ORIGINAL ARTICLE

WILEY

Promotor methylation status of MAPK4 is a novel epigenetic biomarker for prognosis of recurrence in patients with thymic epithelial tumors

Wei Guan[1†] | Songlin Li[2,3†] | Zhimin Zhang[4†] | He Xiao[1] | Juan He[1] | Jian Li[1] | Xuan He[1] | Jia Luo[1] | Yun Liu[1] | Lin Lei[1] | Jungang Ma[1] | Lizhao Chen[5] | Chuan Chen[1]

[1]Department of Cancer Center, Daping Hospital, Army Medical University, Chongqing, China

[2]Cancer Center, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu, China

[3]Chinese Academy of Sciences Sichuan Translational Medicine Research Hospital, Chengdu, China

[4]Cancer Center, Renmin Hospital of Wuhan University, Wuhan, Hubei, China

[5]Department of Neurosurgery, Daping Hospital, Army Medical University, Chongqing, China

Correspondence
Chuan Chen, Department of Cancer Center, Daping Hospital, Army Medical University, Chongqing, 400042, China, No. 10 Changjiang Zhilu, Yuzhong, Chongqing, 400042, China.
Email: sinkriver@126.com

Lizhao Chen, Department of Neurosurgery, Daping Hospital, Army Medical University, Chongqing, 400042, China, No. 10 Changjiang Zhilu, Yuzhong, Chongqing, 400042, China.
Email: chenlizhao1971@163.com

Abstract

Background: The prognosis of thymic epithelial tumors (TETs) currently relies on the commonly adopted WHO classification and Masaoka staging system, which cannot reflect the undefined biological behaviors limiting them as prognostic factors.

Methods: In this study, we first identified 40 genes and 179 genes, respectively that were epigenetically upregulated and silenced, corresponding to a total of 509 functionally methylated CpG sites between thymomas and thymic carcinomas by using the TCGA dataset.

Results: The methylation β-values of cg20068620 in *MAPK4* and cg18770944 in *USP51* were significantly associated with recurrence-free survival (RFS). In the independent validation cohort, only WHO classification and methylation β-values of cg20068620 in *MAPK4* were independent prognostic factors for RFS in Chinese patients with TETs. A linear weighted model including these two factors was used to calculate the recurrence risk score (RRS). Time-dependent ROC curve analysis revealed that RRS was overwhelmingly superior to WHO classification for predicting 3-, 5-, and 10-year RFS and Masaoka stage for 3- and 5-year RFS.

Conclusions: These results suggested that the methylation site cg20068620 in *MAPK4* can improve the accuracy of the WHO classification alone regarding the prognostic value of TETs recurrence.

KEYWORDS
DNA methylation, prognosis, pyrosequencing, recurrence, thymic epithelial tumors

INTRODUCTION

Thymic epithelial tumors (TETs) are the most common epithelial neoplasms of the anterior mediastinum.[1] The World Health Organization (WHO) classification and Masaoka staging are the most commonly used prognostic factors for TETs because they reflect their histological types, clinical findings, and prognosis.[2–4] The WHO classification system divides TETs into thymomas (type A, AB, B1, B2, and B3) and thymic carcinoma based on the morphology of epithelial tumor cells, degree of atypia, and relative proportion of the nontumoral lymphocytic component.[5] In contrast, the

---

† These authors have contributed equally to this work

Masaoka staging system divides TETs into four stages based on invasiveness.[6] However, many biological behaviors of TETs remain unclear.[7] Thus, the current WHO classification and Masaoka staging system cannot reflect these undefined biological behaviors of TETs, limiting them as prognostic factors for TETs. For example, even with complete resection, some TET patients still experience metastatic or local recurrence, which renders significant obstacles to the long-term survival of TET patients.[8] Unfortunately, the recurrence mechanism remains unclear, and no biomarkers can accurately predict recurrence in TETs.

Tumor recurrence is a complex process that involves many factors, including aberrant DNA methylation. DNA methylation is an essential epigenetic mechanism that regulates gene expression.[9] Thus, abnormal DNA methylation leads to abnormal gene expression, which leads to intracellular signal pathway disorders and ultimately tumor development and progression, including recurrence.[10,11] Importantly, dysregulated DNA methylation is also detected in TETs, and several gene methylation statuses are closely correlated with the prognosis of TETs,[12,13] suggesting that abnormal DNA methylation plays a crucial role in TET progression and may be a prognostic biomarker.

This study aimed to identify methylation markers that can predict TET recurrence. Here, we demonstrated different DNA methylation patterns between thymomas and thymic carcinomas, and different methylation levels led to differences in gene expression and signaling pathways between thymomas and thymic carcinomas. In addition, we demonstrated that the WHO classification and methylation site cg20068620 in *MAPK4* are independent predictors of recurrence in TET patients. Significantly, the combination of the WHO classification and methylation site cg20068620 in *MAPK4* can more accurately predict the recurrence of TET patients. Additionally, we identified that the Masaoka stage could predict the recurrence of thymomas.

METHODS

TET dataset of TCGA

The methylation dataset TCGA.THYM.sampleMap/ Human Methylation450 (version 2017-09-08) of TETs was downloaded from UCSC (https://xenabrowser.net/ datapages) and used to identify differentially methylated sites between the thymoma and thymic carcinoma groups. This dataset included 113 cases of thymoma and 11 cases of thymic carcinoma, and the clinicopathological characteristics of the patients are shown in Table 1. A raw count matrix of gene-level RSEM values from 120 TET tumor samples was downloaded from http://gdac.broadinstitute. org/runs/stddata_2015_11_01/ and used to identify differentially expressed genes (DEGs) between thymoma and thymic carcinoma. The clinicopathological characteristics of the 120 patients are shown in Table 2.

