



m⁵C Regulator-Mediated Methylation Modification Patterns and Tumor Microenvironment Infiltration Characterization in Papillary Thyroid Carcinoma

Fei Li^{1*†}, Qingmei Deng^{2†}, Xiaoxi Pang^{1†}, Shan Huang¹, Jingmiao Zhang¹, Xiaxia Zhu¹, Hong Chen¹ and Xiuxia Liu¹

¹ Department of Nuclear Medicine, The Second Affiliated Hospital of Anhui Medical University, Anhui Medical University, Hefei, China, ² Department of Molecular Pathology, Hefei Cancer Hospital, Chinese Academy of Sciences, Hefei, China

OPEN ACCESS

Edited by:

Olorunseun O. Ogunwobi,
Hunter College (CUNY), United States

Reviewed by:

Shuyuan Wang,
Harbin Medical University, China
Wenhao Weng,
Tongji University, China

*Correspondence:

Fei Li
lifei007@139.com

[†]These authors have contributed
equally to this work and share
first authorship

Specialty section:

This article was submitted to
Cancer Genetics,
a section of the journal
Frontiers in Oncology

Received: 24 June 2021

Accepted: 12 October 2021

Published: 03 November 2021

Citation:

Li F, Deng Q, Pang X, Huang S,
Zhang J, Zhu X, Chen H and Liu X
(2021) m⁵C Regulator-Mediated
Methylation Modification Patterns
and Tumor Microenvironment
Infiltration Characterization in
Papillary Thyroid Carcinoma.
Front. Oncol. 11:729887.
doi: 10.3389/fonc.2021.729887

Recently, immune response modulation at the epigenetic level is illustrated in studies, but the possible function of RNA 5-methylcytosine (m⁵C) modification in cell infiltration within the tumor microenvironment (TME) is still unclear. Three different m⁵C modification patterns were identified, and high differentiation degree was observed in the cell infiltration features within TME under the above three identified patterns. A low m⁵C-score, which was reflected in the activated immunity, predicted the relatively favorable prognostic outcome. A small amount of effective immune infiltration was seen in the high m⁵C-score subtype, indicating the dismal patient survival. Our study constructed a diagnostic model using the 10 signature genes highly related to the m⁵C-score, discovered that the model exhibited high diagnostic accuracy for PTC, and screened out five potential drugs for PTC based on this m⁵C-score model. m⁵C modification exerts an important part in forming the TME complexity and diversity. It is valuable to evaluate the m⁵C modification patterns in single tumors, so as to enhance our understanding towards the infiltration characterization in TME.

Keywords: papillary thyroid carcinoma (PTC), subtype, immune infiltration, 5-methylcytosine (m⁵C) modification, tumor microenvironment (TME)

INTRODUCTION

Thyroid cancer (THCA), a frequently-occurring endocrine cancer, takes up approximately 1.7% of human cancers (1). TC can be divided into subtypes, namely, allopastic, follicular, medullary, and papillary thyroid cancer (PTC) (2). Of them, PTC shows the highest morbidity (75%–85% of thyroid cancers) (3). PTC can be cured under general conditions, and its survival rate at 5 years was over 95%, but PTC may sometimes differentiate to THCA, a malignancy with higher aggressiveness and mortality (4). Besides, approximately 30% of PTC cases suffer from tumor recurrence (5). As a result, analyzing the disease features at the molecular level is essential.

It is increasingly suggested that RNA modification at the post-transcriptional level exerts a vital part in a variety of malignancies (6, 7). RNA and histone alterations at epigenetic and genetic levels are extensively investigated in the context of tumor progression; as a result, numerous therapeutic means have been developed, such as the drugs that target the hypoxic pathways and the histone deacetylase inhibitors (8). In the living body, over 150 RNA modifications are modified as the third epigenetics layer, such as N1-methyladenosine (m¹A) and N6-methyladenosine (m⁶A), together with 5-methylcytosine (m⁵C) (9–13).

