

RESEARCH

Open Access



Novel molecular typing reveals the risk of recurrence in patients with early-stage papillary thyroid cancer

Mingyu Sun^{1†}, Bingqing Zhao^{2†}, Tao Chen^{3†}, Lijun Yao⁴, Xiaoxin Li⁵, Shaojun Hu⁴, Chengling Chen^{1*}, Xinbao Gao^{6*} and Chuangang Tang^{1*}

Abstract

Background Papillary thyroid cancer (PTC) is an indolent disease with a favorable prognosis but characterized by a high recurrence rate. We aimed to improve precise stratification of recurrence risk in PTC patients with early stage using multi-gene signatures.

Patients and methods The present study was performed using data from The Cancer Genome Atlas (TCGA) and multi-center datasets. Unsupervised consensus clustering was used to obtain the optimal molecular subtypes and least absolute shrinkage and selection operator (LASSO) analysis was performed to identify potential genes for the construction of recurrence signature. Kaplan-Meier survival analysis and the log-rank test was used to detect survival differences. Harrells concordance index (C-index) was used to assess the performance of the DNA damage repair (DDR) recurrence signature.

Results Through screening 8 candidate gene sets, the entire cohort was successfully stratified into two recurrence-related molecular subtypes based on DDR genes: DDR-high subtype and DDR-low subtype. The recurrence rate of DDR-high subtype was significantly lower than DDR-low subtype [HR = 0.288 (95%CI, 0.084–0.986), $P = 0.047$]. Further, a two-gene DDR recurrence signature was constructed, including PER1 and EME2. The high-risk group showed a significantly worse recurrence-free survival (RFS) than the low-risk group [HR = 10.647 (95%CI, 1.363–83.197), $P = 0.024$]. The multi-center data demonstrated that proportion of patients with low expression of PER1 and EME2 was higher in the recurrence group than those in the non-recurrence group.

Conclusions These findings could help accurately and reliably identify PTC patients with high risk of recurrence so that they could receive more radical and aggressive treatment strategies and more rigorous surveillance practices.

[†]Mingyu Sun, Bingqing Zhao and Tao Chen contributed equally to this work.

*Correspondence:
Chengling Chen
276480517@qq.com
Xinbao Gao
499581400@qq.com
Chuangang Tang
tcg3711@163.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords Thyroid cancer, Recurrence, DNA damage repair, Risk model

Background

Thyroid cancer is the most common malignant tumor of the endocrine system, and can be divided pathologically into four major types: papillary thyroid cancer (PTC), follicular thyroid cancer (FTC), medullary thyroid cancer (MTC), and anaplastic thyroid cancer (ATC) [1]. Both PTC and FTC originate from the thyroid follicular cells and carry a favorable survival, collectively referred to as differentiated thyroid cancer (DTC) [2]. In recent decades, the incidence of thyroid cancer has been annually increased primarily due to the increase of the diagnosis of PTC at early stage [3, 4]. The latest American Joint Committee on Cancer (AJCC) staging system for PTC modified the cut-off value of age from 45 to 55, which further increased the proportion of early-stage PTC [5]. According to statistics, PTC make up the vast majority of all thyroid cancers and PTC with stage I accounted for more than 80% of all PTC [6, 7]. The overall 10-year survival rate of PTC is approximately 90% [8], and the 10-year survival rate of PTC with stage I is up to 95% [4]. However, the recurrence rate is relatively high, ranging from 10 to 30%, even if patients are in the early stage of PTC [9, 10]. Thewjitcharoen et al. [6] showed that 23.7% of PTC patients with stage I fell into the high risk of recurrence group according to the 2015 American Thyroid Association (ATA) risk classification, indicating that PTC patients with the same survival prognosis also had heterogeneous tumors with different risk of recurrence. Thus, there is a great need to stratify early-stage PTC patients by the risk of recurrence more accurately and reduce their recurrence risk with personalized treatment strategies.

Multiple studies have been dedicated to investigating the risk of recurrence in PTC and have successfully developed risk models. The proto-oncogenes *PBF* (*PTTG1IP*) and *PTTG1* were linked to recurrence in PTC and the *PBF/PTTG1* signature displayed a notable discriminatory power [11, 12]. A new combinatorial 5-gene risk score classifier incorporated mRNAs and microRNAs, which conducted thorough explorations to achieve a more accurate prediction of recurrence in PTC [13]. Read et al. [14] developed a 13-gene risk score classifier to assess the risk of recurrence in PTC with ablative radioiodine treatment. In their most recent work, they further identified an AP2 gene-related risk score classifier as an effective tool for predicting the recurrence of BRAF-mutant PTC and PTC with ablative radioiodine treatment [15]. However, these studies did not differentiate subgroups of PTC based on clinical staging, which was unfavorable for facilitating appropriate personalized management for PTC patients. Therefore, we aimed to improve precise stratification of

recurrence risk in PTC patients with stage I aged <55 years using multi-gene signatures.

Patients and methods

Dataset

Patients diagnosed as thyroid cancer were included from the TCGA-THCA dataset, downloaded from The Cancer Genome Atlas (TCGA, <http://tcga-data.nci.nih.gov/tcga/>). The exclusion criteria were as follows: (1) diagnosed as FTC, MTC or ATC; (2) 55 years old or older; (3) distant metastasis; (4) unavailable recurrence information. RNA sequences (FPKM) of 191 samples with basic clinical characteristics, including gender, age, pathological T category, pathological N category, ATA recurrence risk stratification, and recurrence status, were downloaded. Gene names were converted using the Ensemble database (<http://www.ensembl.org/index.html>).

A series of gene sets potentially involved in key pathological mechanisms in PTC were collected from literature, including DNA damage repair (DDR)-related gene set, immune-related gene set, energy metabolism-related gene set, hypoxia-related gene set, ferroptosis-related gene set, lipid metabolism-related gene set, endothelial-mesenchymal transition (EMT)-related gene set, and m6A-related gene set [16–21]. Detailed gene sets were provided in Supplementary Table 1.

Consistency analysis

Unsupervised consensus clustering was performed using R package ConsensusClusterPlus. 80% of the overall samples were randomly subsampled with 100 iterations. The number of clusters ranged from two to six. The cumulative distribution function (CDF) curves were used to obtain the optimal number of clusters. The relative changes in the area under the CDF curves were also calculated. The heatmap of clustering and gene expression was drawn using R package pheatmap. Genes with standard deviation (SD) less than 0.1 were not shown in the gene expression heatmap. Kaplan-Meier survival analysis and the log-rank test was performed to detect survival differences between two or multiple subgroups using R package survival and survminer.

Least Absolute Shrinkage and Selection Operator (LASSO) and Receiver Operating Characteristic (ROC)

LASSO regression analysis was performed to identify potential recurrence-associated genes for the construction of DDR-signature. A 10-fold cross-validation scheme was applied. Each gene was assigned a coefficient and the risk score was calculated as follows: Risk score = $\beta_1 \times \text{expression mRNA}_1 + \beta_2 \times \text{expression mRNA}_2$

+... + β_6 *expression mRNA⁶. More and more coefficients were set gradually to zero, and the top contributing recurrence-associated genes would be screened. The accuracy of DDR-signature to predict the probable 1-year, 5-year and 10-year recurrence rates was evaluated by the ROC curve and the area under the ROC curve (AUC). LASSO regression analysis and ROC curve analysis were performed using R package glmnet and timeROC, respectively. Harrells concordance index (C-index) was used to assess the performance of the DDR recurrence signature and ATA recurrence risk stratification.

