



Integrative Analysis of Methylation and Transcriptome Identified Epigenetically Regulated lncRNAs With Prognostic Relevance for Thyroid Cancer

Qiuying Li¹, Peng Wang¹, Chuanhui Sun², Chao Wang^{1*} and Yanan Sun^{1*}

OPEN ACCESS

Edited by:

Jianzhong Su,
Wenzhou Medical University, China

Reviewed by:

Hengqiang Zhao,
Peking Union Medical College
Hospital (CAMS), China
Yang Wang,
Hubei University, China

*Correspondence:

Chao Wang
hijwangchao@163.com
Yanan Sun
syn2767@126.com

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

Received: 27 October 2019

Accepted: 10 December 2019

Published: 09 January 2020

Citation:

Li Q, Wang P, Sun C, Wang C and Sun Y (2020) Integrative Analysis of Methylation and Transcriptome Identified Epigenetically Regulated lncRNAs With Prognostic Relevance for Thyroid Cancer. *Front. Bioeng. Biotechnol.* 7:439. doi: 10.3389/fbioe.2019.00439

¹ Department of Otorhinolaryngology, Head and Neck Surgery, The Second Affiliated Hospital, Harbin Medical University, Harbin, China, ² Department of Otorhinolaryngology, The First Affiliated Hospital, Guizhou University of Traditional Chinese Medicine, Guiyang, China

Emerging evidence has shown that epigenetic changes in DNA methylation, an important regulator of long non-coding RNA (lncRNA) expression, can disturb the expression patterns of lncRNAs and contribute to carcinogenesis. However, knowledge about crosstalk effects between DNA methylation and lncRNA regulation in thyroid cancer (THCA) remain largely unknown. In this study, we performed an integrated analysis of methylation and the transcriptome and identified 483 epigenetically regulated lncRNAs (EpiLncRNAs) associated with the development and progression of THCA. These EpiLncRNAs can be divided into two categories based on their methylation and expression patterns: 228 HyperLncRNAs and 255 HypoLncRNAs. Then, we identified a methylation-driven 5-lncRNA-based signature (EpiLncPM) to improve prognosis prediction using the random survival forest and multivariate Cox analysis, which were then validated using the training dataset [Hazard ratio (HR) = 50.097, 95% confidence interval (CI): 10.231-245.312, $p < 0.001$] and testing dataset (HR = 4.395, 95% CI: 0.981-19.686, $p = 0.053$). Multivariate analysis suggested that the EpiLncPM is an independent prognostic factor. By performing a functional enrichment analysis of GO and KEGG for mRNAs co-expressed with the EpiLncPM, we found that the EpiLncPM was involved in immune and inflammatory-related biological processes. Finally, *in situ* hybridization analysis in 119 papillary thyroid carcinoma (PTC) tissues and paired adjacent normal tissues revealed that selected candidate lncRNA *AC110011* has significantly higher expression of PTC compared to adjacent non-neoplastic tissues, and was closely related to the tumor size, lymph node metastasis, and extrathyroidal extension. In summary, our study characterized the crosstalk between DNA methylation and lncRNA, and provided novel biomarkers for the prognosis of THCA.

Keywords: thyroid cancer, long non-coding RNAs, DNA methylation, biomarkers, prognosis

INTRODUCTION

Thyroid cancer (THCA) constitutes one of the most frequently diagnosed types of head and neck tumors (Herouiu Cataloio et al., 2013). Incidence rates of THCA had a significant upward trend in recent decades worldwide. It is estimated that the newly diagnosed cases of thyroid cancer are almost 50,000 in the United States in 2019 (Siegel et al., 2019). In China, the incidence rate of THCA in women has ranked fourth among the 10 most common cancers (Chen et al., 2016). Although patients diagnosed with THCA tended to have a favorable prognosis through surgery and radioiodine therapy, more than 10% of THCA patients will have a recurrence leading to considerable morbidity (Soares et al., 2014). Therefore, identifying reliable and accurate biomarkers for diagnosis and prognosis of THCA remains a challenge.

Long non-coding RNAs (lncRNAs) have been recognized as the major class of RNAs with more than 200 nucleotides that do not encode protein (Kopp and Mendell, 2018). A large body of literature has shown that lncRNAs play essential roles in a wide variety of biological processes, such as developmental and differentiation processes by lncRNA-mediated execution of gene expression programs (Fatica and Bozzoni, 2014; Marchese et al., 2017). High-throughput sequencing and profile analysis have identified a large number of differentially expressed lncRNAs in a multitude of cancers compared to normal tissues. In addition, aberrant lncRNA expression contributes to the development and maintenance of human cancers, thereby demonstrating the potential of lncRNAs as novel biomarkers in cancer diagnosis, prognosis, and as therapeutic targets (Spizzo et al., 2012; Sun et al., 2014; Huarte, 2015; Zhou et al., 2015a,b; Sanchez Calle et al., 2018; Zhou M. et al., 2018a,c; Bao et al., 2019). DNA methylation is one of the most common epigenetic mechanisms and is essential for the regulation of gene expression (Moore et al., 2013). It has been reported that epigenetic alterations in DNA methylation are also associated with various human cancers, including THCA (Stephen et al., 2011; Mancikova et al., 2014). Recent studies have reported aberrant DNA methylation of lncRNA promoters leading to perturbations of gene regulatory network, implying complex interplay between lncRNAs and DNA methylation (Zhao et al., 2016; Morlando and Fatica, 2018). However, crosstalk effects between DNA methylation and lncRNA regulation in THCA remain largely unknown.

In this study, we performed a genome-wide integrated analysis of methylation and the transcriptome to characterize the crosstalk between DNA methylation and lncRNA regulation, and identify epigenetically regulated lncRNAs. We further investigated the potential clinical value of these epigenetically regulated lncRNAs in a large number of THCA patients through bioinformatics analysis and experimental methods.

METHODS AND MATERIALS

Patients and Tissue Samples

The clinical information of 507 THCA patients was downloaded from The Cancer Genome Atlas (TCGA, <https://www.cancer.gov/>) database.

TABLE 1 | The relation of *AC110011* expression and clinicopathological characteristics of PTC patients.

Characteristics	Number	AC110011 expression		P-value
		High	Low	
Gender				
Male	13	7	6	0.7935
Female	106	53	53	
Age (Years)				
<45	67	30	37	0.1622
≥45	52	30	22	
Multifocality				
Yes	22	11	11	0.9652
No	97	49	48	
Tumor size (cm)				
≤2	98	44	54	0.0092*
>2	21	16	5	
Lymph node metastasis				
Yes	56	36	20	0.0043*
No	63	24	39	
Extral thyroidal extension				
Yes	23	19	4	0.0006*
No	96	41	55	
TNM stage				
I+II	106	51	55	0.1506
III+IV	13	9	4	

* $P \leq 0.05$ was considered significant.

The bold values indicates that they have statistical significance.