Of them, m⁵C modification, a reversible RNA post-transcriptional modification, exerts an important part in the regulation of mRNA translation, export, alternative splicing (AS) and stabilization localization (14, 15). m⁵C in mRNAs has been extensively studied, and many articles reveal that m⁵C greatly affects mRNAs, tRNAs, and rRNAs (16). The m⁵C methylation is related to various regulators, such as the m⁵C “readers”, demethylases, and methyltransferases. Typically, the methyltransferase “writer” complex enhances RNA methylation at the C5 position, whereas the distinct “reader” proteins are responsible for recognizing and binding to methylated mRNAs, and “eraser” protein is in charge of reversing the m⁵C modification through the degradation of written methylation. The adenosine demethylases, methyltransferases, together with the RNA-binding proteins involved in m⁵C modification are referred to as the m⁵C “erasers” (like TET2), m⁵C “writers” (like NSUN1-7, DNMT1-2, and DNMT3A-3B), as well as m⁵C “readers” (like ALYREF) (17). More and more studies suggest that m⁵C modification exerts an important part in a variety of critical pathophysiological processes, such as the dysregulated cell proliferation and death, abnormal immune modulation, developmental defects, malignant development of tumor, and damaged self-renewal ability (18–20). Nonetheless, the typical gene signatures, together with the diagnostic and prognostic significance of m⁵C-related regulators in PTC, are still unclear.

Immunotherapy based on the immunological checkpoint inhibitors (PD-1/L1, ICB, or CTLA-4) is found to be effective on certain patients who have persistent responses. However, most patients can only gain small or even no benefit from immunotherapy (21). In traditional practice, tumor progression is recognized to be the multi-step process involving variations within tumor cells at epigenetic and genetic levels, but many articles reveal that the tumor microenvironment (TME) for the development and survival of tumor cells also exerts an important part during tumor progression (22). There is a complicated TME in tumor, which contains tumor cells and stromal cells like macrophages and resident fibroblasts [cancer-associated fibroblast (CAF)]. In addition, it also contains distant recruited cells like the infiltrating immunocytes (lymphocytes and myeloid cells), bone marrow-derived cells (BMDCs) like hematopoietic and endothelial progenitor cells, the secretory factors (like chemokines, cytokines, and growth factors), and new vessels (23). Notably, the tumor-associated myeloid cells (TAMCs) are composed of five different myeloid subsets, namely, myeloid-derived suppressor cells (MDSCs), tumor-associated macrophages (TAMs), tumor-associated neutrophils (TANs), Tie2-expressing monocytes, and

dendritic cells (DCs) (24). Tumor cells can trigger changes in biological behaviors *via* directly or indirectly interacting with other components in the TME; for instance, the induction of new vessel formation and proliferation, apoptosis inhibition, hypoxia prevention, and immune tolerance induction (25). The TME complexity and diversity have been increasingly revealed, and TME is found to play an important part in immune escape and tumor progression, together with its impact on immunotherapy response (26, 27). It is critical to predict ICB response according to TME cell infiltration characterization, so as to increase the success rate of current ICBs and to exploit the new immunotherapies. Consequently, the comprehensive analysis of the complexity and diversity of TME landscapes helps to identify the diverse tumor immune phenotypes and to guide and predict responses to immunotherapies (28). Furthermore, it also contributes to revealing the potential biomarkers, thus facilitating to recognize the immunotherapy responses in patients and develop the novel therapeutic targets (29).

Individual recent articles suggest that the TME infiltrating immunocytes are related to m⁵C modification, and such relationship cannot be interpreted through the mechanism of RNA degradation (30, 31). Nonetheless, these articles only focus on holistic 5-hydroxymethylcytosine (5hmC) levels or cell types because of the technical restrictions, and the anticancer efficacy is evaluated based on a number of the highly coordinated tumor suppressor factors. Consequently, it is necessary to comprehensively recognize cell infiltration features within TME under the regulation of several m⁵C regulators, so as to shed more light on the TME immunomodulation. The present work combined genome data from 493 TCGA-PTC samples for the comprehensive evaluation of m⁵C modification patterns, and related them to cell infiltration features within TME. Altogether, three different m⁵C modification patterns were identified, under which the high differentiation degree of TME features were found, indicating the critical part of m⁵C modification in forming individual TME features. On this basis, the scoring system was also established for the quantification of m⁵C modification patterns for individual cases. Finally, this study mined the m⁵C-score-related signature genes to construct the PTC diagnostic model using the support vector machine (SVM) method.