Immunohistochemistry

PTC patients with recurrence were collected between 2014 and 2016 from multi-center. All the patients from the multicenter datasets were followed up for at least 3 years. Patients with recurrence occurring within 6 months were excluded.

All paraffin-embedded tissues were assembled into tissue chips and immunohistochemistry was performed to detect the protein expression levels of DDR-signature genes (*PER1* and *EME2*). The primary antibodies were Rabbit anti-*PER1* (BOSTER Biological Technology, Wuhan, China; 1:100) and Rabbit anti-*EME2* (Bioss Inc., Beijing, China; 1:100). Secondary antibodies were used at a dilution of 1:10000. The Research Resource Identifier (RRID) numbers of primary and secondary antibodies

Table 1 Basic clinical characteristics of PTC patients with stage I aged < 55 years

Characteristics	PTC (n = 191)
Age (median)	39
Gender	
Female	146 (76.4%)
Male	45 (23.6%)
Pathological T category	
T1	61 (31.9%)
T2	64 (33.5%)
T3	63 (33.0%)
T4	2 (1.1%)
TX	1 (0.5%)
Pathological N category	
N0	92 (48.2%)
N1	92 (48.2%)
NX	7 (3.6%)
Pathological M category	
M0	191 (100%)
ATA recurrence risk stratification *	
Low	61 (31.9%)
Intermediate	92 (48.1%)
High	2 (1%)
Recurrence	
Yes	11 (5.8%)
No	180 (94.2%)

*36 cases missing

were provided in Supplementary Table 2. The tissue chips were evaluated by two independent experienced pathological physicians and an additional pathological physician was included in case of inconsistency. Pathological physicians did not have prior knowledge of basic clinical characteristics and recurrence status. The immunohistochemical staining score was calculated by summing the percentage score (proportion of positive stained cells) and the staining intensity score. The percentage score was defined as 0 (0%), 1 (>0, ≤25%), 2 (>25%, ≤50%), 3 (>50%, ≤75%), and 4 (>75%, ≤100%). The staining intensity score was defined as 0 (negative), 1 (weak), 2 (moderate), and 3 (strong). The entire cohort was ranked according to immunohistochemistry scores from low to high. The top 50% patients were considered as high expression. The bottom 50% patients were considered as low expression. Written consent was obtained from all participants. Samples were collected and disposed of in accordance with WHO guidelines. The study was approved by the Ethics Committee of Suzhou Ninth People's Hospital.

Statistical analysis

Statistical analysis was performed with R software version 3.6.2 (The R Foundation). Hazard ratio (HR) and its 95% confidence interval (95% CI) were calculated by the log-rank test and univariate Cox regression analysis. Multivariate Cox regression analysis was used to determine the independent predictors. P value less than 0.05 was considered statistically significant.

Results

Basic clinical characteristics

A total of 191 PTC patients with stage I aged < 55 years were included from the TCGA database (Table 1). The age ranged from 15 to 54 years with the median age of 39 years. Females were responsible for the majority (76.4%, 146/191). The entire cohort was almost equally distributed across three pathological T categories. The number of PTC patients without LN metastasis was the same as those with LN metastasis. The follow-up period ranged from 0 to 14.1 years with the median follow-up period of 2.5 years. 11 patients had recurrences over the course of follow-up.

Molecular typing based on DDR genes

Using the above 8 gene sets, we performed unsupervised consensus clustering to identify distinct subgroups of early PTC with different risks of recurrence. No statistically significant different subgroups were identified with the exception of DDR-based subgroups [HR=0.288 (95%CI, 0.084–0.986), $p=0.047$] (Fig. 1). This suggested that DDR might play a crucial role in tumor recurrence of early-stage PTC. According to the CDF values and the R package SigClust evaluation, $k=2$ was determined as the

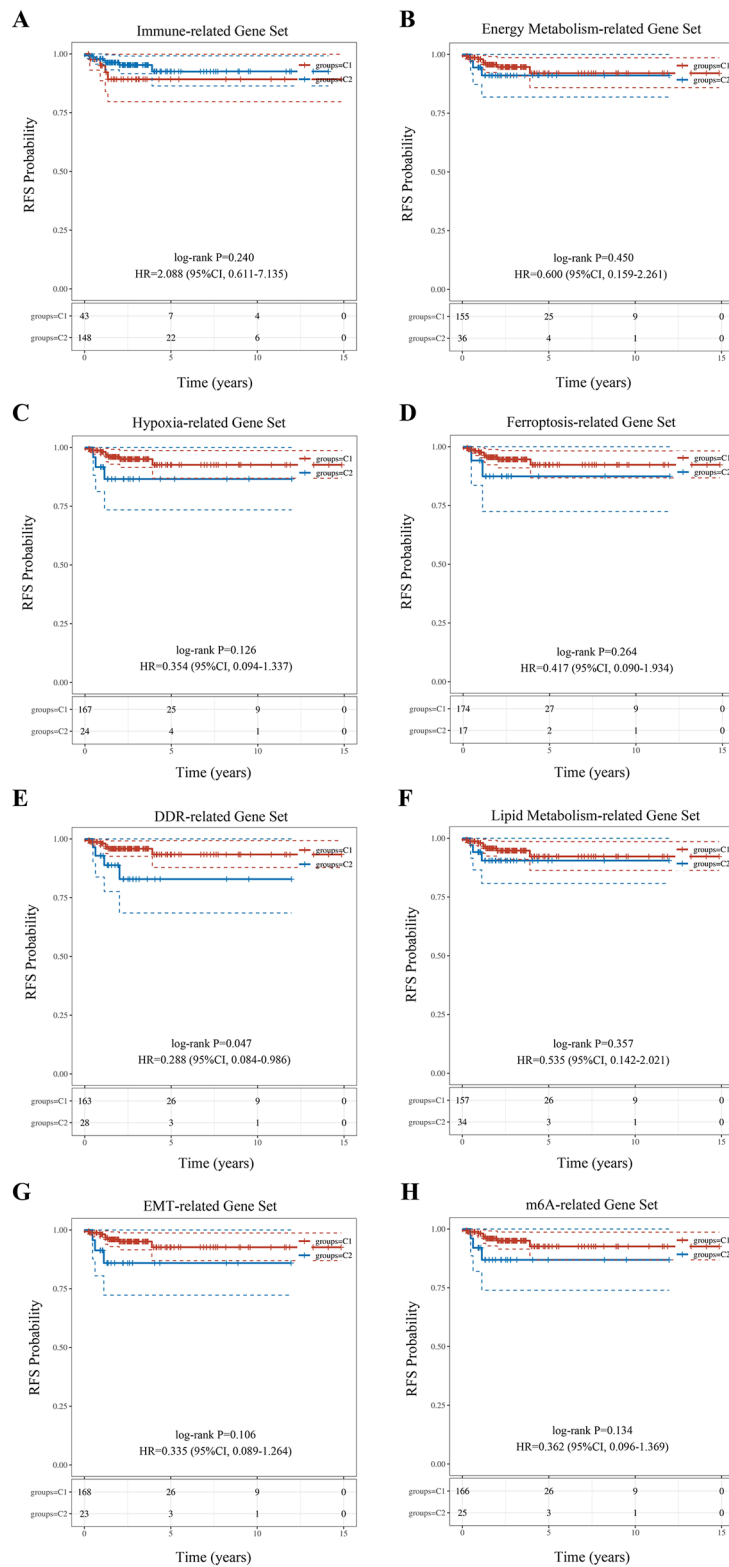


Fig. 1 Survival curves categorized by 8 different gene sets. **A** Survival curve categorized by immune-related gene set. **B** Survival curve categorized by energy metabolism-related gene set. **C** Survival curve categorized by hypoxia-related gene set. **D** Survival curve categorized by ferroptosis-related gene set. **E** Survival curve categorized by DDR-related gene set. **F** Survival curve categorized by lipid metabolism-related gene set. **G** Survival curve categorized by EMT-related gene set. **H** Survival curve categorized by m6A-related gene set