T A B L E 1 Clinicopathological characteristics of 124 cases from TCGA dataset THYM

Clinicopathological characteristics		n (%)
Recurrence-free survival	Nonsensored	109 (92.4)
	Censored	9 (7.6)
Gender	Female	60 (48.4)
	Male	64 (51.6)
WHO classification	A-B3 type	113 (91.1)
	C type	11 (8.9)
History myasthenia gravis	No	87 (71.9)
	Yes	34 (28.1)
Masaoka stage	I–IIB	99 (81.1)
	III–IV	23 (18.9)
Tumor tissue site	Thymus	97 (78.2)
	Anterior mediastinum	27 (21.8)
History of neoadjuvant treatment	No	122 (98.4)
	Yes	2 (1.6)
Postoperative radiotherapy and chemotherapy	No	114 (92.7)
	Yes	9 (7.3)
Radiation therapy	No	80 (65.0)
	Yes	43 (35.0)

Identification of epigenetically regulated genes and candidate CpG sites responsible for prognosis for RFS in TCGA datasets

In total, 392 653 DNA methylation probes were included for methylation site analysis after removing probes with missing values. For each probe, the Wilcoxon rank-sum test was used to evaluate differences in methylation β-values between thymoma and thymic carcinoma. The Bonferroni procedure adjusted per test p-values for multiple comparisons. The annotation for probes was performed by using the package "IMA" (Illumina methylation analyzer, version 3.1.2) and the annotation file "fullannotInd.rda". DEGs in the comparison of thymic carcinomas versus thymomas were determined with the R/Bioconductor limma package after filtering genes with reading counts <10 in at least 80% of cases, leaving 13,581 genes included in the analysis. DNA hypermethylation or hypomethylation events that could functionally regulate mRNA expression were respectively identified through the following criteria: (1) probe at the promoter or the first exon region of one gene (TSS1500, TSS200, 1stExon), (2) the mean methylation in thymic carcinomas increased >50% compared with that in thymomas with adjusted $p < 0.05$ as well as a mean methylation in thymic carcinoma >30%, and (3) log2Ratio < -1 and FDR <0.05 for the corresponding gene; or (1) probe at the promoter or the first exon region of one gene (TSS1500, TSS200, 1stExon), (2) the mean methylation in thymic carcinomas decreased >50% compared with that in thymomas with adjusted $p < 0.05$ as well as a mean methylation in thymomas >30%, and (3) log2Ratio >1 and FDR <0.05 for

TABLE 2 Clinicopathological characteristics of 120 cases from TCGA dataset THYM

Clinicopathological characteristics		n (%)
Recurrence free survival	Nonsensored	106 (93.0)
	Censored	8 (7.0)
Gender	Female	57 (47.5)
	Male	63 (52.5)
WHO classification	A-B3 type	109 (90.8)
	C type	11 (9.2)
History myasthenia gravis	No	83 (70.9)
	Yes	34 (29.1)
Masaoka stage	I–IB	97 (82.2)
	III–IV	21 (17.8)
Tumor tissue site	Thymus	93 (77.5)
	Anterior mediastinum	27 (22.5)
History of neoadjuvant treatment	No	118 (98.3)
	Yes	2 (1.7)
Postoperative radiotherapy and chemotherapy	No	111 (93.3)
	Yes	8 (6.7)
Radiation therapy	No	77 (64.7)
	Yes	42 (35.3)

TABLE 3 Clinicopathological characteristics of 95 patients with thymoma in validation set

Clinicopathological characteristics		n (%) median (IQR)
Recurrence-free survival	Censored	67 (70.5)
	Nonsensored	28 (29.5)
Gender	Female	37 (38.9)
	Male	58 (61.1)
History myasthenia gravis	No	64 (67.4)
	Yes	31 (32.6)
WHO classification	type A-B3	83 (87.4)
	type C	12 (12.6)
Masaoka stage	I-II	54 (56.8)
	III-IV	41 (43.2)
Adjuvant radiotherapy	No	65 (68.4)
	Yes	30 (31.6)
Adjuvant chemotherapy	No	68 (71.6)
	Yes	27 (28.4)
Age		50 (41–62)
cg20068620 (MAPK4)		27 (16–58)
cg18770944 (USP51)		55 (27–67)

mRNA expression of the corresponding gene. MSigDB c2 gene set was used to infer enrichment of genes that were potentially impacted by hypermethylation or hypomethylation with R package "clusterProfiler".

Because none of the clinicopathological characteristics, including sex (male vs. female), age, WHO classification (type C vs. type A-B3), Masaoka stage (III–IV vs. I-II) and adjuvant radiotherapy (yes vs. no), were significantly associated with RFS in the TCGA thymoma datasets revealed by univariate Cox regression, only univariate Cox regression was used to evaluate the prognostic value for each of 509 identified functional CpG sites. Only probes that showed significance with crude p-values < 0.05 were identified as candidate probes for further analysis. Hierarchical clustering and visualization of methylation β-values of 509 functional CpG sites across 124 TET cases was carried out through the "aheatmap" function in the R package "NMF" (version 0.1.3).

Validation of candidate CpG sites in the Daping Hospital cohort

Overall, 95 patients with histologically confirmed thymoma or thymic carcinoma were enrolled and hospitalized between October 2013 and October 2016 for thoracic surgery at the Daping Hospital of the Army Medical University. This study was approved by the Medical Ethics Committee of Daping Hospital. Written informed consent was obtained from all patients prior to their enrollment. The clinicopathological characteristics of the patients are shown in Tables 2 and 3.

Genomic DNA was extracted from formalin-fixed, paraffin-embedded (FFPE) TET tissues using the QIAamp DNA FFPE Tissue kit (Qiagen). The DNA concentration and purity were determined with a spectrophotometer (NanoDrop2000, Thermo Scientific). Bisulfite conversion of 500 ng purified DNA in each sample was performed with an EZ DNA Methylation-Gold kit according to the manufacturer's instructions (cat. no. D5006, Zymo Research Corporation, Orange). The bisulfite-converted DNA was amplified with TaKaRa EpiTaq HS (cat. no. R110A, Takara Biomedical Technology [Beijing] Co., Ltd) with the following reaction: 10 ng bisulfite-treated DNA, 0.4 μM forward primer and reverse primer, 2.5 μl 10 × EpiTap PCR Buffer, 2.5 mM $MgCl_2$, dNTP mixture (0.264 mM each), and EpiTap HS (0.025 U/μl) in 25 μl per reaction. The following thermal cycle conditions were used: denaturation at 98°C for 10 s, annealing at 55°C for 30 s, extension at 72°C for 30 s executed for 35 cycles followed by extension at 72°C for 1 min, and hold at 4°C. The amplicons were then subjected to pyrosequencing with PyroMark Q96 (Qiagen). All primers used are shown in Table S1.