Another in-house dataset including 119 PTC tissues and paired adjacent normal tissues were collected from patients who underwent surgery at the Second Affiliated Hospital of Harbin Medical University (HMU). No patients received any local or systemic treatments before the operation. The adjacent non-cancerous tissues were collected >2 cm from the tumor margins on the same or another lobe. All tissue samples were confirmed independently by two pathologists, blocked of formalin-fixed paraffin-embedded material and stored at 2–8°C with desiccation until use for later experiments. The clinicopathological characteristics of these patients are listed in **Table 1**. Informed consent was obtained for each patient, and the experiments were allowed by the Research Ethics Committee of HMU.

Processing and Analysis of lncRNA Expression Profiles

RNA-seq data of 510 THCA tumor tissues and 58 non-cancer tissues based on the IlluminaHiSeq_RNASeq platform were retrieved from the UCSC Xena Browser (<https://xena.ucsc.edu/>). We annotate 15,873 lncRNAs from the RNA-seq data based on GENCODE (v23) annotations. Differential expression analysis was conducted using the R package “DESeq2.” Those lncRNAs with $|\log_2(\text{fold change})| > 1$ and False Discovery Rate (FDR)-corrected p -values < 0.05 were identified as differentially

expressed lncRNAs. Hierarchical clustering analysis of samples based on lncRNA expression was performed using the R package “pheatmap” with the “ward.D2” method.

Processing and Analysis of DNA Methylation Profiling

DNA methylation data of 515 THCA tumor tissues and 56 non-cancer tissues based on the Illumina Human Methylation 450 platform were retrieved from the UCSC Xena Browser (<https://xena.ucsc.edu/>). After removing probes with the missing value in more than 10% of samples, a total of 372,978 probes were stored for further analysis. Differentially methylated CpG loci between paired tumor tissues and non-cancer tissues were identified using a paired *t*-test with FDR-corrected *p*-values < 0.05 and absolute differences between group methylation mean (DGMB) > 0.04. Differentially methylated CpG loci between tumors with and without recurrence were determined using the R package “minfi” with FDR-corrected *p*-values < 0.05 and DGMB > 0.04.

Survival Analysis

The random survival forest was carried out to identify the optimal combination of epigenetically regulated lncRNAs as novel signature for survival prediction through the R package “randomforestSRC” (Taylor, 2011). Kaplan-Meier survival curves and log-rank tests were used to compare survival differences between high-risk group and low-risk group with the R package “survival.” Univariate and multivariate Cox proportional hazards regression analyses were conducted through the R package “survival.” The time-dependent Receiver Operating Characteristic (ROC) curve was used to evaluate the performance of the signature for survival prediction using the R package “survivalROC.”

Function Enrichment Analysis

Function enrichment analysis of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) was carried out using the R package “clusterProfiler” (Yu et al., 2012). GO terms or KEGG pathways with adjusted *p*-value < 0.05 were taken to be significantly enriched.

RNAscope *in situ* Hybridization

Paraffin sections were disposed according to the agreement developed by the Advanced Cell Diagnostics. ISH was disposed using the RNAscope[®] 2.5 Assay (ACD, Inc. Catalog No. 322335) and RNAscope[®] 2.5HD Detection Kit -BROWN (ACD, Inc. Cat. No. 322310). RNAscope Probe-Hs-AC110011 and positive and negative control probes were ordered from ACD. These probes included positive control probes PPIB (positive control, Cat. No. 313901), and negative control probe DapB (negative control, Cat. No. 310043). The percentages of positive cancer cells were scored as follows: 0: none; 1: <10%; 2: 10–50%; and 3: >50%. A score of 2 was used to distinguish between low (<2) and high (≥2) levels of AC110011 gene expression. Images of the slides were analyzed using an OLYMPUS Dual-CCD microscope digital camera, and relevant semi-quantitative scores were acquired by estimating the punctate staining. In addition, half of the quantitative fraction was acquired by estimating the dot dyeing.

RESULTS

Identification of Altered lncRNAs in the Development and Progression of THCA

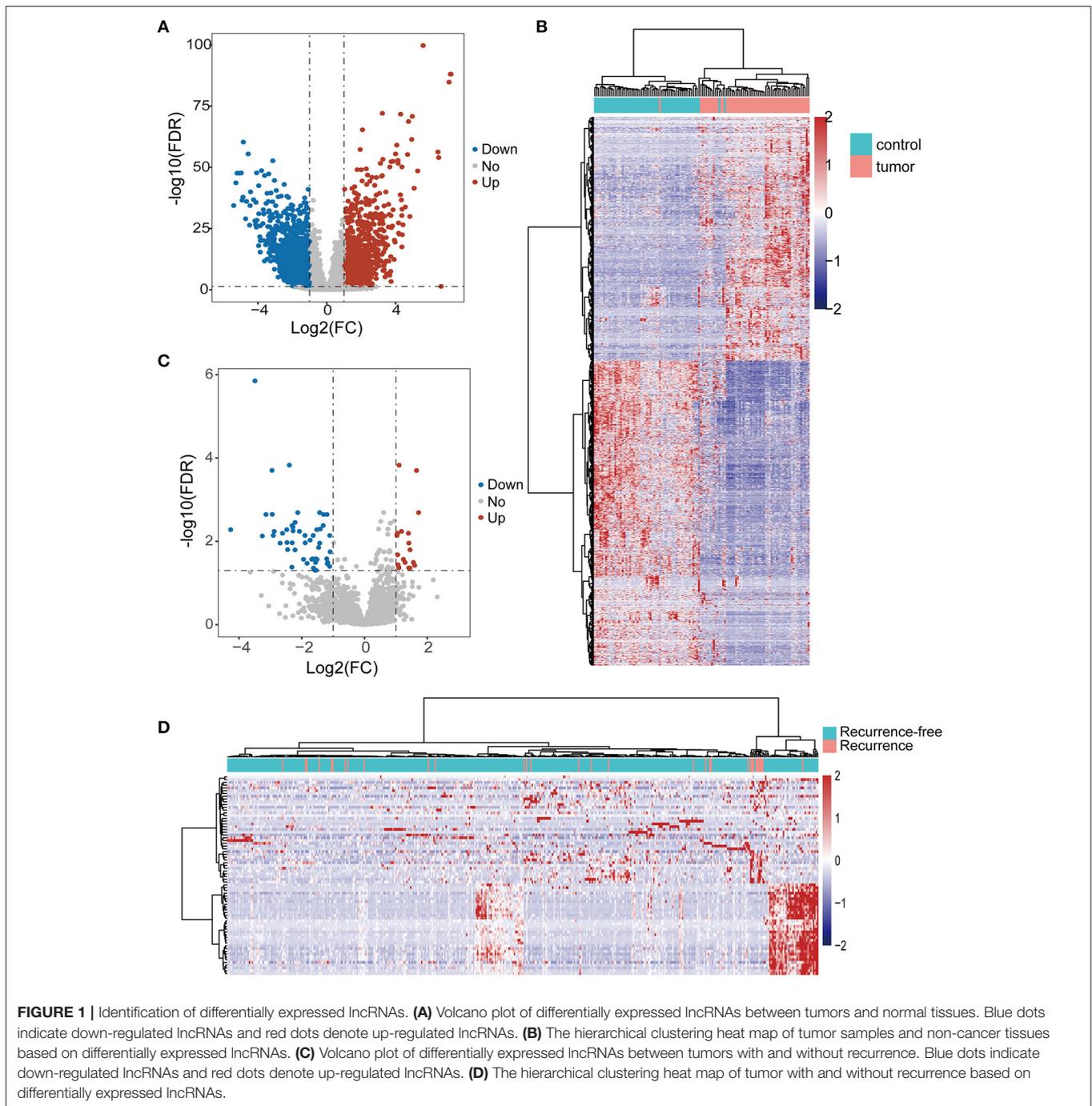
We first performed differential expression analysis for lncRNAs in 58 pairs of tumors and non-cancer tissues of THCA. A total of 1,969 lncRNAs were found to be differentially expressed, including 868 up-regulated and 1,101 down-regulated lncRNAs in tumors compared to non-cancer tissues (Figure 1A). Hierarchical clustering analysis showed that the expression patterns of 1,969 lncRNAs were capable of distinguishing tumor samples from non-cancer tissues (Figure 1B). To identify lncRNAs associated with tumor recurrence, differential expression analysis for lncRNAs was undertaken between 32 recurrent tumors and 401 recurrence-free tumors. Finally, we identified 72 lncRNAs that were significantly altered in recurrent tumors vs. recurrence-free tumors. Of them, 21 lncRNAs were up-regulated and 51 lncRNAs were downregulated in recurrent tumors compared to recurrence-free tumors (Figure 1C). Hierarchical clustering analysis showed that all samples fell into two clusters that are significantly associated with recurrence status (*p* = 0.0008, Chi-square test; Figure 1D).