best clustering solution (Fig. 2A and B). The consensus clustering matrix showed that $k=2$ could clearly stratify samples into two apparently different clusters (Fig. 2C). Cluster 1 (C1) comprised 163 samples and Cluster 2 (C2) comprised 28 samples. As shown in the gene expression heatmap of DDR (Fig. 2D), the majority of DDR genes were highly expressed in C1 and associated with a relatively low risk of recurrence (named DDR-high subtype). Conversely, the majority of DDR genes were down-regulated in C2 and associated with a relatively high risk of recurrence (named DDR-low subtype).

Construction of DDR recurrence signature

According to the expression level of each DDR gene, the patient cohort was correspondingly classified into high-expression and low-expression groups. The differences

in RFS between the two groups were evaluated using Kaplan–Meier survival analysis with log-rank test. The results revealed 6 DDR genes with prognostic significance in relation to patient RFS, including *EME2*, *MNATI*, *MUS81*, *PER1*, *POLM*, and *RAD9A* (Fig. 3, $p < 0.05$). Interestingly, the lower expression levels of these DDR genes were associated with a higher probability of experiencing recurrence in early-stage PTC. The detailed information of all DDR genes with HRs and 95%CI was provided in Supplementary Table 3. To be clear, the p -values shown were unadjusted values.

Based on the above 6 DDR genes in relation to RFS, a LASSO regression model was constructed to develop a DDR recurrence signature. The minimal λ was 0.0164, and two genes (*PER1* and *EME2*) were screened for the final signature (Fig. 4A and B). The recurrence risk score

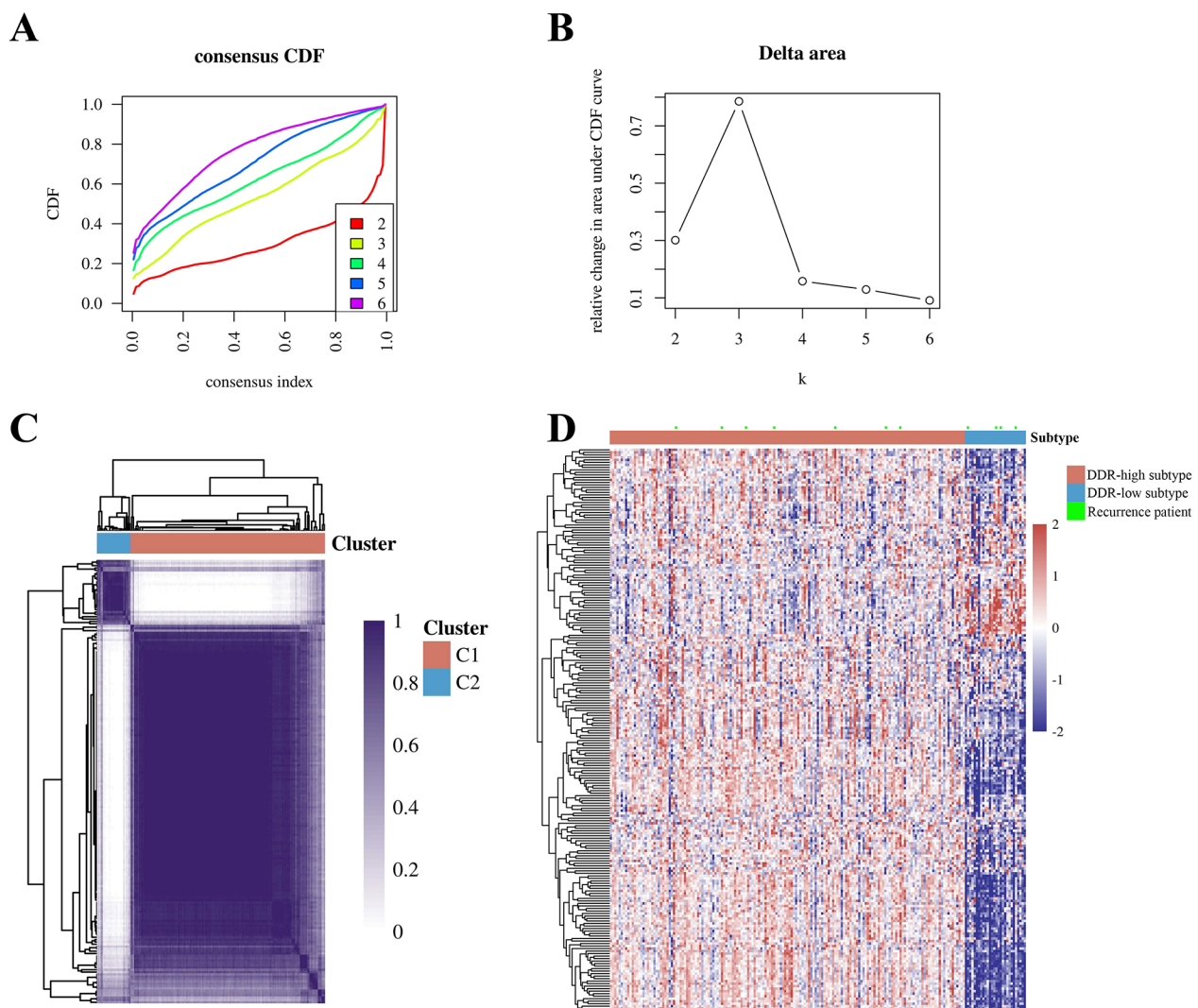


Fig. 2 Two molecular subtypes categorized by DDR-related gene set. **A** Consensus clustering model with CDF for $k=2-6$. **B** Relative change in the area under the CDF curve $k=2-6$. **C** Entire cohort was divided into two distinct clusters ($k=2$). **D** Heatmap of gene expression in two distinct clusters (cluster 1 was named DDR-high subtype and cluster 2 was named DDR-low subtype)