MATERIALS AND METHODS

Source and Preprocessing of PTC Data

The work flow chart in the present work is presented in **Supplementary Figure 1**. The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO) databases were searched to obtain the clinical annotation and related gene expression data. Patients who had no survival data were eliminated from this study. The eligible PTC cohorts [including GSE33630 (32), GSE65144 (33), and GSE29265, together with TCGA-PTC (The Cancer Genome Atlas-papillary thyroid carcinoma)] were collected into the present work. With regard to Affymetrix[®] microarray data, raw “CEL”

files were downloaded to adjust the background and normalize the quantile using the multiarray averaging approach by affy and simpleaffy packages. In terms of microarray data of additional sources, matrix files after normalization were collected directly. For TCGA datasets, the RNA sequencing information (FPKM values) of gene levels was obtained based on the Genomic Data Commons (GDC, <https://portal.gdc.cancer.gov/>) by TCGAAbiolinks of R package, a software designed to comprehensively analyze GDC data (34). Thereafter, the FPKM values were converted to the transcripts per kilobase million (TPM) values. At the same time, the GSE65144 (12 tumor and 13 normal samples), GSE33630 (60 tumor along with 45 normal samples), and GSE29265 (29 tumor and 20 normal samples) datasets were also downloaded. R package (version 3.6.1) was utilized for data analysis.

Consensus Clustering of the 13 m⁵C Regulators

Altogether, 13 regulators were obtained based on TCGA datasets to identify the diverse m⁵C regulator-mediated m⁵C modification patterns. All the 13 genes, except for ALYREF and NSUN1, were with available expression profiles. The remaining 11 m⁵C regulators contained one eraser (TET2) and 10 writers (NSUN2-7, DNMT1-2, and DNMT3A-3B). Of our 493 patients from TCGA-PTC, 9 among those 11 genes were differentially expressed between tumor and normal tissues (with the exception for NSUN3 and DNMT3A) (**Supplementary Table 1** and **Supplementary Figure 2**). Later, consensus clustering was adopted for identifying the different m⁵C modification patterns according to nine m⁵C regulator expression levels, and then patients were classified accordingly. The above procedure was performed using the ConsensusClusterPlus package (35) for 1000 iterations to guarantee the classification stability.

Gene Set Variation Analysis together With Functional Annotation

To investigate the heterogeneities in biological process among the m⁵C modification patterns, GSVA was carried out by the “GSVA” R package. Notably, GSVA is the unsupervised, non-parametric approach usually used to estimate variations of pathways and biological processes within the expression dataset samples (36). The “c2.cp.kegg.v7.0.symbols” gene sets were extracted based on MSigDB database to conduct GSVA. The adjusted $p < 0.05$ indicated statistical significance. Meanwhile, functional annotation was performed using WebGestaltR package (37), and the threshold was $FDR < 0.05$.

TME Cell Infiltration Estimation

Estimate R package was utilized to calculate immune and stromal scores for all samples to reflect the immune and stromal cell infiltration degrees on the whole. Besides, CIBERSORT algorithm (38) was adopted for quantifying cell infiltration relative abundance within the TME of PTC. Thereafter, the gene set used to mark the TME-infiltrating immunocyte type was acquired to score different human immunocyte subtypes,

like the activated CD8 T cells, regulatory T cells, natural killer T cells, activated dendritic cells (DCs), and macrophages (39).

Discovery of Differentially Expressed Genes Across the Different m⁵C Phenotypes

To identify the m⁵C-associated genes, the patients were divided into three different m⁵C modification patterns according to nine m⁵C regulator expression levels. DEGs were determined across the diverse modification patterns using the empirical Bayesian method in the limma R package (40). The adjusted $p < 0.05$ served as the significance criterion to determine DEGs.

m⁵C Gene Signature Construction

To quantify m⁵C modification patterns among individual tumors, the scoring system, m⁵C-score, was built based on the m⁵C gene signature, as shown below:

First of all, DEGs obtained based on the diverse m⁵C-clusters were subjected to normalization across all samples, and then, those overlapped genes were selected. Afterwards, all cases were divided into different groups *via* the unsupervised clustering approach, so as to analyze the overlapped DEGs. In addition, the gene cluster number and the stability were defined using the consensus clustering algorithm. Later, prognostic analysis was carried out for all genes selected in our constructed signature by the use of the univariate Cox regression model. Later, those significant genes were obtained in subsequent analysis. In this study, $p < 0.01$ was selected as the criterion to screen 49 genes. **Supplementary Table 2** shows the results of single factor survival analysis for the 49 genes. Then, principal component analysis (PCA) was utilized for constructing the m⁵C signature. PC1 and PC2 were adopted as the signature scores; as a result, the score was focused on the set that had the greatest number of well-correlated (or anticorrelated) genes, and the contributions of genes not tracking with other members in the set were down-weighted. Later, the m⁵C-score was defined by the GGI-like approach (41):

$$m^5C\text{-score} = \sum(PC1_i + PC2_i)$$

In the formula, i represents the 49 m⁵C phenotype-associated gene expression levels.

m⁵C-Score Based PTC Diagnostic Model Establishment

First of all, this study mined the signature genes significantly correlated with the m⁵C-score (correlation coefficient > 0.4), and the PTC diagnostic model was constructed by the SVM method. Thereafter, the accuracy of this model was verified using samples from TCGA and GEO databases.

Statistical Methods

The Spearman and distance correlation analysis was adopted to calculate the correlation coefficients of TME-infiltrating immunocytes with m⁵C regulator expression levels. Thereafter, Kruskal-Wallis test and one-way ANOVA were applied in comparing the heterogeneities among three groups. Based on correlations of the m⁵C-score with patient survival, the survminer R package was utilized to determine the threshold value for every

dataset. In addition, the “surv-cutpoint” function was used for dichotomizing the m⁵C-score by testing the possible threshold values to find the maximal rank statistic. Later, all cases were classified as high or low m⁵C-score group according to the maximal log-rank statistics for reducing the calculation batch effect. In the meantime, the log-rank test and Kaplan–Meier approach were adopted for identifying the significance of differences, so as to generate survival curves. The univariate Cox regression model was used for calculating hazard ratios (HRs) for the m⁵C regulators as well as the m⁵C phenotype-associated genes. Meanwhile, the receiver operating characteristic (ROC) curve was plotted to assess the sensitivity and specificity of our diagnostic model and the m⁵C-score, and pROC R package was utilized to quantify the area under the curve (AUC). The two-sided statistical $p < 0.05$ indicated statistical significance. The R 3.6.1 software was employed for data processing.

RESULTS

The Nine-Regulator-Mediated m⁵C Methylation Modification Patterns

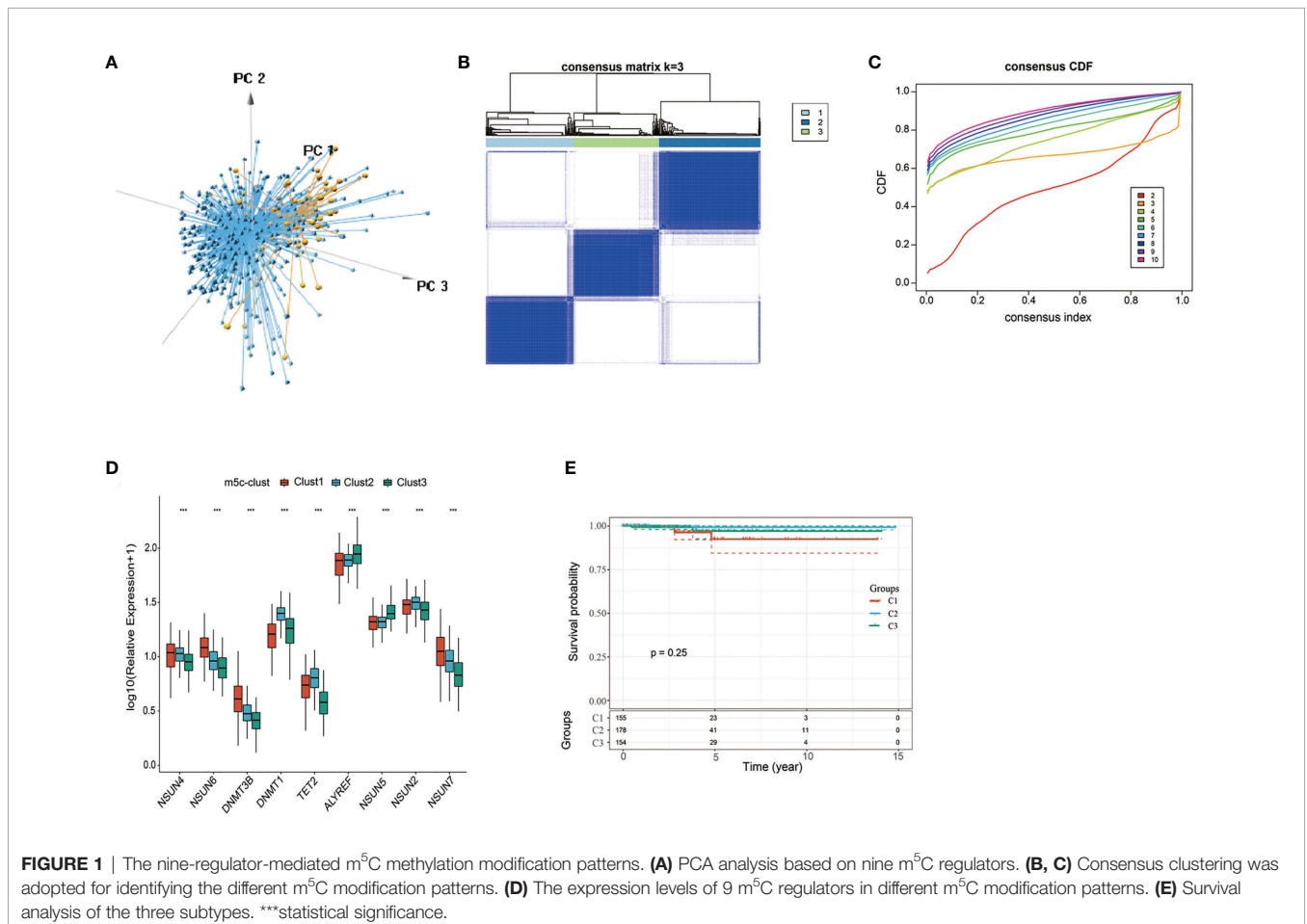
According to nine m⁵C regulators with expression profiles in the TCGA-PTC dataset, PTC samples were identified from normal