Differential DNA Methylation Profiling During THCA Development and Progression

Further comparison of genome-wide DNA methylation on 56 pairs of tumor and adjacent normal tissues of THCA identified a total of 41,157 (11.03%) differentially methylated CpG sites, including 18,848 (45.8%) hypermethylated CpG sites (FDR < 0.05 and DGMB > 0.04) and 22,309 (54.2%) hypomethylated CpG sites (FDR < 0.05 and DGMB < -0.04) in tumors compared to adjacent normal tissue. To identify novel CpG islands aberrantly methylated in THCA recurrence, we also performed differential DNA methylation analysis between tumors with and without recurrence. We identified 20,283 (5.44%) differentially methylated CpG sites including 14,654 (72.25%) hypermethylated CpG sites (FDR < 0.05 and DGMB > 0.04) and 5,629 (27.75%) hypomethylated CpG sites (FDR < 0.05 and DGMB < -0.04) in tumors with recurrence compared to recurrence-free tumors. A Manhattan plot showed that these differentially methylated CpG sites are significantly associated with THCA development and progression is distributed across the methylome (Figures 2A,B).

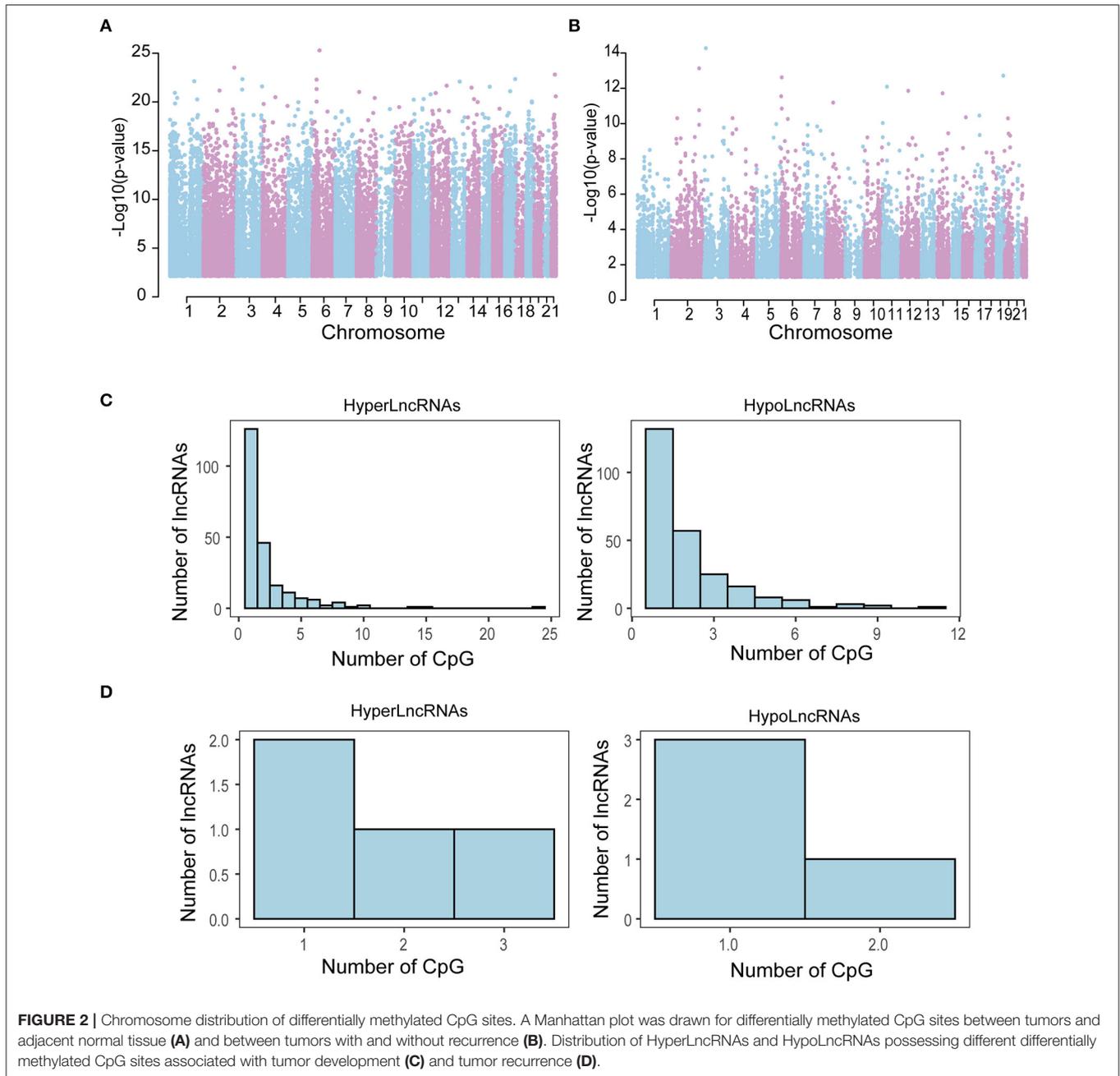
The Integrated Analysis Identifies Epigenetically Regulated lncRNAs in THCA

To identify epigenetically regulated lncRNAs (Epi-lncRNAs) in THCA, these differentially methylated CpG sites were mapped to 3 kb upstream and 3 kb downstream of differentially expressed lncRNAs. A total of 7,765 differentially methylated CpG sites were located in 689 up-regulated lncRNAs and 10,349 differentially methylated CpG sites were located in 873 down-regulated lncRNAs in tumors compared to adjacent



normal tissues. For tumors with recurrence, there were 427 differentially methylated CpG sites located in 20 up-regulated lncRNAs and 334 differentially methylated CpG sites located in 40 down-regulated lncRNAs. Then we grouped these lncRNAs into categories based on DNA methylation in lncRNA: HyperLncRNAs and HypoLncRNAs. HyperLncRNAs are down-regulated lncRNAs with high DNA methylation levels, whereas HypoLncRNAs are up-regulated lncRNAs with low DNA methylation levels. Finally, we identified 475 methylation-driven