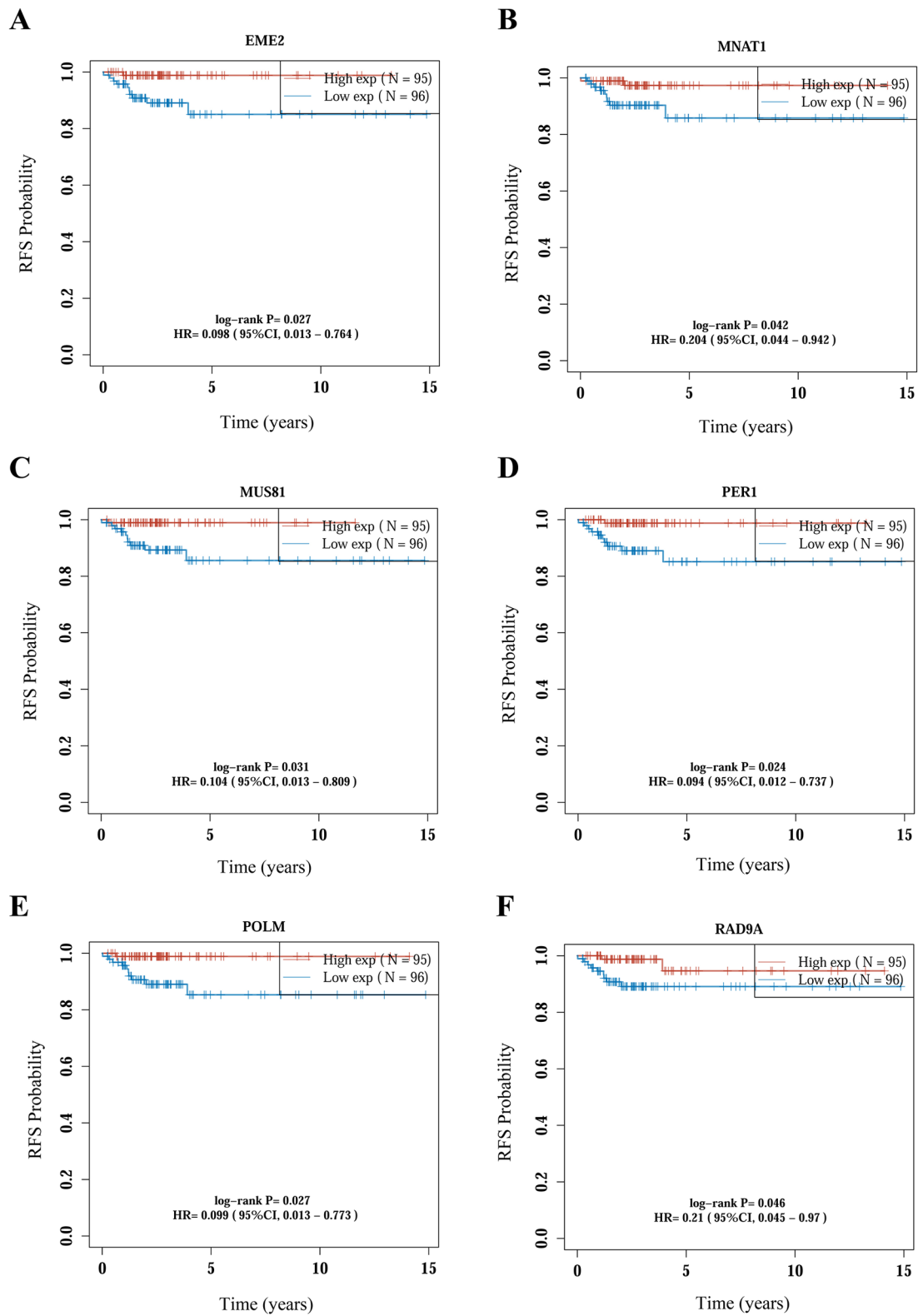


Fig. 3 The RFS curves of 6 DDR genes with prognostic significance in early-stage PTC. **A** The RFS curve according to the expression of EME2. **B** The RFS curve according to the expression of MNAT1. **C** The RFS curve according to the expression of MUS81. **D** The RFS curve according to the expression of PER1. **E** The RFS curve according to the expression of POLM. **F** The RFS curve according to the expression of RAD9A

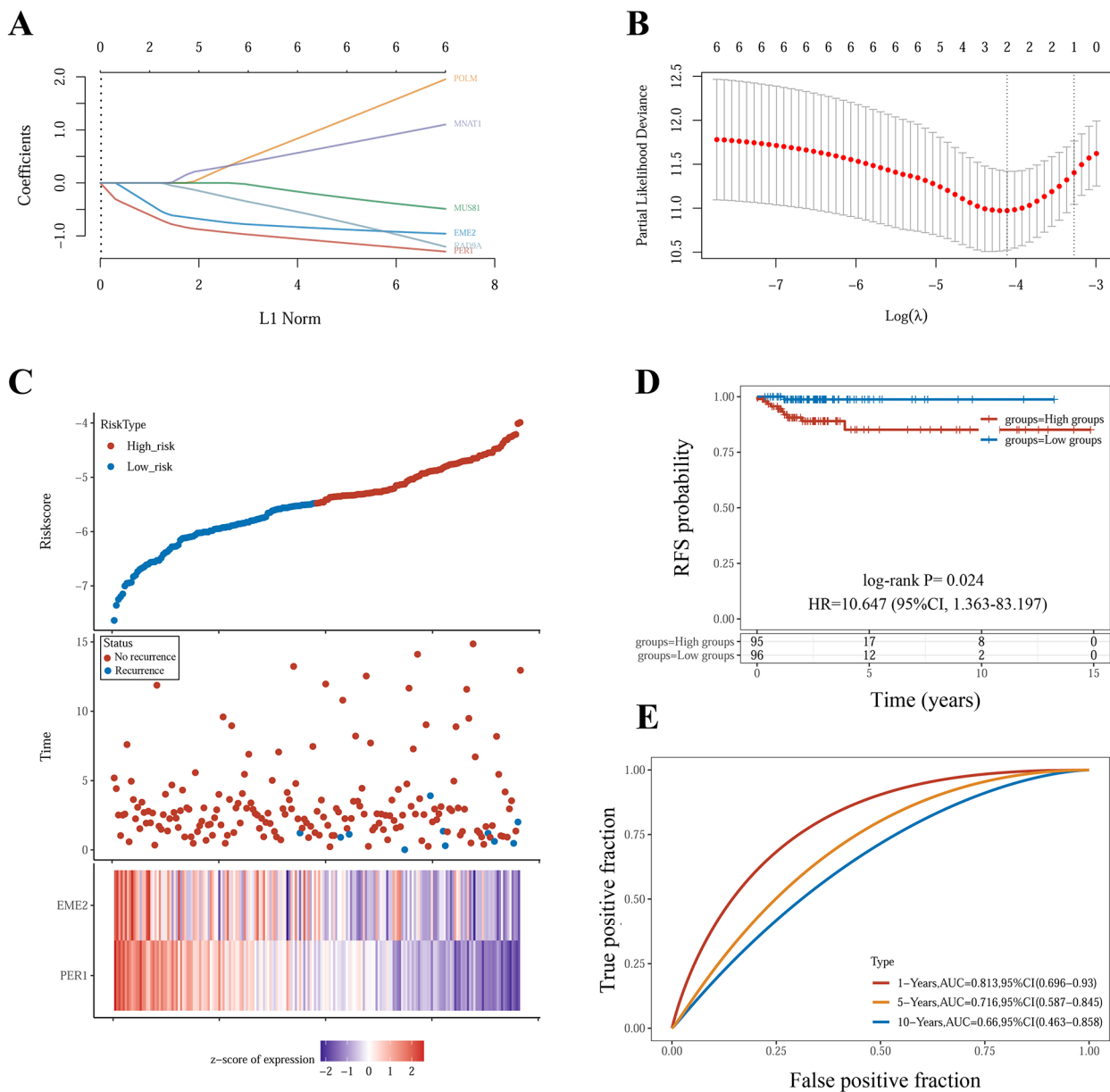


Fig. 4 Construction of DDR recurrence signature. **A** Lasso regression analysis. **B** Partial likelihood deviance for different genes and numbers. **C** Patient distribution according to risk score from small to large. **D** The RFS curve according to different risk groups. **E** The AUC of DDR recurrence signature for the prediction of 1-year, 5-year, and 10-year RFS.