samples (**Figure 1A**). Afterwards, the expression profile data of these nine m⁵C regulators were carried out with z -score standardization using the scale function in mosaic package. Then, three different m⁵C modification patterns were discovered according to those nine m⁵C regulator expression patterns (**Figures 1B, C**). These three patterns were named m⁵C-clusters 1–3. It was observed from **Figure 1D** that the expression level of these nine m⁵C regulators showed significant differences among the three distinct subtype samples.

Prognostic analysis was also carried out for these three major m⁵C modification patterns, which suggested that the m⁵C-cluster 2 modification pattern showed survival advantage (**Figure 1E**). However, due to the speciality of PTC and the good overall prognosis, there was no significant statistical difference among these three subtypes. Besides, average survival time of samples in these three subtypes was also analyzed, which discovered that the average survival time of C2 subtype samples was 1307.657 days, that of C1 subtype samples was 1125.877 days, and that of C3 subtype samples was 1202.695, with that in C2 higher than those in C1 and C3.

TME-Infiltrating Cell Features in Different m⁵C Modification Patterns

To explore those biological behaviors in the different m⁵C modification patterns, GSVA was conducted. It was illustrated



from **Figure 2A** that m⁵C-cluster 1 significantly associated with the amino acid metabolic pathways; m⁵C-cluster 2 was enriched to the endocrine system, lipid metabolism, and cancer, whereas m⁵C-cluster 3 was associated with cell cycle, DNA repair, and nucleic acid metabolism.

Furthermore, the distribution of clinical features of samples in the above three subtypes was statistically analyzed. The statistical results are displayed in **Supplementary Table 3** and **Figure 3**. It was found from the results that, multiple clinical features in the three subtype samples were randomly distributed, with no significant difference.

In addition, the ESTIMATE algorithm was applied in quantifying the differences in stromal cell infiltration among the three subtype samples. As shown in **Figure 2B**, the stromal score in m⁵C-cluster 2 was the highest, followed by m⁵C-cluster 3, while m⁵C-cluster 1 had the lowest score. In addition, there were significant differences among them. Thereafter, the CIBERSORT deconvolution algorithm was utilized for comparing the heterogeneities in immunocyte components of the three m⁵C modification patterns (**Figure 2C**). Meanwhile, the support vector regression was used to determine the immunocyte types in tumors. As a result, high levels of Tregs and monocytes were detected in m⁵C-cluster 1 and m⁵C-cluster 3, whereas excessive resting/activated DCs were found in m⁵C-cluster 2. Recently, research has particularly focused on the RNA modification mechanism in the regulation of DC activation. DCs function to present antigen and to activate the naive T cells, which connect the intrinsic immunity with the adaptive one (42).

