. 8). The network provided valuable information for reveal potential RNA regulation mechanism in PTC. The upregulated expression of lncRNA HOX transcript antisense RNA (HOTAIR), surfactant associated 1, pseudogene (SFTA1P), urothelial cancer associated 1 (UCA1) or the downregulated expression of Prader-Willi region non-protein coding RNA 2 (PWRN2) and DIO3 opposite strand/antisense RNA (DIO3OS) result in the downregulation of miR-206 expression, which in turn may subsequently lead to the upregulated gene expression of kallikrein-related peptidase 10 (*KLK10*), dachshund family transcription factor 2 (*DACH2*), 5-hydroxytryptamine receptor 1D (*HTR1D*), zinc-finger CCHC-type containing 16 (*ZCCHC16*) and transient receptor potential cation channel subfamily C member 5 (*TRPC5*). These five genes were also revealed to be potentially significantly upregulated by the decreased expression level of miR-18b, which was induced by the upregulation of C16orf18 and the downregulation of MYCN opposite strand (MYCNOS), T cell leukemia/lymphoma 6 (TCL6) and DIO3OS.

Discussion

Using the RNA-seq data downloaded from TCGA, DE-miRNA, DEGs and DE-lncRNAs in PTC samples were identified. The identified factors were subjected to known interaction databases to obtain the lncRNA-miRNA and the miRNA-mRNA regulatory networks, as well as a comprehensive ceRNA network. By conducting BC value calculation and RFE algorithm, 15 feature genes were screened. The SVM classifier constructed by these 15 feature genes revealed an accuracy of 96.05% (486/506) for the PTC samples downloaded from TCGA, and high accuracy (96.81 and 98.46%) for the PTC samples downloaded from GEO. The lncRNA-miRNA-mRNA interaction of HOTAIR-miR-206-*KLK10* (or HOTAIR-miR-206-*DACH2*, *-HTR1D*, *-ZCCHC16* or *-TRPC5)* is some of the interactions in the ceRNA regulatory network of the selected PTC feature genes; the five genes aforementioned may also be significantly upregulated by a decrease in the expression level of miR-18b. It has been reported that miR-18b and miR-206 induce cell cycle arrest and inhibit estrogen-induced proliferation, with inhibition levels comparable to those achieved by estrogen receptor (*ER*)*α* small interfering RNA (8). Therefore, it may be concluded that miR-18b and miR-206 may serve a similar regulatory role for their target genes. In addition, genes in the comprehensive ceRNA regulatory network were mainly enriched in neuronal development and differentiation related functions, and in four pathways.

To date, the functions and roles of lncRNAs in thyroid remain to be further explored. HOTAIR is a tumor-related lncRNA, and its overexpression is believed to be associated with poor prognosis and increased invasiveness in cancer (24). The expression status of HOTAIR may be used to predict metastasis and mortality in patients with breast tumors and hepatocellular carcinoma (25,26). However, the connections between HOTAIR and PTC have not been reported previously. The inhibition of HOTAIR expression may induce the downregulation of neuronal growth-related genes (27); notably,

feature genes in the comprehensive ceRNA regulatory network were related to neuron development functions. Therefore, the present study hypothesized that HOTAIR may be involved in the development of PTC by interrupting neuronal growth.

Alterations in miRNA expression levels have been identified in thyroid tumors, which indicated their involvement in thyroid carcinoma. For example, the expression of miR-146b-5p was previously revealed to be upregulated, and miR-335 expression downregulated in PTC samples (28). There are many reports about the roles of miRNA in the diagnosis and prognosis of PTC (29-31). miR-206 is a member of the skeletal muscle specific miRNA (myomiR) family, which exert important roles in the pathogenesis of various types of cancers (32). miR-206 was reported to be significantly deregulated in PTC tissues (33). miR-1 is another member in the myomiR family, and was revealed to be downregulated in aggressive PTC and associated with thyroid cell growth and migration (34). miR-1 functions as a tumor suppressor in thyroid carcinoma, targeting C-X-C motif chemokine receptor 4, cyclin D2 and stromal cell-derived factor 1α (35). Therefore, the present study hypothesized that miR-206 may serve similar roles with miR-1 in PTC.