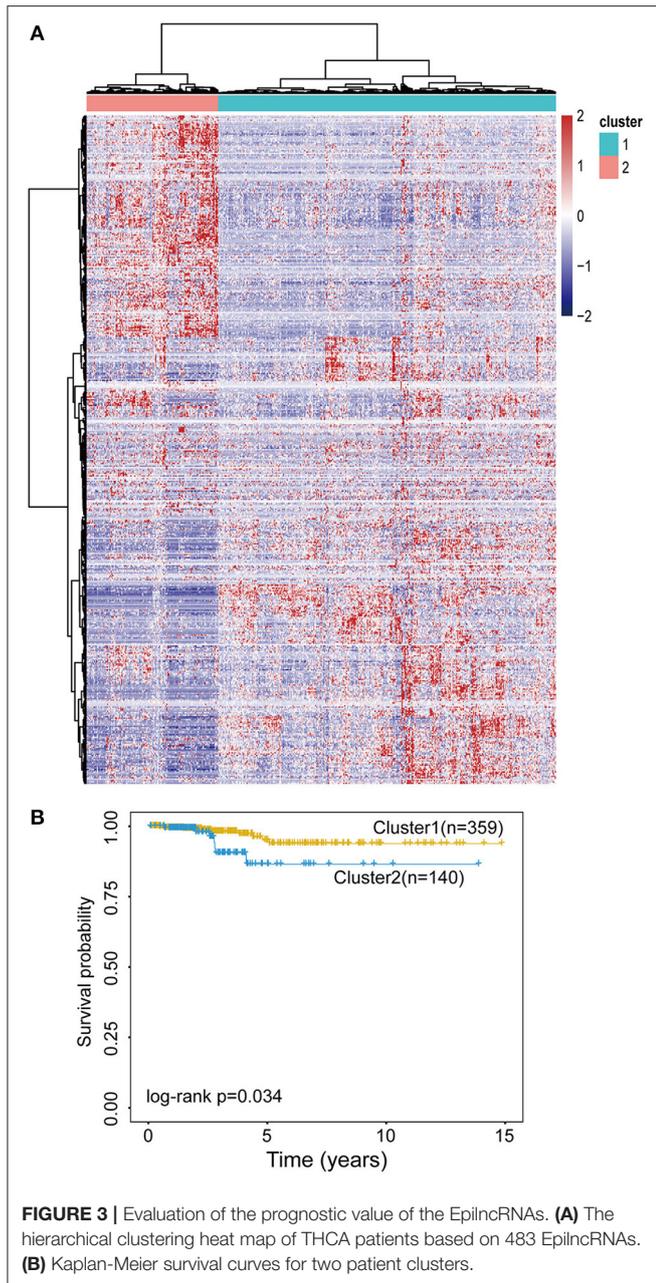
lncRNAs (224 HyperLncRNAs and 251 HypoLncRNAs) associated with tumor development and 8 methylation-driven lncRNAs (4 HyperLncRNAs and 4 HypoLncRNAs) associated with tumor recurrence. These results demonstrated that methylation-mediated dysregulated lncRNA expression patterns involved in THCA development and progression. Furthermore, we found that 228 HyperLncRNAs displayed dissimilar DNA methylation patterns with 255 HypoLncRNAs (**Figures 2C,D**).



Construction and Validation of an Epigenetically Regulated lncRNA-Based Prognostic Model (EpiLncPM) in THCA

To examine whether epigenetically regulated lncRNAs have prognostic value for THCA patients, we first performed a hierarchical clustering analysis of all THCA patients based on the 483 methylation-driven lncRNAs. As shown in **Figure 3A**, all THCA were grouped into two patient clusters with significantly different survival ($p = 0.034$, log-rank test; **Figure 3B**), suggesting that epigenetically regulated lncRNAs may be used as biomarkers in survival prediction of THCA. Therefore, all

THCA patients were randomly split into the training dataset ($n = 250$) and testing dataset ($n = 249$) and performed univariate Cox regression analysis for 475 methylation-driven lncRNAs in the training dataset. A total of 42 methylation-driven lncRNAs were significantly associated with survival. Thereafter, we performed the RSF on 42 candidate methylation-driven lncRNA biomarkers and identified an optimal combination of 5 methylation-driven lncRNAs as a prognostic signature. Finally an epigenetically regulated lncRNA-based prognostic scoring model (EpiLncPM) was developed in the training dataset using the expression levels of five optimal methylation-driven



lncRNAs weighted by their coefficients from multivariate Cox regression analysis as follows: $\text{EpiLncPM} = 13.1 * AP006248.2 + 2.53 * AC068580.3 + 33.2 * AC016396.2 + 3.12 * LINC01140 + 1.19 * LINC01135$ (Table 2). Patients in the training dataset were divided into the high-risk group ($n = 26$) and low-risk group ($n = 224$) according to the optimal risk cutoff point from the R package “maxstat.” As expected, the survival analysis showed that patients with a low-risk score have a better prognosis than those with a high-risk score ($p < 0.001$, log-rank test; Figure 4A). The time-dependent ROC analysis showed that the AUC of the performance of the EpiLncPM for survival prediction at three, 5 and 10-year survival rates in the training dataset reached 0.948, 0.965, and 0.949, respectively (Figure 4B).

TABLE 2 | Detailed information of five optimal methylation-driven prognostic lncRNAs.

Ensembl version	Ensembl name	Genomic location	P-value
ENSG00000279271.1	AP006248.2	Chromosome 8: 17,498,647-17,499,238(-)	0.008
ENSG00000235027.1	AC068580.3	Chromosome 11: 1,760,348-1,762,486 (+)	0.009
ENSG00000261076.1	AC016396.2	Chromosome 10: 58,325,614-58,327,030 (-)	0.002
ENSG00000267272.5	LINC01140	Chromosome 1: 87,129,765-87,169,198 (+)	0.023
ENSG00000234807.7	LINC01135	Chromosome 1: 58,785,128-58,901,109 (+)	0.030

When the EpiLncPM was applied to the testing dataset, patients in the high-risk group had significantly shorter survival than those in the low-risk group ($p = 0.034$, log-rank test; Figure 4C). The time-dependent ROC analysis showed that the AUC of the performance of the EpiLncPM for survival prediction at three, 5 and 10-year survival rates in the testing dataset reached 0.542, 0.625 and 0.688, respectively (Figure 4D).

Independence of the EpiLncPM From Other Clinical Factors

Univariate Cox regression analysis showed that the EpiLncPM and stage are both significantly correlated with survival in the training and testing datasets (HR = 50.097, 95% CI 10.231-245.312, $p < 0.001$ for training dataset; HR = 4.395, 95% CI 0.981-19.686, $p = 0.053$; Table 3). Therefore, to investigate whether the EpiLncPM is an independent factor in predicting survival, we performed multivariate analysis including age, gender, stage, and the EpiLncPM. Results of multivariate analyses showed that the EpiLncPM still maintained a significant association with survival adjusting by other clinical factors (Table 3). These observations showed that EpiLncPM is an independent factor in predicting the survival of THCA patients.