Prognostic model development and statistical analysis

A weighted model was constructed for the prognostic model.[14] First, the prognostic values of sex (male vs. female), age, history of myasthenia gravis (yes vs. no), WHO classification (C vs. A-B3), Masaoka stage (III–IV vs. I–II), adjuvant radiotherapy (yes vs. no), adjuvant chemotherapy (yes vs. no) and methylation β-values in two

candidate CpG sites, cg20068620 and cg18770944, were determined using univariate Cox regression. Only factors significantly associated with RFS were included in the multivariable Cox regression with a stepwise forward selection procedure for the identification of independent prognostic factors in which one covariate was included with criteria $p < 0.05$ and excluded with criteria $p > 0.10$ based on the likelihood ratio test. The proportional hazards assumption for the Cox proportional hazards regression model was assessed through the Schoenfeld residuals test. The recurrence risk score (RRS) was constructed using the linear predictor of the finalized model. The whole cohort was dichotomized into low- and high-risk subgroups by median RRS.

All β values and other continuous variables are represented by the median values and interquartile ranges and visualized with box plots. The differences in methylation β-values in two candidate CpG sites between thymoma and thymic carcinoma were evaluated by the Kruskal–Wallis test. The cutoff value of methylation level that was used to define low- and high-methylation subgroups with maximum log-rank statistics in terms of RFS was determined by function "surv_cutpoint" in R package "survminer". The Kaplan–Meier method and the log-rank test were used to compare the RFS between low- and high-methylation subgroups or low- and high-risk subgroups. The predictive efficiency of RRS, Masaoka's stage and WHO classification for 3-, 5- and 10-year RFS was determined with time-dependent ROC curve analysis using the "time ROC" function. Comparisons between two time-dependent AUCs were performed with the "compare" function embedded in the R package "timeROC" (version 0.3 published in 2015-03-25).[15] All other statistical analyses were performed using SPSS 17.0 (IBM SPSS). All tests were bilateral, and $p < 0.05$ was considered statistically significant.

RESULTS

Identification of methylation sites, genes, and signaling pathways that are differentially expressed between thymomas and thymic carcinomas

Because a previous report showed that recurrences occurred more frequently in thymic carcinomas than thymomas, we first identified differentially methylated sites between thymomas and thymic carcinomas using the TCGA dataset. In total, 17 384 probes were identified as differentially methylated sites between thymic carcinomas and thymomas, of which 9021 probes were annotated (Figure 1a, Table S2). Among them, 1530 CpG sites were hypomethylated and 7491 CpG sites hypermethylated in thymic carcinomas, and these CpG sites covered 3460 genes (Table S2). We used Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analysis to demonstrate that these genes are closely associated with several signaling pathways,

including the neuroactive ligand-receptor interaction (hsa04080, adjusted $p = 1.86 \times 10^{-15}$), MAPK signaling pathway (hsa04024, adjusted $p = 1.60 \times 10^{-5}$), and cAMP signaling pathway (hsa04024, adjusted $p = 1.20 \times 10^{-10}$) (Figure 1b, Table S3).

Next, we determined the differentially expressed genes between thymomas and thymic carcinomas from the TGCA dataset. The results showed that 2267 genes were differentially expressed between thymomas and thymic carcinomas (Table S4). Among these genes, 1229 were downregulated, and 1038 were upregulated in thymic carcinomas compared to thymomas. After overlapping with genes that were found to have differentially methylated CpG sites in their promotor or the first exon regions, 40 genes and 179 genes were considered epigenetically upregulated and silenced, respectively, corresponding to a total of 509 functionally methylated CpG sites (Tables S5 and S6). The hierarchical clustering of these methylation β-values across all 124 patients is shown in Figure 1c.

Enrichment analysis revealed that epigenetically silenced by hypermethylation genes were mainly involved in tumorigenesis such as mammary stem cell, gastric cancer and prostate cancer (Table S7, Figure S1a). The genes impacted by hypomethylation were enriched in more diverse pathways, including thyroid cancer, TNF pathway, and glioblastoma mesenchymal (Table S7, Figure S1b).

This analysis revealed distinct methylation profiles that distinguish thymic carcinomas from thymomas. These findings also suggest that distinct methylation profiles may contribute to the malignant behavior of thymic carcinomas by dysregulating gene expression and signaling pathways.

Identification of candidate methylation sites for RFS prognosis of TETs using TCGA THYM dataset