of each sample was obtained using the following formula: Risk score = $(-0.664) \cdot PER1 + (-0.483) \cdot EME2$. Multivariate Cox regression analysis showed that the risk score could be taken as an independent prognostic factor for RFS in early-stage PTC patients (Table 2). The patient distribution was shown in Fig. 4C according to risk score from small to large. The PTC patients with recurrence were largely distributed in the region with higher risk score. The mRNA expression levels of *PER1* and *EME2* decreased gradually with the elevation of recurrence risk score. According to the median risk score, the

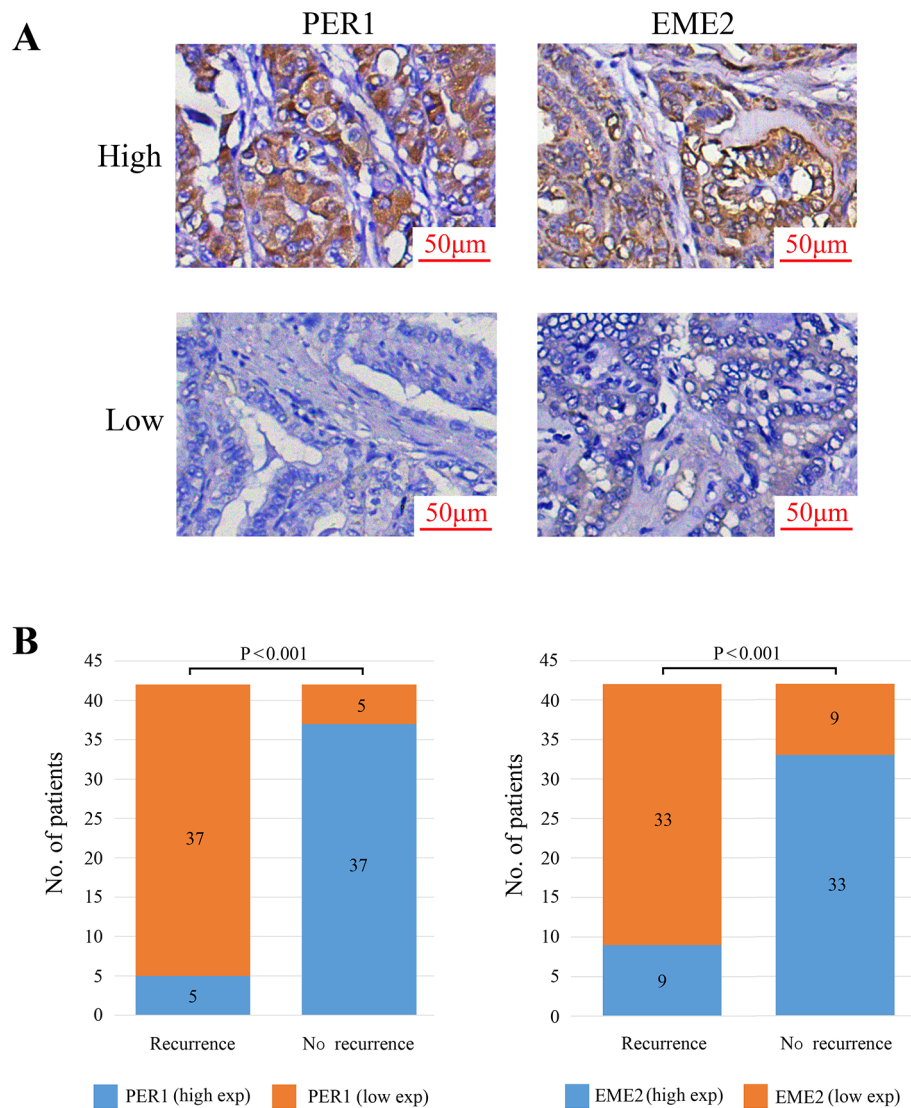
entire cohort was divided into high-risk group and low-risk group. The high-risk group showed a significantly worse RFS than the low-risk group [HR=10.647 (95%CI, 1.363–83.197), $p=0.024$] (Fig. 4D). The AUC of the DDR recurrence signature for the prediction of 1-year RFS was 0.813 (95%CI, 0.696–0.930), indicating that the signature could accurately reflect the recurrence of early-stage PTC (Fig. 4E). Due to lower recurrence rate, however, the signature did not exhibit as highly predictive power for 10-year RFS as 1-year RFS.

Table 2 Multivariate Cox analysis for RFS of PTC patients with stage I aged < 55 years

Characteristics	HR (95%CI)	P value
Age	1.034(0.963–1.110)	0.359
Gender		
Male	Reference	
Female	0.951(0.191–4.741)	0.951
Pathological T category		
T1	Reference	
T2	2.606(0.502–13.519)	0.254
T3	1.846(0.313–10.876)	0.498
Pathological N category		
N0	Reference	
N1	1.182(0.317–4.402)	0.803
Risk score	5.286(1.780–15.699)	0.003

Multicenter validation

42 early-stage PTC patients with recurrence were collected in the multicenter datasets between 2014 and 2016. Over the same period, a total of 629 early-stage PTC patients without recurrence were recorded. To eliminate intergroup selection bias, 1:1 propensity score matching (PSM) was performed by adjusting baseline characteristics, including age, gender, pathological T category, and pathological N category. 42 pairs of patients (42 cases in the recurrence group and 42 cases in the non-recurrence group) were finally included in the further validation using immunohistochemical analysis (Fig. 5A, Supplementary Fig. 1, Supplementary Table 4). In the recurrence group, 88.1% of patients had low expression of *PER1*, which was significantly higher than

**Fig. 5** Immunohistochemical analysis of clinical samples from multi-center datasets. **A** Immunohistochemical staining for PER1 and EME2. **B** Statistical analysis for PER1 and EME2 protein staining

that in the non-recurrence group (Fig. 5B, $p < 0.001$). Similarly, the proportion of patients with low expression of *EME2* in the recurrence group was also significantly higher than that in the non-recurrence group (78.6% vs. 21.4%, Fig. 5B, $p < 0.001$).

Comprehensive assessment of DDR recurrence signature

We compared the discriminatory abilities of the DDR recurrence signature and ATA recurrence risk stratification. The C-index for the DDR recurrence signature was 0.701 (95% CI=0.561–0.841), whereas for the ATA recurrence risk stratification it was 0.509 (95% CI=0.378–0.640) (Fig. 6A). These indicated the DDR recurrence signature outperformed the ATA recurrence risk stratification in the prediction of recurrence risk for early-stage PTC.

Next, the performance of the DDR recurrence signature and published *PBF/PTTG1* signature were compared [11]. Although the *PBF/PTTG1* combined signature displayed a notable discriminatory power for all-stage PTC, its performance in distinguishing early-stage PTC

remained to be evaluated. Regrettably, the *PBF/PTTG1* signature was not a reliable predictor for recurrence risk in early-stage PTC (Fig. 6B).

Finally, we further investigated the effectiveness of the DDR recurrence signature in predicting recurrence for all-stage PTC using the entire TCGA cohort. Unfortunately, the DDR recurrence signature did not exhibit strong predictive capability for recurrence in all-stage PTC (Fig. 6C). Even for BRAF-mutated PTC with a higher recurrence rate, the predictive performance of the DDR recurrence signature was not sufficiently satisfactory either (Fig. 6D), suggesting that the DDR recurrence signature may function as a recurrence risk model only for early-stage PTC.