KLK10 serves an important role in the tumor microenvironment, invasion and angiogenesis (36). KLK10 regulates normal cell growth and probably functions as a tumor suppressor (37,38). Expression of KLK10 is increased in PTC samples, and there is a specific hypomethylation of KLK10 in *BRAF*-mutated PTC (39).

In conclusion, ceRNA regulatory network is a good way to identify the signature RNAs for cancers. HOTAIR, miR-206, and KLK10 are potential gene markers for targeting therapy and diagnosis of PTC.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

WZ conceived and designed the study. SC performed the bioinformatics analyses and was a major contributor in writing the manuscript. XF assisted in bioinformatics analysis. HG and LZ selected and downloaded the datasets. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. PDQ Cancer Information Summaries: Health Professional Version. In: PDQ Cancer Information Summaries, Bethesda (MD), 2002.
2. Sosa JA and Udelsman R: Papillary thyroid cancer. Surg Oncol Clin N Am 15: 585-601, 2006.
3. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M, *et al*: 2015 American Thyroid Association Management Guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The American Thyroid Association Guidelines task force on thyroid nodules and differentiated thyroid cancer. Thyroid 26: 1-133, 2016.
4. Xing M, Haugen BR and Schlumberger M: Progress in molecular-based management of differentiated thyroid cancer. Lancet 381: 1058-1069, 2013.
5. Song X, Wang B, Bromberg M, Hu Z, Konigsberg W and Garen A: Retroviral-mediated transmission of a mouse VL30 RNA to human melanoma cells promotes metastasis in an immunodeficient mouse model. Proc Natl Acad Sci USA 99: 6269-6273, 2002.
6. Yotsukura S, duVerle D, Hancock T, Natsume-Kitatani Y and Mamitsuka H: Computational recognition for long non-coding RNA (lncRNA): Software and databases. Brief Bioinform 18: 9-27, 2017.
7. Bartel DP: MicroRNAs: Target recognition and regulatory functions. Cell 136: 215-233, 2009.
8. Salmena L, Poliseno L, Tay Y, Kats L and Pandolfi PP: A ceRNA hypothesis: The Rosetta Stone of a hidden RNA language? Cell 146: 353-358, 2011.
9. Jeggari A, Marks DS and Larsson E: miRcode: A map of putative microRNA target sites in the long non-coding transcriptome. Bioinformatics 28: 2062-2063, 2012.
10. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, *et al*: Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495: 333-338, 2013.
11. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK and Kjems J: Natural RNA circles function as efficient microRNA sponges. Nature 495: 384-388, 2013.
12. Karreth FA, Tay Y, Perna D, Ala U, Tan SM, Rust AG, DeNicola G, Webster KA, Weiss D, Perez-Mancera PA, *et al*: In vivo identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. Cell 147: 382-395, 2011.
13. Wang J, Liu X, Wu H, Ni P, Gu Z, Qiao Y, Chen N, Sun F and Fan Q: CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res 38: 5366-5383, 2010.
14. Wang H, Niu L, Jiang S, Zhai J, Wang P, Kong F and Jin X: Comprehensive analysis of aberrantly expressed profiles of lncRNAs and miRNAs with associated ceRNA network in muscle-invasive bladder cancer. Oncotarget 7: 86174-86185, 2016.
15. Zhang J, Fan D, Jian Z, Chen GG and Lai PB: Cancer specific long noncoding RNAs show differential expression patterns and competing endogenous RNA potential in hepatocellular carcinoma. PLoS One 10: e0141042, 2015.
16. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D and Altman RB: Missing value estimation methods for DNA microarrays. Bioinformatics 17: 520-525, 2001.
17. Rao Y, Lee Y, Jarjoura D, Ruppert AS, Liu CG, Hsu JC and Hagan JP: A comparison of normalization techniques for microRNA microarray data. Stat Appl Genet Mol Biol 7: Article22, 2008.
18. Robinson MD, McCarthy DJ and Smyth GK: edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140, 2010.