In order to identify candidate methylation sites that had a potential impact on prognosis, we used univariate Cox regression to identify methylation sites that were closely related to RFS in TETs from 509 functional methylation sites. The results showed that 52 CpG sites were significantly associated with RFS in TETs (Table S8). Based on their possible involvement in the progression of carcinomas, as revealed by previously published articles,[16,17] cg20068620 in MAPK4 and cg18770944 in USP51 were ultimately selected as candidate methylation sites for further analysis (Table S9). The methylation β-values of 0.2030 in the cg20068620 CpG site and 0.3636 in the cg18770944 CpG site were chosen as cutoff values to categorize patients into low and high methylation subgroups by using curv_cutpoint in terms of RFS. Patients in the low methylation subgroup exhibited extremely superior RFS to those with high methylation in the cg20068620 CpG site (HR = 10.64, 95% CI: 1.309–86.42, $p = 0.027$, Log-rank $\chi^2 = 7.421$, $p = 6.448 \times 10^{-3}$). This result was more pronounced in the cg18770944 CpG site (HR = 16.94, 95% CI: 3.411–84.17,
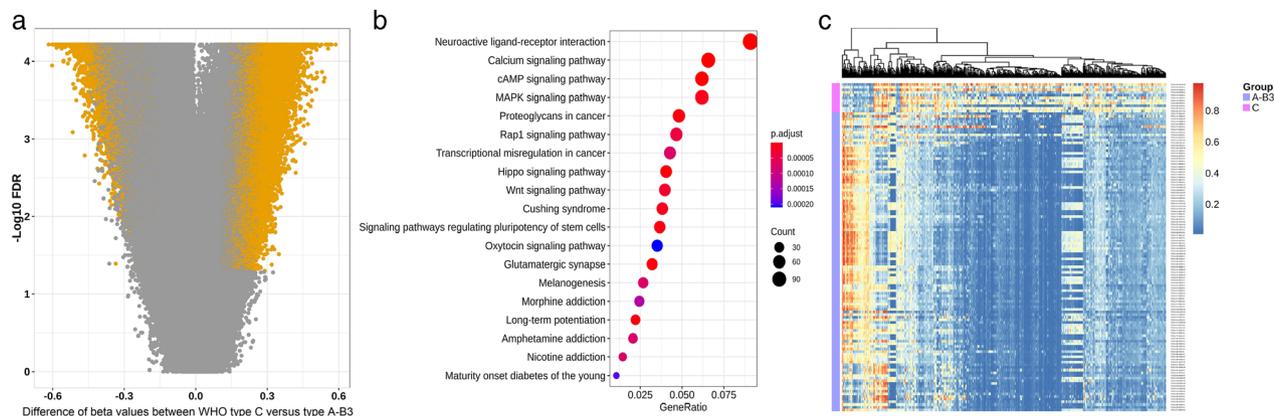
FIGURE 1 Epigenetic characteristics between thymoma with WHO classification type C and type A to B3. (a) Volcano plots showing significantly expressed methylation sites in 392 653 probes in the HumanMethylation450K array between thymoma with WHO classification type C and type A to B3. The yellow dots represent the significantly expressed methylation sites according to the criteria proposed in the "Materials and Methods". The abscissa axis represents the difference in methylation β-values of WHO classification type C versus WHO classification type A to B3, and the vertical axis represents the minus logarithm of crude p-values deduced from the Wilcoxon rank-sum test. (b) Dot plots showing the top 20 most significantly enriched pathways evaluated from the KEGG enrichment analysis. (c) A heatmap showing the methylation profiles of 509 functional methylation sites across all 124 cases. The first 11 rows represent patients with WHO classification type C and the rest of the WHO classification types A to B3

$p = 0.00054$, log-rank $\chi^2 = 20.878$, $p = 4.894 \times 10^{-6}$) (Figure 2a–b). Moreover, the whole population was further divided into three groups: patients with low methylation levels in both two candidate CpG sites were categorized into the low-risk group, patients with both high methylation levels into the high-risk group and the rest into the intermediate-risk group. This stratification had a strong prognostic capacity (log rank $\chi^2 = 24.839$, $p = 4.037 \times 10^{-6}$) (Figure 2c). Moreover, the prognostic efficacy of these two candidate CpG sites was further evaluated as a linear combination of coefficients of Cox regression of two CpG sties and corresponding β-values with the formula: $7.861 \times$ methylation β-value of cg20068620 +- $6.778 \times$ methylation β-value of cg18770944. Univariate Cox regression showed that the linear combination value was significantly associated with RFS (HR = 2.454, 95% CI: 1.436–4.193, $p = 0.00103$). After adjusting for age, sex, WHO classification, Masaoka stage, and adjuvant radiotherapy, the linear combination value as a continuous covariate was the only independent prognostic factor for RFS (HR = 2.728, 95% CI: 1.278–5.823, $p = 0.0095$) (Table S10). Thus, these two candidate methylation sites were finally chosen for further validation.

Validation of candidate methylation sites for the prognosis of TET patients with RFS

The results observed from the TCGA dataset were further confirmed in our patient cohort. Our data show that the recurrence rate of thymic carcinomas was 75.0% (9/12), which was significantly higher than the 22.9% recurrence rate of thymomas (19/83). The 3-, 5-, and 10-year RFS rates were 98.6, 92.8, and 50.4%, respectively, in thymoma patients and 83.3, 37.4, and 16.7%, respectively, in

thymic carcinoma patients. Consistently, the RFS time was significantly shorter in thymic carcinoma patients than in thymoma patients (log-rank $p < 0.001$), suggesting that thymic carcinoma recurs more frequently than thymoma.