Functional Implication of the EpiLncPM

To gain insights into the functional role of the EpiLncPM, we calculated the Pearson correlation coefficient between expression levels of lncRNAs and mRNAs using the EpiLncPM and identified 537 mRNAs that correlated with the EpiLncPM (Pearson correlation coefficient > 0.5). GO functional enrichment analysis revealed that the correlated mRNAs were significantly clustered in immune and inflammatory-related biological processes (Figure 5A). In addition, the KEGG enrichment analysis showed that the correlated mRNAs were enriched in known cancer-related pathways (Figure 5B).

In situ Hybridization Analysis

The expression of AC110011 in 119 paraffin-embedded tissue samples of PTC and adjacent non-neoplastic tissues was semi-quantitatively examined by ISH. ISH showed that AC110011 was expressed in the nucleus (Figure 6). PTC tissues showed a significant increase in AC110011 expression as compared to that observed in adjacent non-neoplastic tissues. The AC110011

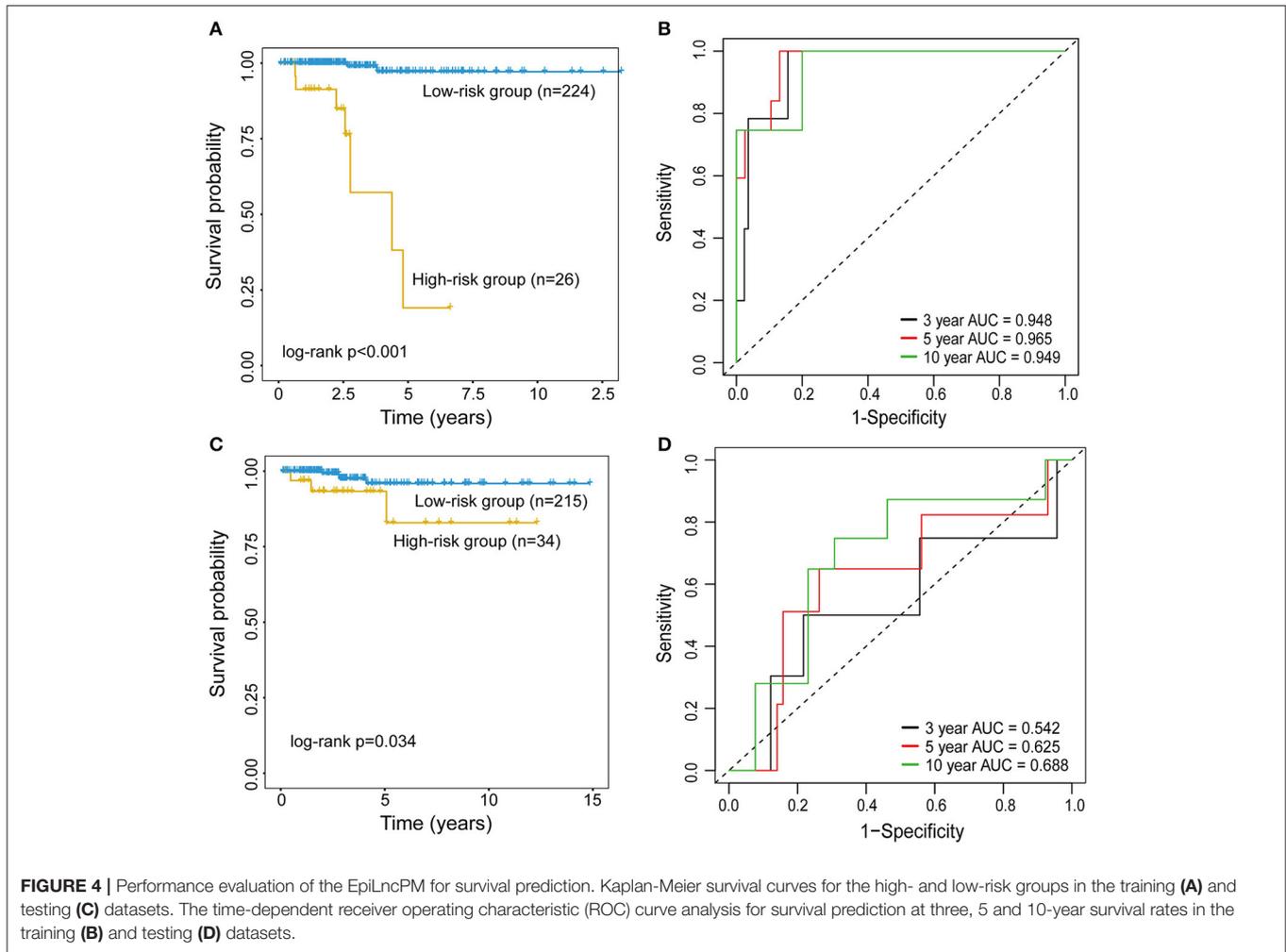


FIGURE 4 | Performance evaluation of the EpiLncPM for survival prediction. Kaplan-Meier survival curves for the high- and low-risk groups in the training (A) and testing (C) datasets. The time-dependent receiver operating characteristic (ROC) curve analysis for survival prediction at three, 5 and 10-year survival rates in the training (B) and testing (D) datasets.

TABLE 3 | Univariate and multivariate analysis for survival in the training and testing datasets.

Variables		Univariate analysis			Multivariate analysis		
		HR	95% CI	p-value	HR	95% CI	p-value
Training dataset (n = 250)							
EpiLncPM	High vs. low	50.097	10.231–245.312	<0.001	29.050	5.94–142.074	<0.001
Age	Old vs. young	Inf	0–Inf	0.998	Inf	0–Inf	0.990
Gender	Female vs. male	1.307	0.271–6.305	0.739	0.237	0.247–6.509	0.776
Stage	III, IV vs. I, II	9.377	1.933–45.492	0.005	0.302	0.26–7.051	0.720
Testing dataset (n = 249)							
EpiLncPM	High vs. low	4.395	0.981–19.686	0.053	3.699	0.813–16.833	0.091
Age	Old vs. young	Inf	0–Inf	0.998	Inf	0–Inf	0.999
Gender	Female vs. male	0.202	0.045–0.915	0.038	0.268	0.058–1.233	0.091
Stage	III, IV vs. I, II	5.416	1.048–27.997	0.044	1.083	0.207–5.662	0.924

expression was statistically associated with tumor size (cm), lymph node metastasis, and extrathyroidal extension (Table 1).