19. Li D and Dye TD: Power and stability properties of resampling-based multiple testing procedures with applications to gene oncology studies. Comput Math Methods Med 2013: 610297, 2013.
20. Kivimäki I, Lebichot B, Saramäki J and Saerens M: Two betweenness centrality measures based on Randomized Shortest Paths. Sci Rep 6: 19668, 2016.
21. Breitkreutz D, Hlatky L, Rietman E and Tuszynski JA: Molecular signaling network complexity is correlated with cancer patient survivability. Proc Natl Acad Sci USA 109: 9209-9212, 2012.
22. van den Heuvel MP and Sporns O: Network hubs in the human brain. Trends Cogn Sci 17: 683-696, 2013.
23. Baur B and Bozdag S: A feature selection algorithm to compute gene centric methylation from probe level methylation data. PLoS One 11: e0148977, 2016.
24. Berrondo C, Flax J, Kucherov V, Siebert A, Osinski T, Rosenberg A, Fucile C, Richheimer S and Beckham CJ: Expression of the long non-coding RNA HOTAIR correlates with disease progression in bladder cancer and is contained in bladder cancer patient urinary exosomes. PLoS One 11: e0147236, 2016.
25. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, *et al*: Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464: 1071-1076, 2010.
26. Geng YJ, Xie SL, Li Q, Ma J and Wang GY: Large intervening non-coding RNA HOTAIR is associated with hepatocellular carcinoma progression. J Int Med Res 39: 2119-2128, 2011.
27. Ono H, Motoi N, Nagano H, Miyauchi E, Ushijima M, Matsuura M, Okumura S, Nishio M, Hirose T, Inase N, *et al*: Long noncoding RNA HOTAIR is relevant to cellular proliferation, invasiveness, and clinical relapse in small-cell lung cancer. Cancer Med 3: 632-642, 2014.
28. Zhang J, Liu Y, Liu Z, Wang XM, Yin DT, Zheng LL, Zhang DY and Lu XB: Differential expression profiling and functional analysis of microRNAs through stage I-III papillary thyroid carcinoma. Int J Med Sci 10: 585-592, 2013.
29. Suresh R, Sethi S, Ali S, Giorgadze T and Sarkar FH: Differential expression of MicroRNAs in papillary thyroid carcinoma and their role in racial disparity. J Cancer Sci Ther 7: 145-154, 2015.
30. Chen YT, Kitabayashi N, Zhou XK, Fahey TJ III and Scognamiglio T: MicroRNA analysis as a potential diagnostic tool for papillary thyroid carcinoma. Mod Pathol 21: 1139-1146, 2008.
31. Nikiforova MN, Tseng GC, Steward D, Diorio D and Nikiforov YE: MicroRNA expression profiling of thyroid tumors: Biological significance and diagnostic utility. J Clin Endocrinol Metab 93: 1600-1608, 2008.
32. McCarthy JJ: MicroRNA-206: The skeletal muscle-specific myomiR. Biochim Biophys Acta 1779: 682-691, 2008.
33. Liu X, He M, Hou Y, Liang B, Zhao L, Ma S, Yu Y and Liu X: Expression profiles of microRNAs and their target genes in papillary thyroid carcinoma. Oncol Rep 29: 1415-1420, 2013.
34. Yip L, Kelly L, Shuai Y, Armstrong MJ, Nikiforov YE, Carty SE and Nikiforova MN: MicroRNA signature distinguishes the degree of aggressiveness of papillary thyroid carcinoma. Ann Surg Oncol 18: 2035-2041, 2011.
35. Leone V, D'Angelo D, Rubio I, de Freitas PM, Federico A, Colamaio M, Pallante P, Medeiros-Neto G and Fusco A: MiR-1 is a tumor suppressor in thyroid carcinogenesis targeting CCND2, CXCR4, and SDF-1alpha. J Clin Endocrinol Metab 96: E1388-E1398, 2011.
36. Borgoño CA, Michael IP and Diamandis EP: Human tissue kallikreins: Physiologic roles and applications in cancer. Mol Cancer Res 2: 257-280, 2004.
37. Liu XL, Wazer DE, Watanabe K and Band V: Identification of a novel serine protease-like gene, the expression of which is down-regulated during breast cancer progression. Cancer Res 56: 3371-3379, 1996.
38. Goyal J, Smith KM, Cowan JM, Wazer DE, Lee SW and Band V: The role for NES1 serine protease as a novel tumor suppressor. Cancer Res 58: 4782-4786, 1998.
39. Mancikova V, Buj R, Castelblanco E, Inglada-Pérez L, Diez A, de Cubas AA, Curras-Freixes M, Maravall FX, Mauricio D, Matias-Guiu X, *et al*: DNA methylation profiling of well-differentiated thyroid cancer uncovers markers of recurrence free survival. Int J Cancer 135: 598-610, 2014.