Finally, this study analyzed the expression of the 34 known immune checkpoints in the three subtype samples. As found from **Figure 2D**, there were significant differences in the expression of these 34 immune checkpoints among the three subtypes. Most immune checkpoint genes were highly expressed in m⁵C-cluster 2, followed by m⁵C-cluster 3, while m⁵C-cluster 1 had the lowest expression, which was consistent with the average survival time of samples in the three subtypes.

m⁵C Gene Signature Establishment Along With Functional Annotation

To better investigate the possible biological behaviors of all the m⁵C modification patterns, the limma package was used to determine 690 m⁵C phenotype-associated DEGs (**Supplementary Figure 3**). In addition, KEGG pathway enrichment analysis was carried out on DEGs using the WebGestaltR package. It was surprising that these genes were enriched to cell cycle, DNA repair, cell adhesion molecules, and immune inflammatory response-related pathways. These findings verified the important role of m⁵C modification in cancer cells themselves and in TME immunomodulation (**Figure 4A**). To better validate such regulatory mechanism, the unsupervised clustering analysis was performed using those 690 m⁵C phenotype-associated genes, for the sake of classifying cases to distinct genome subtypes. Similar to clustering analysis of m⁵C modification patterns, three different m⁵C modification genome phenotypes were found, which were referred to as m⁵C gene-clusters A–C, separately (**Supplementary Figure 4**). According to such results, there were three different m⁵C

methylation modification patterns in PTC. Besides, there were diverse signature genes in the three different gene clusters (**Supplementary Figure 4**). The m⁵C regulator expression levels were significantly different among the three m⁵C gene-clusters (**Supplementary Figure 5**), consistent with the results obtained for m⁵C methylation modification patterns. The expression quantities of these nine genes were the highest in gene-cluster B samples, followed by gene-cluster A samples, and were the lowest in the gene-cluster C samples.

Clinical Features and Transcriptome Traits of the m⁵C-Associated Phenotypes

First of all, we analyzed the stromal scores of three m⁵C gene-cluster subtypes. The results suggested that (**Figure 4B**) there were significant differences in the stromal score of three subtypes, among which, gene-cluster C had the highest score, followed by gene-cluster B, while gene-cluster A had the lowest score. Then, we analyzed the distribution of 22 immunocytes in the three m⁵C gene-cluster subtypes. As observed from **Figure 4C**, the distribution of 15 immunocytes in three subtypes showed statistically significant differences. These findings revealed the important role of m⁵C methylation modification in the formation of diverse TME landscapes and tumor-related immune regulation.

Nonetheless, the above results were obtained from patient population alone, which might not precisely estimate the m⁵C methylation modification patterns of individual cases. Due to the m⁵C modification complexity and heterogeneity in individual samples, this study established the scoring system (m⁵C-score) using the phenotype-associated genes for quantifying m⁵C modification patterns in individual PTC cases. Besides, those attribute alterations in individual patients were visualized by the alluvial diagram (**Figure 5A**). It was discovered from the figure that, among the 3 m⁵C-cluster subtypes, samples in m⁵C-cluster 2 and m⁵C-cluster 3 subtypes were mostly distributed in the low m⁵C-score group, while those in the high m⁵C-score group were basically derived from the m⁵C-cluster 1 subtype. In the three m⁵C gene-cluster subtypes, the m⁵C-score values of Cluster A and Cluster C samples were lower. Samples aged over 40 years were mostly classified into the low m⁵C-score group, while females mostly belonged to the high m⁵C-score group.

To further evaluate the differences between low and high score samples, the limma package was used to analyze the DEGs between the two groups. Using the thresholds of log₂FC > log₂(1.2) and *p* < 0.05, 67 DEGs were screened, including 58 upregulated and 9 downregulated ones (**Figure 5B**). Moreover, the WebGestaltR package was utilized for the GO and KEGG enrichment analyses of DEGs, with *p* < 0.05 as the threshold. A total of 62 biological processes (BP), 2 cellular components (CC), 6 molecular functions (MF), and 9 pathways were selected. As shown in **Supplementary Figure 6**, these genes were mainly involved in tumor proliferation and immune response-related biological processes/molecular functions and signaling pathways, such as MAPK, TNF, and IL-17.