DISCUSSION

Recent multi-omics analysis has demonstrated that THCA is a heterogeneous disease characterized by a high degree

of molecular heterogeneity (Killock, 2014; Yoo et al., 2019), implying the potential of molecular profiling as a predictive factor in the diagnosis, prognosis and treatment response for THCA which will overcome the limitations of conventional clinical and histopathological features. mRNA and miRNA expression patterns have been widely investigated for a long time and many mRNA- or miRNA-based signatures have been identified

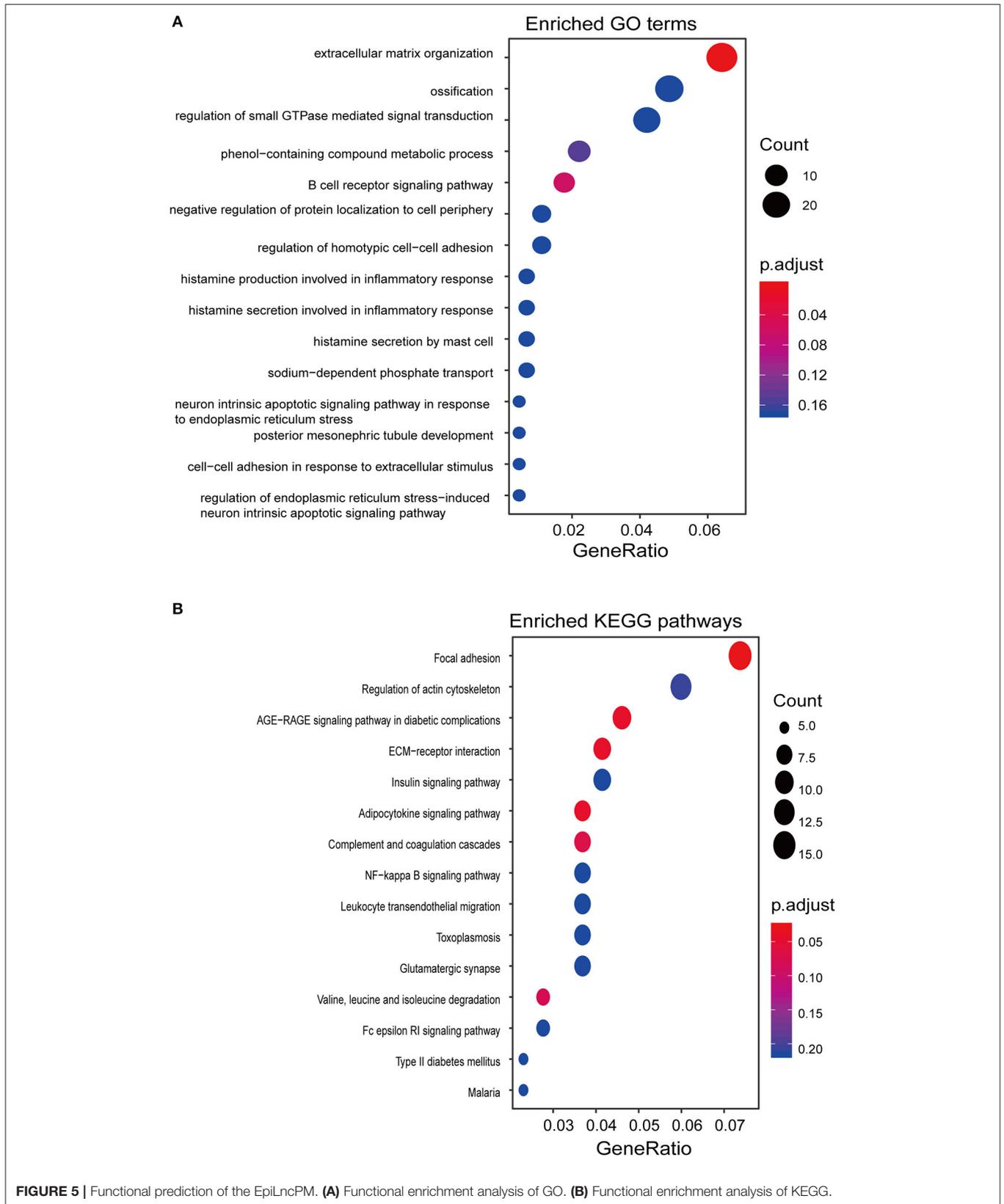
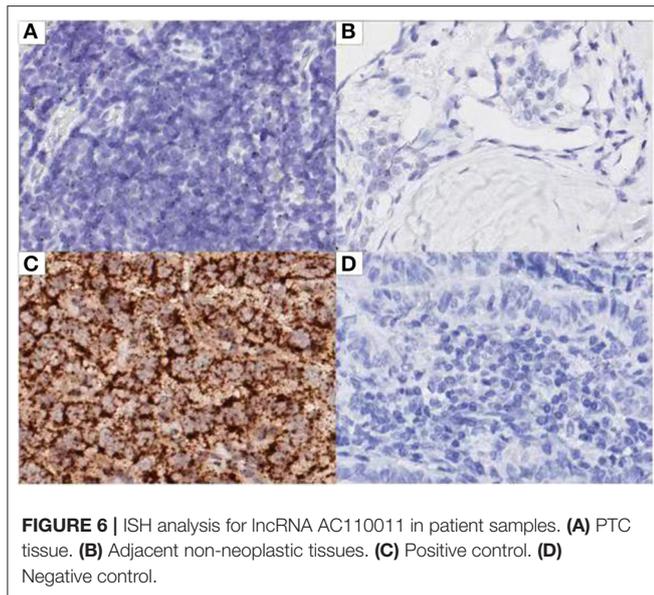


FIGURE 5 | Functional prediction of the EpiLncPM. **(A)** Functional enrichment analysis of GO. **(B)** Functional enrichment analysis of KEGG.



to improve the diagnosis and prognosis of THCA (Santarpi et al., 2013; Han et al., 2018; Kim et al., 2018; Teng et al., 2018; Wang et al., 2019). In recent years, a new ncRNA class, lncRNA, has gradually become a research hotspot in diverse cancer fields. Numerous studies have demonstrated the superiority of lncRNA as diagnostic and predictive biomarkers in a diverse range of cancers compared to mRNAs and miRNAs (Zhou et al., 2015c, 2017, 2019; Zhou M. et al., 2018b). Aberrant lncRNA expression has also been observed in the development and progression of THCA (Yang et al., 2016; Lu et al., 2018), and several lncRNA biomarkers have been identified for THCA diagnosis and prognosis (Li et al., 2017; Liu et al., 2018; Liang and Sun, 2019).

Emerging evidence has shown that DNA methylation is an important epigenetic regulator of lncRNA expression and epigenetic changes in DNA methylation can disturb the expression pattern of lncRNAs and contributes to carcinogenesis (Heilmann et al., 2017; Tang, 2018; Zhou Z. et al., 2018; Bao et al., 2019). However, the interplay between lncRNA regulation and DNA methylation in the development and progression of THCA is still largely unknown. In this study, we performed an integrated analysis of methylation and the transcriptome to characterize the dysregulated DNA methylation pattern of lncRNAs. By performing a differential expression analysis, we observed genome-wide changes in lncRNA expression and DNA methylation during THCA development and progression. By directly mapping altered DNA methylation to the flanking regions of lncRNAs, we found that a total of 483 differentially expressed lncRNAs were epigenetically deregulated, which were defined as methylation-driven lncRNAs. These lncRNAs can be separated into two categories based on their methylation patterns and expression levels, between tumor and normal samples or between tumors with and without recurrence.