Discussion

PTC is generally an indolent disease with extremely good survival outcomes and concern arises as to whether some PTC patients are over-treated, particularly PTC patients with early stage. It was for the reason that the ATA guidelines recommended lobectomy as an

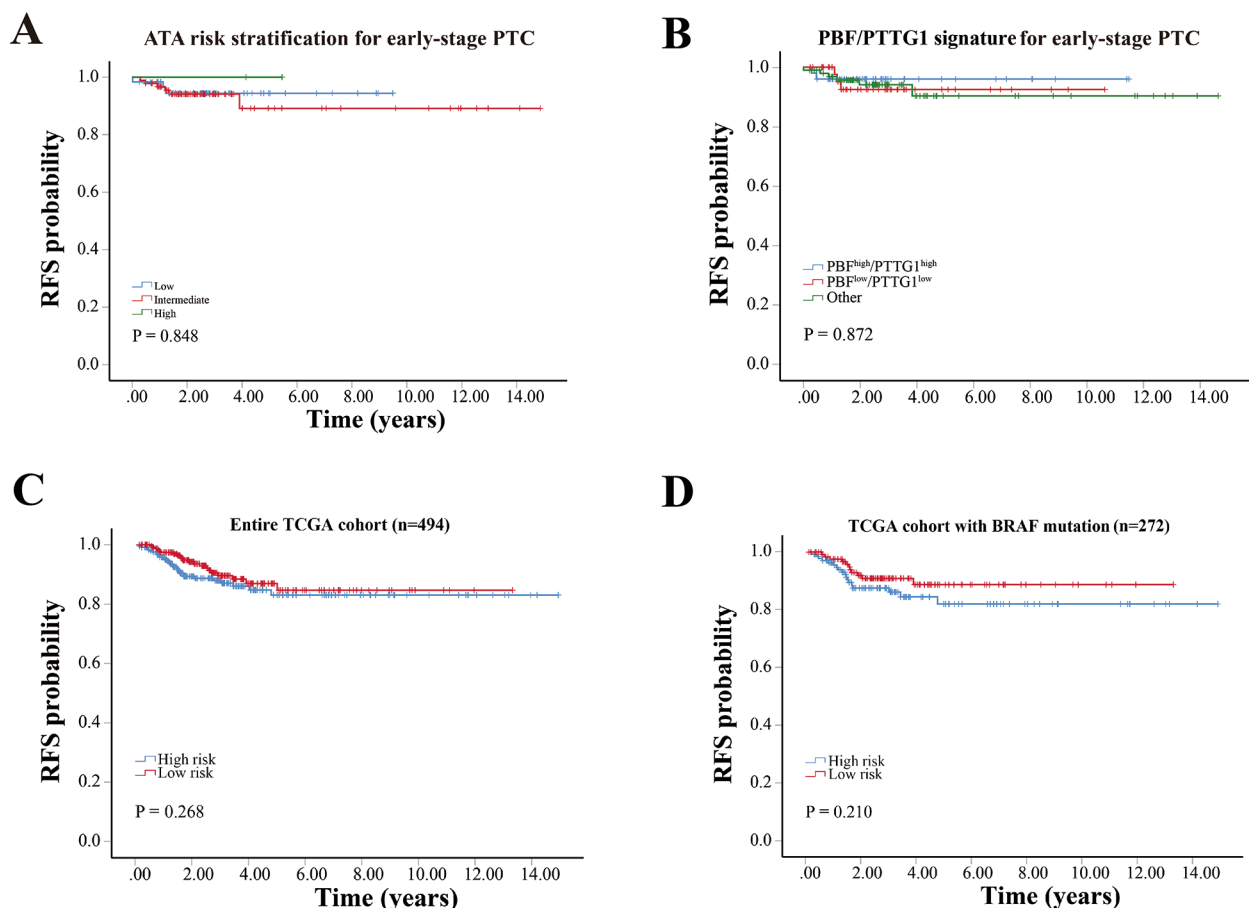


Fig. 6 Comprehensive assessment of DDR Recurrence Signature. **A** The RFS curve for early-stage PTC according to the ATA risk stratification. **B** The RFS curve for early-stage PTC according to the *PBF/PTTG1* signature. **C** The application of the DDR signature in the entire TCGA cohort. **D** The application of the DDR signature in the TCGA cohort with BRAF mutation

appropriate treatment method for solitary intrathyroidal PTC without lymph node or distant metastases (low risk of recurrence patients according to the ATA guidelines). However, there is currently no strong supporting clinical evidence for the recommendation. Huang et al. [22] classified the low risk of recurrence patients based on the BRAF V600E status, with a view to finding a subgroup of patients with high risk of recurrence. The study also has several shortcomings that should be addressed. First, the prevalence of BRAF V600E mutation in early-stage PTC was significantly lower than the average of all PTC patients [22–24]. Second, multi-gene signature might provide more accurate stratification of risk. Using the TCGA database and multi-center datasets, we proposed a novel molecular subtype of PTC patients with stage I aged <55 years based on DDR, which were associated with disparate risks of recurrence. Further, a two-gene DDR recurrence signature (*PER1* and *EME2*) was constructed, which could accurately and reliably predict the risk of recurrence. These findings indicated that more radical and aggressive treatment strategies and more rigorous surveillance practices might be appropriate for early-stage PTC patients with low expression of *PER1* or *EME2*.

PER1 (period 1), one of circadian clock genes, was expressed at low levels in multiple cancers and reported to be involved in tumor progression and prognosis via multiple biological pathways, in particular via DDR pathways and cell apoptosis [25, 26]. Jiang et al. [27] developed a ten-gene signature including *PER1*, which could effectively predict the disease-free survival (DFS) of patients with hepatocellular cancer. High expression of *PER1* was also found to be associated with metastasis-free survival (MFS) of patients with breast cancer [28], suggesting an important role of *PER1* in the early stages of tumor progression. However, there are no reports about the effect of *PER1* on the prognosis of PTC patients at present. Our study found that the low expression of *PER1* was significantly associated with high risk of recurrence in early-stage PTC, which was in good agreement with previous studies for other cancers [29]. *PER1* promoted DDR-induced apoptosis via *p53/Myc* pathway in different tumor cells, while *p53* repressed the transcription of *PER1* [30, 31]. These indicated that *PER1* might decelerate the relapse process of early-stage PTC by interacting with *p53*.

EME2, essential meiotic structure-specific endonuclease subunit 2, is localized on chromosome 16p13.3. *EME2*, *EME1*, and *MUS81* belong to the *XPF/MUS81* family of structure-selective endonucleases [32]. It has previously been suggested that *MUS81* participated in various cellular functions only through *MUS81-EME1* complex [33]. Recent studies indicated that *MUS81-EME2* complex was involved in the process of fork

cleavage and restart during DNA replication [34]. Our study also showed that both *EME2* and *MUS81* were significant predictors of early-stage PTC recurrence, suggesting *EME2* and *MUS81* might act in the form of complex in PTC. Kiminori et al. [35] showed that depletion of *MUS81-EME2* complex and *Polη* (a Y-family translesion synthesis polymerase) would synergistically enhanced *Myc*-induced replication stress, contributing to carcinogenesis. Additionally, simultaneous deficiency of *MUS81* and *p53* greatly increased the likelihood of sarcoma development in mice. These suggested that *MUS81-EME2* complex, *PER1*, and *p53/Myc* pathway may collectively contribute to the molecular mechanisms involved in the recurrence of early-stage PTC. It has been shown that the well-known proto-oncogenes *PBF* and *PTTG1* not only modulated the expression of DDR genes, but also exerted effects on the activity of *p53* [11]. Therefore, we speculate that *PBF/PTTG1* may be upstream regulatory genes for complex molecular networks including *MUS81-EME2*, *PER1*, and *p53/Myc* in early-stage PTC. Certainly, further experiments are needed to confirm this deduction.