Subsequently, this study observed the correlation of the m⁵C-score with patient survival and analyzed the difference in

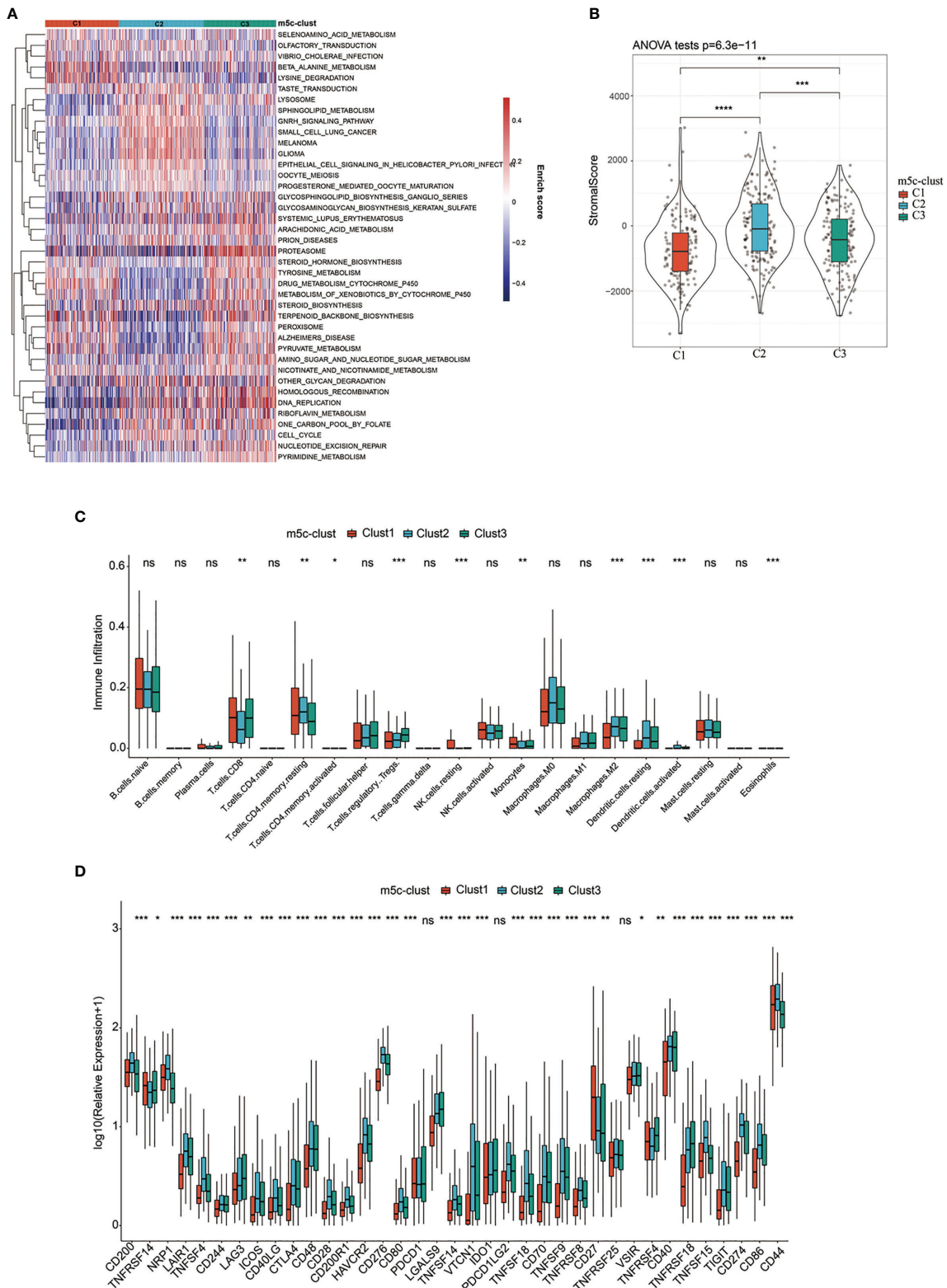


FIGURE 2 | TME-infiltrating cell features in different m⁵C modification patterns. **(A)** Biological behaviors in the different m⁵C modification patterns were conducted by GSVA. **(B)** Levels of stromal scores in different m⁵C modification patterns. **(C)** The levels of infiltration of 22 immune cells in different m⁵C modification patterns. **(D)** The expression levels of 37 immune checkpoints in different m⁵C modification patterns. *, **, ***, **** statistic difference at different levels; ns, no significance.