Next, our patient cohort further confirmed the correlation of cg20068620 in *MAPK4* and cg18770944 in *USP51* with RFS of TETs. Consistent with TCGA dataset analysis results, our data also showed that the β values of both cg20068620 in *MAPK4* and cg18770944 in *USP51* were significantly increased in patients with thymic carcinoma compared to patients with thymomas (median (IQR), cg20068620: 65 (42–86) versus 20 (16–45), $p < 0.001$; cg18770944: 71 (62–99) versus 40 (25–66), $p < 0.001$). We also indicated significantly increased β-values of cg20068620 in *MAPK4* and cg18770944 in *USP51* in advanced stage (Masaoka stage III–IV) compared to early stage (Masaoka stage I–II) (median (IQR), cg20068620: 59 (20–64) versus 20 (14–29), $p < 0.001$; cg18770944: 66 (40–71) versus 32 (21–62), $p < 0.001$). To illustrate the prognostic efficiency, the entire population was categorized into two groups with low or high methylation levels according to optimal cut-off β-values 0.51 of cg20068620 and 0.67 of cg18770944, respectively. As shown in Figure 2d–f, in line with the results obtained from TCGA THYM dataset, the patient with low methylation levels in cg20068620 or cg18770944 had superior RFS than those with high methylation levels (cg20068620: Log-rank $\chi^2 = 83.260$, $p = 7.195 \times 10^{-20}$, HR = 94.57, 95% CI: 12.8–701.3, $p = 8.34 \times 10^{-6}$, cg18770944: Log-rank $\chi^2 = 49.465$, $p = 2.019 \times 10^{-12}$, HR = 9.867, 95% CI: 4.524–21.52, $p = 8.73 \times 10^{-9}$). The patients with high methylation in the both CpG sites had the worse RFS ($\chi^2 = 85.531$, $p = 2.674 \times 10^{-19}$). Taken together, these results suggested that hypermethylation of cg20068620 in *MAPK4* and
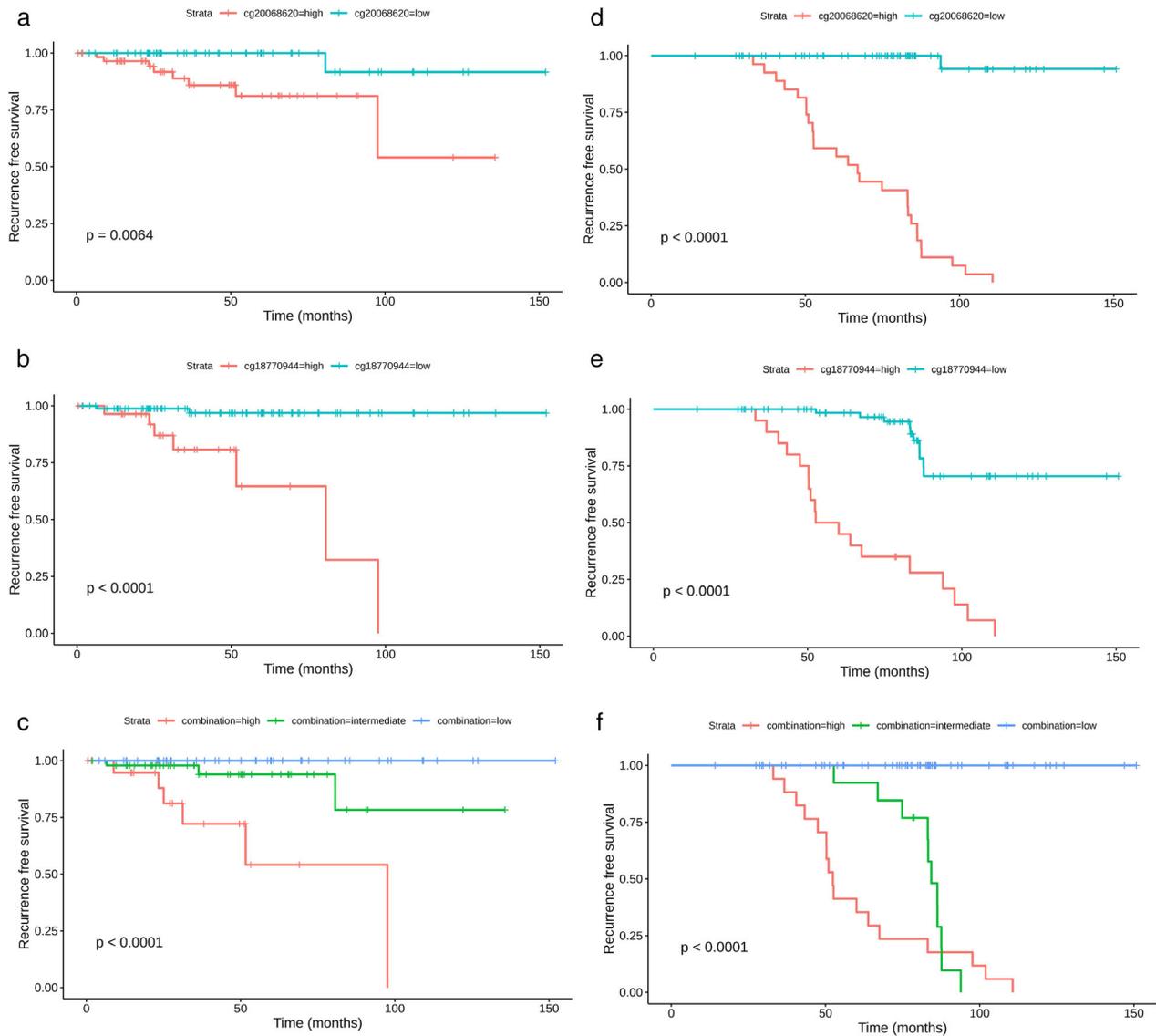
FIGURE 2 Kaplan–Meier curves show the methylation profiles of thymoma patients stratified into subgroups with different recurrence-free survival (RFS) in the TCGA thymoma dataset (a–c) and the validation set (d–f). (a) High methylation of cg20068620 in *MAPK4* appears to be associated with shorter RFS. (b) High methylation of cg18770944 in *USP51* is significantly associated with shorter RFS. (c) The patients were categorized into three subgroups according to the low and high methylation levels of cg20068620 in *MAPK4* and cg18770944 in *USP51*. Patients with low methylation levels in both two CpG sites were stratified into the low-risk subgroup, patients with high methylation levels in both two CpG sites into the high-risk subgroup, and the remaining patients were stratified into the intermediate-risk subgroup. (d) High methylation of cg20068620 in *MAPK4* at the optimal cutoff value of 0.51 appears to be associated with shorter RFS. (e) High methylation of cg18770944 in *USP51* at the optimal cutoff value 0.67 is significantly associated with shorter RFS. (f) The patients were categorized into three subgroups according to the low and high methylation levels of cg20068620 in *MAPK4* and cg18770944 in *USP51*. Patients with low methylation levels in both two CpG sites were stratified into the low-risk subgroup, patients with high methylation levels in both two CpG sites into the high-risk subgroup, and the remaining patients were stratified into the intermediate-risk subgroup

cg18770944 in *USP51* is closely associated with the aggressiveness of TETs.

In addition, univariate Cox regression analysis revealed that age, WHO classification, Masaoka stage, adjuvant radiotherapy, adjuvant chemotherapy and two candidate methylation sites (cg20068620 and cg18770944) were significantly associated with RFS (Table 4). However, multivariable Cox regression with forwarding selection for covariates showed that only WHO classification and methylation β-values of cg20068620 in *MAPK4* were independent

prognostic factors for RFS in TETs (Table 4). Therefore, the recurrence risk score (RRS) was calculated using these two factors as a weighted linear combination: $1.468 \times$ WHO classification (C vs. A-B3) $+ 0.097 \times$ methylation β-value in cg20068620. The median value and IQR of RRS in the whole validation set was 2.62 (1.55–5.72). The whole population of the validation set was categorized into low- and high-risk subgroups according to the median value of RRS. This grouping had robust discriminative efficiency for the prognosis of RFS. The RFS in the high-risk group was