Although the other genes related to RFS were not included in the DDR signature, they were reported to have an impact on the tumor progression in malignancies. The single nucleotide polymorphism (SNP) of *MNAT1* (rs12888332) was a significant predictor of recurrence of head and neck cancer [36]. Long et al. [37] established a RFS nomogram model of prostate cancer including *POLM*, which could accurately discriminated patients at high and low risk of recurrence. Eva et al. [38] found that reduced expression of *RAD9A* was associated with high risk of developing secondary cancer in children with a history of cancer.

There were a few limitations in the present study. First, all the data from TCGA and multi-center were retrospective. Prospective, multi-center and large-sample clinical trials are needed to provide guidance for individualized precision therapy. Second, as PTC patients with stage I aged <55 years usually have an excellent prognosis, the duration of follow-up seemed not sufficiently long. Third, the study did not address animal or in vitro cellular experiments, leading to insufficient mining of molecular mechanisms behind recurrence in early-stage PTC.

Conclusions

In sum, we proposed a novel molecular subtype of PTC patients with stage I aged <55 years based on DDR, and further constructed a two-gene signature (*PER1* and *EME2*) for early-stage PTC recurrence prediction. These findings could help accurately and reliably identify PTC patients with high risk of recurrence so that they could receive more radical and aggressive treatment strategies and more rigorous surveillance practices.

Abbreviations

PTC	Papillary thyroid cancer
FTC	Follicular thyroid cancer
MTC	Medullary thyroid cancer
ATC	Anaplastic thyroid cancer
AJCC	American Joint Committee on Cancer
ATA	American Thyroid Association
EMT	Mesenchymal transition
DDR	DNA damage repair
CDF	Cumulative distribution function
SD	Standard deviation
LASSO	Least absolute shrinkage and selection operator
ROC	Receiver operating characteristic
AUC	Area under the ROC curve
HR	Hazard ratio
CI	Confidence interval
PSM	Propensity score matching
DFS	Disease-free survival
MFS	Metastasis-free survival
SNP	Single nucleotide polymorphism
C-index	Concordance index

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13044-024-00193-9>.

Supplementary Fig.1. Representative images of IHC
Supplementary Table 1. Detailed list of 8 gene sets
Supplementary Table 2. List of antibodies with RRID numbers
Supplementary Table 3. All DDR genes with HRs and 95%CI
Supplementary Table 4. Detailed data of multicenter cohort

Author contributions

Chuangang Tang, Mingyu Sun, Xinbao Gao, and Bingqing Zhao did the data collections and the major part of data analyses/statistics. Shaojun Hu supervised the process. All authors contributed in writing and revising the manuscript. All authors read and approved the final manuscript.

Funding

None.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request. The training dataset was obtained from the TCGA data portal (<https://portal.gdc.cancer.gov>).

Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Committee of Suzhou Ninth People's Hospital. Samples from our own hospital were collected and disposed of in accordance with WHO guidelines. The participants provided their written informed consent to participate in this study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Breast Surgery, Xuzhou Central Hospital, The Affiliated Xuzhou Hospital of Medical College of Southeast University, Xuzhou 221009, China

²Department of Plastic and Reconstructive Surgery, Tianjin Hospital of ITCWM Nankai Hospital, Tianjin 300100, China

³The Xuzhou Clinical College of Xuzhou Medical University, Jiangsu 221009, China

⁴Department of Oncology, Suzhou Ninth People's Hospital, Suzhou 215200, China

⁵Department of Pathology, Xuzhou Central Hospital, The Affiliated Xuzhou Hospital of Medical College of Southeast University, Xuzhou 221009, China

⁶Department of Surgery for Vascular Thyroid and Hernia, Xuzhou Central Hospital, The Affiliated Xuzhou Hospital of Medical College of Southeast University, Xuzhou 221009, China

Received: 9 November 2023 / Accepted: 6 February 2024

Published online: 01 April 2024

References

1. Xing M. Molecular pathogenesis and mechanisms of thyroid cancer. *Nat Rev Cancer*. 2013;13(3):184–99.
2. Xing M, Haugen BR, Schlumberger M. Progress in molecular-based management of differentiated thyroid cancer. *Lancet*. 2013;381(9871):1058–69.
3. Huang K, Gao N, Bian D, Zhai Q, Yang P, Zhang Y. Associations of BRAF V600E, clinical pathology and imaging factors with the recurrence rate of papillary thyroid microcarcinoma. *Exp Ther Med*. 2020;20(6):243.
4. van Velsen EFS, Stegenga MT, van Kemenade FJ, Kam BLR, van Ginhoven TM, Visser WE, Peeters RP. Comparing the Prognostic Value of the Eighth Edition of the American Joint Committee on Cancer/Tumor Node Metastasis Staging System between Papillary and follicular thyroid Cancer. *Thyroid*. 2018;28(8):976–81.
5. Adam MA, Thomas S, Hyslop T, Scheri RP, Roman SA, Sosa JA. Exploring the relationship between patient age and Cancer-specific survival in papillary thyroid Cancer: Rethinking Current Staging systems. *J Clin Oncol*. 2016;34(36):4415–20.
6. Thewjitchareon Y, Chatchomchuan W, Karndumri K, Porramatikul S, Krittiyawong S, Wanothayaroj E, Butadej S, Nakasatien S, Veerasomboonsin V, Kanchanapituk A, et al. Impacts of the American Joint Committee on Cancer (AJCC) 8(th) edition tumor, node, metastasis (TNM) staging system on outcomes of differentiated thyroid cancer in Thai patients. *Heliyon*. 2021;7(3):e06624.
7. Davies L, Welch HG. Increasing incidence of thyroid cancer in the United States, 1973–2002. *JAMA*. 2006;295(18):2164–7.
8. Grozinsky-Glasberg S, Benbassat CA, Tsvetov G, Feinmesser R, Peretz H, Shimon I, Lapidot M. Medullary thyroid cancer: a retrospective analysis of a cohort treated at a single tertiary care center between 1970 and 2005. *Thyroid* 2007, 17(6):549–556.
9. Jung KW, Won YJ, Oh CM, Kong HJ, Lee DH, Lee KH. Community of Population-based Regional Cancer R: Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2014. *Cancer Res Treat*. 2017;49(2):292–305.
10. Landa I, Ibrahimspasic T, Boucai L, Sinha R, Knauf JA, Shah RH, Dogan S, Ricarte-Filho JC, Krishnamoorthy GP, Xu B, et al. Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers. *J Clin Invest*. 2016;126(3):1052–66.
11. Read ML, Fong JC, Modasia B, Fletcher A, Imruetaichareonchoke W, Thompson RJ, Nieto H, Reynolds JJ, Bacon A, Mallick U, et al. Elevated PTTG and PBF predicts poor patient outcome and modulates DNA damage response genes in thyroid cancer. *Oncogene*. 2017;36(37):5296–308.
12. Read ML, Seed RI, Fong JC, Modasia B, Ryan GA, Watkins RJ, Gagliano T, Smith VE, Stratford AL, Kwan PK, et al. The PTTG1-binding factor (PBF/PTTG1IP) regulates p53 activity in thyroid cells. *Endocrinology*. 2014;155(4):1222–34.
13. Nieto HR, Thornton CEM, Brookes K, Nobre de Menezes A, Fletcher A, Alshahrani M, Kocbiyik M, Sharma N, Boelaert K, Cazier JB, et al. Recurrence of papillary thyroid Cancer: a systematic Appraisal of Risk factors. *J Clin Endocrinol Metab*. 2022;107(5):1392–406.
14. Read ML, Brookes K, Thornton CEM, Fletcher A, Nieto HR, Alshahrani M, Khan R, Borges de Souza P, Zha L, Webster JRM, et al. Targeting non-canonical pathways as a strategy to modulate the sodium iodide symporter. *Cell Chem Biol*. 2022;29(3):502–516e507.
15. Read ML, Brookes K, Zha L, Manivannan S, Kim J, Kocbiyik M, Fletcher A, Gorvin CM, Firth G, Fruhwirth GO et al. Combined Vorinostat and Chloroquine inhibit Sodium Iodide Symporter endocytosis and enhance Radionuclide Uptake in vivo. *Clin Cancer Res* 2023.