Despite the fact that several previous studies have identified a host of lncRNAs with prognostic roles in THCA, knowledge about the clinical value of methylation-driven lncRNAs remains limited. We performed clustering analysis for THCA patients based on the expression patterns of methylation-driven lncRNAs,

which were found to be able to distinguish patients with different prognoses, thereby demonstrating a potential function in the prognosis of THCA. Therefore, we identified a methylation-driven 5-lncRNA-based signature (EpiLncPM) to improve the prognosis prediction utilizing the RSF and multivariate Cox analysis. This EpiLncPM was validated by the training and testing datasets. Moreover, the prognostic capacity of the EpiLncPM is independent of other clinical and pathological factors for the survival of patients with THCA. By performing function enrichment analysis for mRNAs co-expressed with the EpiLncPM, we gained insight into the potential functional relevance of methylation-driven lncRNAs in THCA. We found that the EpiLncPM is involved in immune and inflammatory-related biological processes, such as B cell receptor signaling pathways, histamine production involved in inflammatory responses, histamine secretion by mast cells, and known cancer-related pathways. It has been reported that chronic inflammation and the tumor microenvironment play critical roles in cancer development and progression, including THCA (Cunha et al., 2014; Rotondi et al., 2018; Ferrari et al., 2019). For example, mast cell infiltrates are linked to extrathyroidal extension and invasiveness (Visciano et al., 2015). NF- κ B is associated with inflammatory and immune responses and its increased activity correlate with a more aggressive phenotype of THCA (Giuliani et al., 2018). Finally, we validated the functional roles of one selected methylation-driven lncRNA signature by *in situ* hybridization analysis in 119 PTC tissues and paired with adjacent normal tissues.

In summary, we performed a genome-wide integrated analysis of methylation and the transcriptome to characterize the crosstalk between DNA methylation and lncRNA, and identify epigenetically regulated lncRNAs. Additionally, we identified a methylation-driven 5-lncRNA-based signature (EpiLncPM) with potential clinical application in predicting the prognosis of THCA.

DATA AVAILABILITY STATEMENT

Clinical information of THCA patients was downloaded from The Cancer Genome Atlas (TCGA, <https://www.cancer.gov/>) database. RNA-seq data and DNA methylation data of tumor tissues and non-cancer tissues were retrieved from the UCSC Xena Browser (<https://xena.ucsc.edu/>).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Committee of Harbin medical university. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CW and YS conceived and designed the experiments. QL, PW, and CS performed the experiments and analyzed the data. QL, CW, and YS wrote the paper. All authors read and approved the final manuscript.

FUNDING

This study was supported by the scientific research project of Heilongjiang health and Family Planning Commission (Grant no. 2017-081); the Innovation

Research Fund for young and middle-aged in the Second Affiliated Hospital of Harbin Medical University: (Grant no. kycx2018-04); the scientific funds of National Natural Science Foundation of China (81772874 and 81272965).

REFERENCES

- Bao, S., Zhao, H., Yuan, J., Fan, D., Zhang, Z., Su, J., et al. (2019). Computational identification of mutator-derived lncRNA signatures of genome instability for improving the clinical outcome of cancers: a case study in breast cancer. *Brief. Bioinform.* doi: 10.1093/bib/bbz118. [Epub ahead of print].
- Chen, W., Zheng, R., Baade, P. D., Zhang, S., Zeng, H., Bray, F., et al. (2016). Cancer statistics in China, 2015. *CA Cancer J. Clin.* 66, 115–132. doi: 10.3322/caac.21338
- Cunha, L. L., Marcello, M. A., and Ward, L. S. (2014). The role of the inflammatory microenvironment in thyroid carcinogenesis. *Endocr. Relat. Cancer* 21, R85–R103. doi: 10.1530/ERC-13-0431
- Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* 15, 7–21. doi: 10.1038/nrg3606
- Ferrari, S. M., Fallahi, P., Galdiero, M. R., Ruffilli, I., Elia, G., Ragusa, F., et al. (2019). Immune and inflammatory cells in thyroid cancer microenvironment. *Int. J. Mol. Sci.* 20:4413. doi: 10.3390/ijms20184413
- Giuliani, C., Bucci, L., and Napolitano, G. (2018). The role of the transcription factor nuclear factor-kappa b in thyroid autoimmunity and cancer. *Front. Endocrinol.* 9:471. doi: 10.3389/fendo.2018.00471
- Han, L. O., Li, X. Y., Cao, M. M., Cao, Y., and Zhou, L. H. (2018). Development and validation of an individualized diagnostic signature in thyroid cancer. *Cancer Med.* 7, 1135–1140. doi: 10.1002/cam4.1397
- Heilmann, K., Toth, R., Bossmann, C., Klimo, K., Plass, C., and Gerhauser, C. (2017). Genome-wide screen for differentially methylated long noncoding RNAs identifies *Esrp2* and lncRNA *Esrp2-as* regulated by enhancer DNA methylation with prognostic relevance for human breast cancer. *Oncogene* 36, 6446–6461. doi: 10.1038/onc.2017.246
- Heroui Cataloiu, A. D., Danciu, C. E., and Popescu, C. R. (2013). Multiple cancers of the head and neck. *Maedica* 8, 80–85.
- Huarte, M. (2015). The emerging role of lncRNAs in cancer. *Nat. Med.* 21, 1253–1261. doi: 10.1038/nm.3981
- Killock, D. (2014). Genetics: the cancer genome atlas maps papillary thyroid cancer. *Nat. Rev. Clin. Oncol.* 11:681. doi: 10.1038/nrclinonc.2014.193
- Kim, K., Jeon, S., Kim, T. M., and Jung, C. K. (2018). Immune gene signature delineates a subclass of papillary thyroid cancer with unfavorable clinical outcomes. *Cancers* 10:E494. doi: 10.3390/cancers10120494
- Kopp, F., and Mendell, J. T. (2018). Functional classification and experimental dissection of long noncoding RNAs. *Cell* 172, 393–407. doi: 10.1016/j.cell.2018.01.011
- Li, Q., Li, H., Zhang, L., Zhang, C., Yan, W., and Wang, C. (2017). Identification of novel long non-coding RNA biomarkers for prognosis prediction of papillary thyroid cancer. *Oncotarget* 8, 46136–46144. doi: 10.18632/oncotarget.17556
- Liang, W., and Sun, F. (2019). Identification of pivotal lncRNAs in papillary thyroid cancer using lncRNA-mRNA-miRNA ceRNA network analysis. *PeerJ* 7:e7441. doi: 10.7717/peerj.7441
- Liu, G., Zheng, J., Zhuang, L., Lv, Y., Zhu, G., Pi, L., et al. (2018). A prognostic 5-lncRNA expression signature for head and neck squamous cell carcinoma. *Sci. Rep.* 8:15250. doi: 10.1038/s41598-018-33642-1
- Lu, W., Xu, Y., Xu, J., Wang, Z., and Ye, G. (2018). Identification of differential expressed lncRNAs in human thyroid cancer by a genome-wide analyses. *Cancer Med.* 7, 3935–3944. doi: 10.1002/cam4.1627
- Mancikova, V., Buj, R., Castelblanco, E., Inglada-Pérez, L., Diez, A., de Cubas, A. A., et al. (2014). DNA methylation profiling of well-differentiated thyroid cancer uncovers markers of recurrence free survival. *Int. J. Cancer* 135, 598–610. doi: 10.1002/ijc.28703
- Marchese, F. P., Raimondi, I., and Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.* 18:206. doi: 10.1186/s13059-017-1348-2
- Moore, L. D., Le, T., and Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology* 38, 23–38. doi: 10.1038/npp.2012.112
- Morlando, M., and Fatica, A. (2018). Alteration of epigenetic regulation by long noncoding RNAs in cancer. *Int. J. Mol. Sci.* 19:E570. doi: 10.3390/ijms19020570
- Rotondi, M., Coperchini, F., Latrofa, F., and Chiovato, L. (2018). Role of chemokines in thyroid cancer microenvironment: is CXCL8 the main player? *Front. Endocrinol.* 9:314. doi: 10.3389/fendo.2018.00314
- Sanchez Calle, A., Kawamura, Y., Yamamoto, Y., Takeshita, F., and Ochiya, T. (2018). Emerging roles of long non-coding RNA in cancer. *Cancer Sci.* 109, 2093–2100. doi: 10.1111/cas.13642
- Santarpia, L., Calin, G. A., Adam, L., Ye, L., Fusco, A., Giunti, S., et al. (2013). A miRNA signature associated with human metastatic medullary thyroid carcinoma. *Endocr. Relat. Cancer* 20, 809–823. doi: 10.1530/ERC-13-0357
- Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *CA Cancer J. Clin.* 69, 7–34. doi: 10.3322/caac.21551
- Soares, P., Celestino, R., Melo, M., Fonseca, E., and Sobrinho-Simões, M. (2014). Prognostic biomarkers in thyroid cancer. *Virchows Arch.* 464, 333–346. doi: 10.1007/s00428-013-1521-2
- Spizzo, R., Almeida, M. I., Colombatti, A., and Calin, G. A. (2012). Long non-coding RNAs and cancer: a new frontier of translational research? *Oncogene* 31, 4577–4587. doi: 10.1038/onc.2011.621
- Stephen, J. K., Chitale, D., Narra, V., Chen, K. M., Sawhney, R., and Worsham, M. J. (2011). DNA methylation in thyroid tumorigenesis. *Cancers* 3, 1732–1743. doi: 10.3390/cancers3021732
- Sun, J., Shi, H., Wang, Z., Zhang, C., Liu, L., Wang, L., et al. (2014). Inferring novel lncRNA-disease associations based on a random walk model of a lncRNA functional similarity network. *Mol. Biosyst.* 10, 2074–2081. doi: 10.1039/C3MB70608G
- Tang, B. (2018). Inference of crosstalk effects between DNA methylation and lncRNA regulation in NSCLC. *Biomed Res. Int.* 2018:7602794. doi: 10.1155/2018/7602794
- Taylor, J. M. (2011). Random survival forests. *J. Thorac. Oncol.* 6, 1974–1975. doi: 10.1097/JTO.0b013e318233d835
- Teng, H., Mao, F., Liang, J., Xue, M., Wei, W., Li, X., et al. (2018). Transcriptomic signature associated with carcinogenesis and aggressiveness of papillary thyroid carcinoma. *Theranostics* 8, 4345–4358. doi: 10.7150/thno.26862
- Visciano, C., Prevede, N., Liotti, F., and Marone, G. (2015). Tumor-associated mast cells in thyroid cancer. *Int. J. Endocrinol.* 2015:705169. doi: 10.1155/2015/705169
- Wang, Z., Lv, J., Zou, X., Huang, Z., Zhang, H., Liu, Q., et al. (2019). A three plasma microRNA signature for papillary thyroid carcinoma diagnosis in Chinese patients. *Gene* 693, 37–45. doi: 10.1016/j.gene.2019.01.016
- Yang, M., Tian, J., Guo, X., Yang, Y., Guan, R., Qiu, M., et al. (2016). Long noncoding RNA are aberrantly expressed in human papillary thyroid carcinoma. *Oncol. Lett.* 12, 544–552. doi: 10.3892/ol.2016.4653
- Yoo, S. K., Song, Y. S., Lee, E. K., Hwang, J., Kim, H. H., Jung, G., et al. (2019). Integrative analysis of genomic and transcriptomic characteristics associated with progression of aggressive thyroid cancer. *Nat. Commun.* 10:2764. doi: 10.1038/s41467-019-10680-5
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterprofiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118
- Zhao, Y., Sun, H., and Wang, H. (2016). Long noncoding RNAs in DNA methylation: new players stepping into the old game. *Cell Biosci.* 6:45. doi: 10.1186/s13578-016-0109-3