TABLE 4 Results from univariate and stepwise Cox regression

	HR (95% CI)	*p*-value	HR (95% CI)	*p*-value
Gender (male vs. female)	0.701 (0.332–1.477)	0.350		
Age[a]	0.972 (0.952–0.993)	0.009		
History myasthenia gravis (yes vs. no)	1.710 (0.785–3.722)	0.177		
WHO classification (C vs. A-B3)	5.598 (2.510–12.487)	<0.001	4.339 (1.662–11.329)	0.003
Masaoka stage (III–IV vs. I–II)	27.118 (6.397–114.96)	<0.001		
Adjuvant radiotherapy (yes vs. no)	0.304 (0.105–0.879)	0.028		
Adjuvant chemotherapy (yes vs. no)	0.186 (0.044–0.785)	0.022		
cg20068620[a]	2.653 (1.941–3.626)[b]	<0.001	2.635 (1.907–3.642) ‡	<0.001
cg18770944[a]	2.718 (1.958–3.771)[b]	<0.001		

[a]As continuous variables into the equations.
[b]The hazard ratios and corresponding 95% confidential intervals was calculated for per 10 percentage increment of methylation β-value.
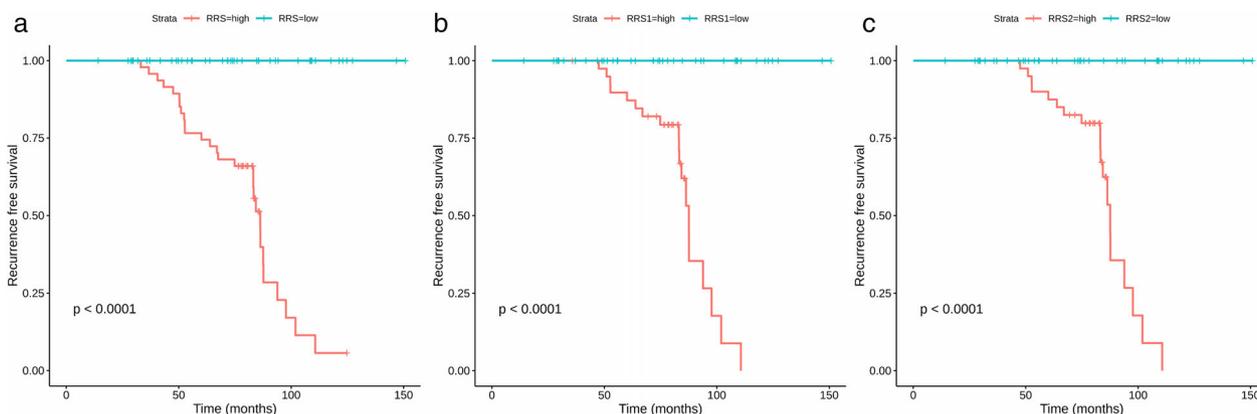

FIGURE 3 Kaplan–Meier (K-M) curves show prognostic efficiency of recurrence risk score (RRS) in the whole validation set (a) and the subset consisted of patients of WHO classification type A-B3 (b–c). (a) K-M curves showing dramatic difference in recurrence-free survival (RFS) between subgroups with different recurrence risks according to RRS in the whole validation set. (b) K-M curves showing dramatic difference in RFS between subgroups with different recurrence risks from model 1. (c) K-M curves showing dramatic difference in RFS between subgroups with different recurrence risks from model 2

TABLE 5 Area under curves and comparison of different prognostic factors for 3, 5, and 10 years RFS

Prognostic factors	3 years	5 years	10 years
WHO classification	0.9368 (0.902–0.972)	0.7770 (0.629–0.925)	0.5430 (0.393–0.693)
Masaoka stage	0.7874 (0.735–0.840)	0.8209 (0.763–0.879)	0.8907 (0.750–1.000)
cg20068620	0.9943 (0.983–1.000)	0.9341 (0.876–0.992)	1.000 (1.000–1.000)
cg18770944	0.9770 (0.945–1.000)	0.9479 (0.881–1.000)	1.000 (1.000–1.000)
RRS	1.000 (1.000–1.000)	0.9590 (0.911–1.000)	1.000 (1.000–1.000)
Adjusted *p* (RRS vs. WHO classification)	1.24×10^{-3}	1.05×10^{-2}	6.98×10^{-9}
Adjusted *p* (RRS vs. Masaoka stage)	3.33×10^{-15}	1.34×10^{-4}	0.3147

significantly shorter than that in the low-risk group (median RFS: 84.3 months vs. not reached, log-rank $p = 1.04 \times 10^{-9}$) (Figure 3a). In particular, all 28 recurrence events occurred in the high-risk group. Multivariable Cox regression with forward selection for covariates verified that RRS was the only prognostic factor for RFS among all clinicopathological characteristics (HR = 2.718, 95% CI: 2.005–3.685, $p = 1.173 \times 10^{-10}$). Furthermore, time-dependent ROC curve analysis revealed that RRS was overwhelmingly superior to WHO classification for predicting 3-, 5-, and 10-year RFS and Masaoka stage for 3- and 5-year RFS (Table 5, Figure S2a–c). These results suggested that

TABLE 6 Summarization of prognostic models developed in the subgroup comprising WHO classification type A to type B3

Prognostic models	Covariates	HR 95% CI	Wald p-value	- 2 log likelihood	χ^2/p
Model 0	Masaoka stage (III–IV vs. I–II)	26.289 (5.986–115.46)	<0.001	99.769	–
Model 1	Masaoka stage (III–IV vs. I–II)	5.808 (1.164–28.993)	0.032	89.835	9.934/0.001623[b]
	cg18770944	1.978 (1.188–3.295)[a]	0.009		
Model 2	cg20068620	2.569 (1.732–3.809)[a]	<0.001	83.704	–

[a]The hazard ratios and corresponding 95% confidential intervals was calculated for per 10 percentage increment of methylation β-value.
[b]The probability was calculated based on likelihood ratio test compared with model 0 which merely contains Masaoka stage as independent variate.

integrating the β values of cg20068620 in *MAPK4* into the commonly adopted clinical prognostic factors can yield a more precise prognosis in patients with TETs.