16. Yi L, Wu G, Guo L, Zou X, Huang P. Comprehensive Analysis of the PD-L1 and Immune infiltrates of m(6)a RNA methylation regulators in Head and Neck squamous cell carcinoma. *Mol Ther Nucleic Acids*. 2020;21:299–314.
17. Liu Z, Zhao Q, Zuo ZX, Yuan SQ, Yu K, Zhang Q, Zhang X, Sheng H, Ju HQ, Cheng H et al. Systematic Analysis of the Aberrances and Functional Implications of Ferroptosis in Cancer. *iScience* 2020, 23(7):101302.
18. Li Y, Xiao J, Bai J, Tian Y, Qu Y, Chen X, Wang Q, Li X, Zhang Y, Xu J. Molecular characterization and clinical relevance of m(6)a regulators across 33 cancer types. *Mol Cancer*. 2019;18(1):137.
19. Zhang Z, Lin E, Zhuang H, Xie L, Feng X, Liu J, Yu Y. Construction of a novel gene-based model for prognosis prediction of clear cell renal cell carcinoma. *Cancer Cell Int*. 2020;20:27.
20. Li B, Cui Y, Nambiar DK, Sunwoo JB, Li R. The Immune subtypes and Landscape of squamous cell carcinoma. *Clin Cancer Res*. 2019;25(12):3528–37.
21. Karasinska JM, Topham JT, Kalloger SE, Jang GH, Denroche RE, Culibrk L, Williamson LM, Wong HL, Lee MKC, O’Kane GM, et al. Altered gene expression along the glycolysis-cholesterol synthesis Axis is Associated with Outcome in Pancreatic Cancer. *Clin Cancer Res*. 2020;26(1):135–46.
22. Huang Y, Qu S, Zhu G, Wang F, Liu R, Shen X, Viola D, Elisei R, Puxeddu E, Fugazzola L, et al. BRAF V600E mutation-assisted risk stratification of Solitary Intrathyroidal Papillary thyroid Cancer for Precision Treatment. *J Natl Cancer Inst*. 2018;110(4):362–70.
23. Xing M. BRAF mutation in thyroid cancer. *Endocr Relat Cancer*. 2005;12(2):245–62.
24. Xing M. BRAF mutation in papillary thyroid cancer: pathogenic role, molecular bases, and clinical implications. *Endocr Rev*. 2007;28(7):742–62.
25. Sahar S, Sassone-Corsi P. Metabolism and cancer: the circadian clock connection. *Nat Rev Cancer*. 2009;9(12):886–96.
26. Sancar A, Lindsey-Boltz LA, Gaddameedhi S, Selby CP, Ye R, Chiou YY, Kemp MG, Hu J, Lee JH, Ozturk N. Circadian clock, cancer, and chemotherapy. *Biochemistry*. 2015;54(2):110–23.
27. Jiang HY, Ning G, Wang YS, Lv WB. Ahypoxia-related signature enhances the prediction of the prognosis in hepatocellular carcinoma patients and correlates with sorafenib treatment response. *Am J Transl Res*. 2020;12(12):7762–81.
28. Cadenas C, van de Sandt L, Edlund K, Lohr M, Hellwig B, Marchan R, Schmidt M, Rahnenfuhrer J, Oster H, Hengstler JG. Loss of circadian clock gene expression is associated with tumor progression in breast cancer. *Cell Cycle*. 2014;13(20):3282–91.
29. Kelleher FC, Rao A, Maguire A. Circadian molecular clocks and cancer. *Cancer Lett*. 2014;342(1):9–18.
30. Gery S, Komatsu N, Baldjyan L, Yu A, Koo D, Koeffler HP. The circadian gene *per1* plays an important role in cell growth and DNA damage control in human cancer cells. *Mol Cell*. 2006;22(3):375–82.
31. Bellet MM, Stincardini C, Costantini C, Gargaro M, Pieroni S, Castelli M, Piobico D, Sassone-Corsi P, Della-Fazia MA, Romani L et al. The circadian protein *PER1* modulates the Cellular response to Anticancer treatments. *Int J Mol Sci* 2021, 22(6).
32. Amangyeld T, Shin YK, Lee M, Kwon B, Seo YS. Human *MUS81-EME2* can cleave a variety of DNA structures including intact Holliday junction and nicked duplex. *Nucleic Acids Res*. 2014;42(9):5846–62.
33. Calzetta NL, Gonzalez Besteiro MA, Gottifredi V. *Mus81-Eme1*-dependent aberrant processing of DNA replication intermediates in mitosis impairs genome integrity. *Sci Adv* 2020, 6(50).
34. Pepe A, West SC. *MUS81-EME2* promotes replication fork restart. *Cell Rep*. 2014;7(4):1048–55.
35. Kurashima K, Sekimoto T, Oda T, Kawabata T, Hanaoka F, Yamashita T. Poleta, a Y-family translesion synthesis polymerase, promotes cellular tolerance of myc-induced replication stress. *J Cell Sci* 2018, 131(12).
36. Wu X, Spitz MR, Lee JJ, Lippman SM, Ye Y, Yang H, Khuri FR, Kim E, Gu J, Lotan R, et al. Novel susceptibility loci for second primary tumors/recurrence in head and neck cancer patients: large-scale evaluation of genetic variants. *Cancer Prev Res (Phila)*. 2009;2(7):617–24.
37. Long G, Ouyang W, Zhang Y, Sun G, Gan J, Hu Z, Li H. Identification of a DNA repair gene signature and establishment of a Prognostic Nomogram Predicting biochemical-recurrence-free survival of prostate Cancer. *Front Mol Biosci*. 2021;8:608369.
38. Weis E, Schoen H, Victor A, Spix C, Ludwig M, Schneider-Raetzke B, Kohlschmidt N, Bartsch O, Gerhold-Ay A, Boehm N, et al. Reduced mRNA and protein expression of the genomic caretaker *RAD9A* in primary fibroblasts of individuals with childhood and independent second cancer. *PLoS ONE*. 2011;6(10):e25750.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.