- Zhou, M., Guo, M., He, D., Wang, X., Cui, Y., Yang, H., et al. (2015a). A potential signature of eight long non-coding RNAs predicts survival in patients with non-small cell lung cancer. *J. Transl. Med.* 13:231. doi: 10.1186/s12967-015-0556-3
- Zhou, M., Hu, L., Zhang, Z., Wu, N., Sun, J., and Su, J. (2018a). Recurrence-associated long non-coding RNA signature for determining the risk of recurrence in patients with colon cancer. *Mol. Ther. Nucleic Acids* 12, 518–529. doi: 10.1016/j.omtn.2018.06.007
- Zhou, M., Wang, X., Li, J., Hao, D., Wang, Z., Shi, H., et al. (2015b). Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network. *Mol. Biosyst.* 11, 760–769. doi: 10.1039/C4MB00511B
- Zhou, M., Zhang, Z., Zhao, H., Bao, S., Cheng, L., and Sun, J. (2018c). An immune-related six-lncRNA signature to improve prognosis prediction of glioblastoma multiforme. *Mol. Neurobiol.* 55, 3684–3697. doi: 10.1007/s12035-017-0572-9
- Zhou, M., Zhang, Z., Zhao, H., Bao, S., and Sun, J. (2018b). A novel lncRNA-focus expression signature for survival prediction in endometrial carcinoma. *BMC Cancer* 18:39. doi: 10.1186/s12885-017-3983-0
- Zhou, M., Zhao, H., Wang, X., Sun, J., and Su, J. (2019). Analysis of long noncoding RNAs highlights region-specific altered expression patterns and diagnostic roles in Alzheimer's disease. *Brief. Bioinform.* 20, 598–608. doi: 10.1093/bib/bby021
- Zhou, M., Zhao, H., Wang, Z., Cheng, L., Yang, L., Shi, H., et al. (2015c). Identification and validation of potential prognostic lncRNA biomarkers for predicting survival in patients with multiple myeloma. *J. Exp. Clin. Cancer Res.* 34:102. doi: 10.1186/s13046-015-0219-5
- Zhou, M., Zhao, H., Xu, W., Bao, S., Cheng, L., and Sun, J. (2017). Discovery and validation of immune-associated long non-coding RNA biomarkers associated with clinically molecular subtype and prognosis in diffuse large B cell lymphoma. *Mol. Cancer* 16:16. doi: 10.1186/s12943-017-0580-4
- Zhou, Z., Lin, Z., Pang, X., Tariq, M. A., Ao, X., Li, P., et al. (2018). Epigenetic regulation of long non-coding RNAs in gastric cancer. *Oncotarget* 9, 19443–19458. doi: 10.18632/oncotarget.23821

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Li, Wang, Sun, Wang and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.