Subgroup analysis and model development in thymomas

Most patients with TETs are diagnosed with thymomas, some patients with thymomas still experience recurrence, and no biomarkers can predict the recurrence of thymomas. Thus, we used our cohort ($n = 83$ thymoma patients) to investigate the possibility of cg20068620 in *MAPK4* and cg18770944 in *USP51* as predictors of thymoma recurrence. Multivariable Cox regression with forwarding selection demonstrated that Masaoka stage (III–IV vs. I–II) was the only prognostic factor for RFS in thymomas (HR = 26.289, 95% CI: 5.986–115.46, $p = 1.491 \times 10^{-5}$) among the clinicopathological features that showed a significant association with RFS in univariate Cox regression analysis. The combination of the Masaoka stage and cg18770944 provided more prognostic information than Masaoka stage alone based on the likelihood ratio test (Table 6). However, Cox regression with the Masaoka stage and cg20068620 as covariates showed that the Masaoka stage was no longer significant and that cg20068620 was strongly associated with RFS, suggesting that cg20068620 alone could more precisely predict RFS in this clinical set (Table 6). The patients with low- and high-risk in this subgroup, according to median values of RRS calculated from model 1 and model 2, showed significant differences in RFS (Figure 3b–c). However, the AUCs for predicting 5- and 10-year RFS of model 1 were not significantly higher than those of Masaoka stage alone (mean 95% CI: 0.943 (0.862–1.000) versus 0.847 (0.789–0.904), $p = 0.055$; 1.000 (1.000–1.000) versus 0.951 (0.882–1.000), $p = 0.303$). The same conclusion held true for model 2 (mean 95% CI: 0.909 (0.815–1.000) versus 0.847 (0.789–0.904), $p = 0.356$; 1.000 (1.000–1.000) versus 0.951 (0.882–1.000), $p = 0.296$).

DISCUSSION

The WHO classification divides thymic epithelial tumors into thymomas and thymic carcinoma histologically, while clinical data show that thymic carcinoma is more aggressive than thymomas. For example, the 5-year survival rate of patients with thymomas is approximately 78%; however, in patients with thymic carcinomas, it is only approximately 40%.[18] According to Khandelwal et al., the recurrence rate (including metastasis) of thymoma patients is 20%, while the recurrence rate of thymic carcinoma patients is as high as 67%.[19] Consistent results were also observed in the present study. Notably, such clinical outcome differences between thymic carcinoma and thymomas may be caused by their different molecular biology.[20] Comprehensive genomic analysis suggests that thymic carcinoma is molecularly distinct from thymomas.[21] Enkner et al. also observed genetic differences between thymic carcinomas and thymomas.[22] In the present study, we demonstrated that thymic carcinoma and thymomas have distinct DNA methylation profiles and that distinct DNA methylation may cause different gene expression patterns and signaling pathway activation. Consistently, Hirose et al. reported that aberrant DNA methylation was more frequent in thymic carcinomas than in thymomas.[23] Altogether, these findings suggested that thymic carcinomas are highly aggressive tumors and that aberrant DNA methylation may be closely associated with the malignant behavior of thymic carcinoma.

Studies have shown that metastatic or local recurrence occurs in 5%–31% of TET patients, suggesting that recurrence is one challenge of TET treatment.[19,24] Thus, reliable and accurate predictive markers to identify which subsets of TET patients are vulnerable to recurrence are urgently needed.[25] Previously, Marx et al. reported that the WHO classification is a prognostic factor for recurrence and survival in TET patients,[4] but the findings are controversial. Other studies have shown that WHO classification does not significantly predict TET recurrence.[25] In this study, we used the TCGA dataset and the cohort analysis demonstrated that the WHO classification and DNA methylation site cg20068620 in *MAPK4* are independent predictors of RFS in TET patients, and combining the WHO classification and cg20068620 in *MAPK4* can more accurately predict the recurrence of TET patients than the WHO classification alone. Notably, our independent cohort analysis showed that all recurrent cases occurred in the high recurrence risk group set according to the combination of the WHO classification and DNA methylation site cg20068620 in *MAPK4*. These findings suggest that combining the WHO classification and cg20068620 in *MAPK4* may be a valuable method for predicting recurrence among TET patients. However, the

sample size of this study was limited. Thus, further confirmation in larger sample size is needed before clinical application. In particular, more thymic carcinoma patients need to be included. It has been reported that several CpG sites in the MAPK4 promoter region, including cg20068620, have been associated with overall survival in low-grade glioma in a previously published article.[17] These results suggested that epigenetic regulation of kinases, including MAPK4, may play an essential role in the biology of carcinomas.

Although the survival time of thymoma patients is long, 20% of thymoma patients experience postoperative recurrence. In the present study, we identified that Masaoka stage was closely correlated with thymoma recurrence. We also indicated that the methylation sites cg20068620 in *MAPK4* and cg18770944 in *USP51* were associated with thymoma recurrence, suggesting that Masaoka stage, cg20068620 in *MAPK4* and cg18770944 in *USP51* are useful candidates as predictors of thymoma recurrence. USP51 is identified as a ZEB1 deubiquitinase in serval reports. USP51 can stabilize ZEB1 protein to enhance epithelial-mesenchymal transition and metastasis.[16,26] However, the impact of epigenetic regulation of USP51 on cancer progression has not been reported. Recently, several genes with hypermethylation in thymic carcinoma compared with thymoma have been identified, including GAD1, GNG4, GHST, HOXD9, and SALL3.[27,28] Soejima et al. reported that patients with TETs with high GAD1 DNA hypermethylation and high mRNA and protein expression levels had significantly shorter relapse-free survival rates than those with low levels.[27] In our analysis, GAD1 was also identified as a gene impacted by hypermethylation in thymic carcinoma (Table S2). Further investigation into the mechanism by which methylation of these genes regulates malignant progression was needed.

In conclusion, our findings indicate that different DNA methylation patterns are closely associated with the malignancy of TETs and that the combination of the WHO classification and methylation site cg20068620 in *MAPK4* is better than the WHO classification alone in terms of the prognostic value of TET recurrence. In addition, we identified that the Masaoka stage, cg20068620 in *MAPK4* and cg18770944 in *USP51*, are valuable candidates as prognostic factors for the recurrence of thymomas.

CONFLICT OF INTEREST
The authors have no conflict of interest.

ORCID
He Xiao https://orcid.org/0000-0003-2298-1121
Chuan Chen https://orcid.org/0000-0003-0047-9560

REFERENCES

1. Kuwano H, Amano J, Yokomise H. Thoracic and cardiovascular surgery in Japan during 2010: annual report by the Japanese Association for Thoracic Surgery. Gen Thorac Cardiovasc Surg. 2012;60(10): 680–708.
2. Strobel P, Marx A, Zettl A, Muller-Hermelink HK. Thymoma and thymic carcinoma: an update of the WHO classification 2004. Surg Today. 2005;35(10):805–11.
3. Lee GD, Kim HR, Choi SH, Kim Y, Kim DK, Park S. Prognostic stratification of thymic epithelial tumors based on both Masaoka-Koga stage and WHO classification systems. J Thorac Dis. 2016;8(5): 901–10.
4. Marx A, Chan JKC, Coindre J, Detterbeck F, Girard N, Harris NL, et al. The 2015 World Health Organization classification of tumors of the thymus continuity and changes. J Thorac Oncol. 2015;10(10): 1383–95.
5. Travis WD, Brambilla E, Burke AP, Marx A, Nicholson AG. Introduction to the 2015 World Health Organization classification of tumors of the lung, pleura, thymus, and heart. J Thorac Oncol. 2015;10(9): 1240–2.
6. Girard N. Thymic epithelial tumours: from basic principles to individualised treatment strategies. Eur Respir Rev. 2013;22(127):75–87.
7. Miura K, Hamanaka K, Uehara T, Momose M, Kanai Y, Matsuoka S, et al. Bone morphogenetic protein-7 expression reflects the high proliferative ability and aggressiveness of thymic epithelial tumors. J Thorac Dis. 2020;12(3):448–56.
8. Hishida T, Nomura S, Yano M, Asamura H, Yamashita M, Ohde Y, et al. Long-term outcome and prognostic factors of surgically treated thymic carcinoma: results of 306 cases from a Japanese Nationwide Database Study. Eur J Cardiothorac Surg. 2016;49(3):835–41.
9. Anastasiadi D, Esteve-Codina A, Piferrer F. Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. Epigenetics Chromatin. 2018;11(1):37.
10. Ahuja N, Sharma AR, Baylin SB. Epigenetic therapeutics: a new weapon in the war against cancer. Annu Rev Med. 2016;67:73–89.
11. Long J, Chen P, Lin J, Bai Y, Yang X, Bian J, et al. DNA methylation-driven genes for constructing diagnostic, prognostic, and recurrence models for hepatocellular carcinoma. Theranostics. 2019;9(24): 7251–67.
12. Li S, Yuan Y, Xiao H, Dai J, Ye Y, Zhang Q, et al. Discovery and validation of DNA methylation markers for overall survival prognosis in patients with thymic epithelial tumors. Clin Epigenetics. 2019; 11(1):38.
13. Kont V, Murumagi A, Tykocinski L, Kinkel SA, Webster KE, Kisand K, et al. DNA methylation signatures of the AIRE promoter in thymic epithelial cells, thymomas and normal tissues. Mol Immunol. 2011;49(3):518–26.
14. Li YQ, Tian YM, Tan SH, Liu MZ, Kusumawidjaja G, Ong EHW, et al. Prognostic model for stratification of Radioresistant nasopharynx carcinoma to curative salvage radiotherapy. J Clin Oncol. 2018; 36(9):891–9.
15. Blanche P, Dartigues J, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Stat Med. 2013; 32(30SI):5381–97.
16. Zhang Z, Li J, Ou Y, Yang G, Deng K, Wang Q, et al. CDK4/6 inhibition blocks cancer metastasis through a USP51-ZEB1-dependent deubiquitination mechanism. Signal Transduct Target Ther. 2020;5(1):25.
17. Southekal S, Mishra NK, Guda C. Pan-cancer analysis of human kinome gene expression and promoter DNA methylation identifies dark kinase biomarkers in multiple cancers. Cancers. 2021;13(6):1189.
18. Ettinger DS, Riely GJ, Akerley W, Borghaei H, Chang AC, Cheney RT, et al. Thymomas and thymic carcinomas clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2013;11(5):562–76.
19. Khandelwal A, Sholl LM, Araki T, Ramaiya NH, Hatabu H, Nishino M. Patterns of metastasis and recurrence in thymic epithelial tumours: longitudinal imaging review in correlation with histological subtypes. Clin Radiol. 2016;71(10):1010–7.

20. Kelly RJ. Thymoma versus Thymic carcinoma: differences in biology impacting treatment. J Natl Compr Canc Netw. 2013;11(5):577–83.

21. Girard N, Shen R, Guo T, Zakowski MF, Heguy A, Riely GJ, et al. Comprehensive genomic analysis reveals clinically relevant molecular distinctions between thymic carcinomas and Thymomas. Clin Cancer Res. 2009;15(22):6790–9.

22. Enkner F, Pichlhoefer B, Zaharie AT, Krunic M, Holper TM, Janik S, et al. Molecular profiling of thymoma and thymic carcinoma: genetic differences and potential novel therapeutic targets. Pathol Oncol Res. 2017;23(3):551–64.

23. Hirose Y, Kondo K, Takizawa H, Nagao T, Nakagawa Y, Fujino H, et al. Aberrant methylation of tumour-related genes in thymic epithelial tumours. Lung Cancer. 2009;64(2):155–9.

24. Liu H, Gu Z, Qiu B, Detterbeck FC, Roden AC, Ruffini E, et al. A recurrence predictive model for thymic tumors and its implication for postoperative management: a Chinese Alliance for research in Thymomas database study. J Thorac Oncol. 2020;15(3):448–56.

25. Su Y, Chen Y, Tian Z, Lu C, Chen L, Ma X. lncRNAs classifier to accurately predict the recurrence of thymic epithelial tumors. Thorac Cancer. 2020;11(7):1773–83.

26. Zhou ZC, Zhang PJ, Hu XY, Kim JC, Yao F, Xiao ZN, et al. USP51 promotes deubiquitination and stabilization of ZEB1. Am J Cancer Res. 2017;7(10):2020–31.

27. Soejima S, Kondo K, Tsuboi M, Muguruma K, Tegshee B, Kawakami Y, et al. GAD1 expression and its methylation as indicators of malignant behavior in thymic epithelial tumors. Oncol Lett. 2021;21(6):483.

28. Kishibuchi R, Kondo K, Soejima S, Tsuboi M, Kajiura K, Kawakami Y, et al. DNA methylation of GHSR, GNG4, HOXD9 and ASLL3 is a common epigenetic alteration in thymic carcinoma. Int J Oncol. 2020;56(1):315–26.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.