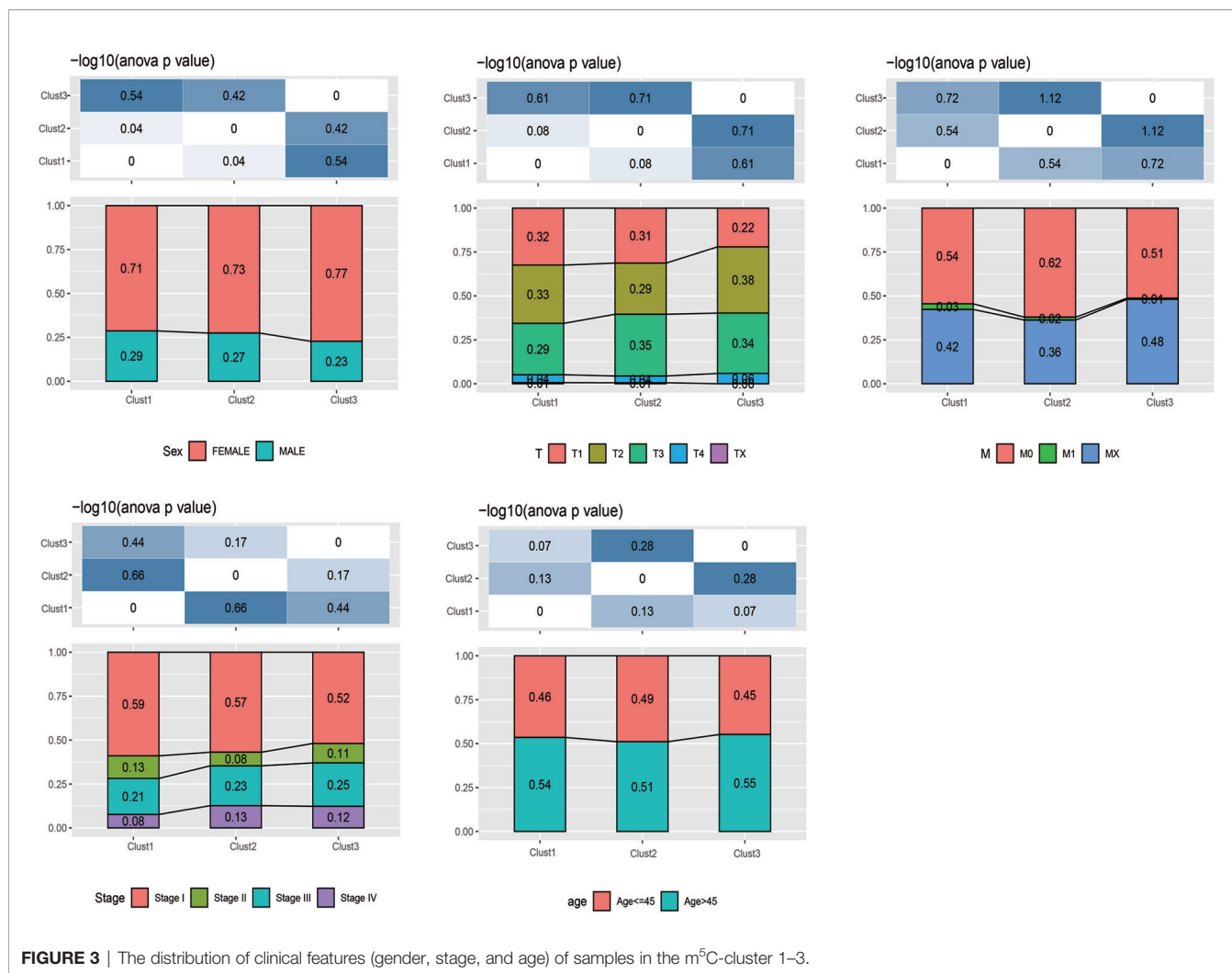


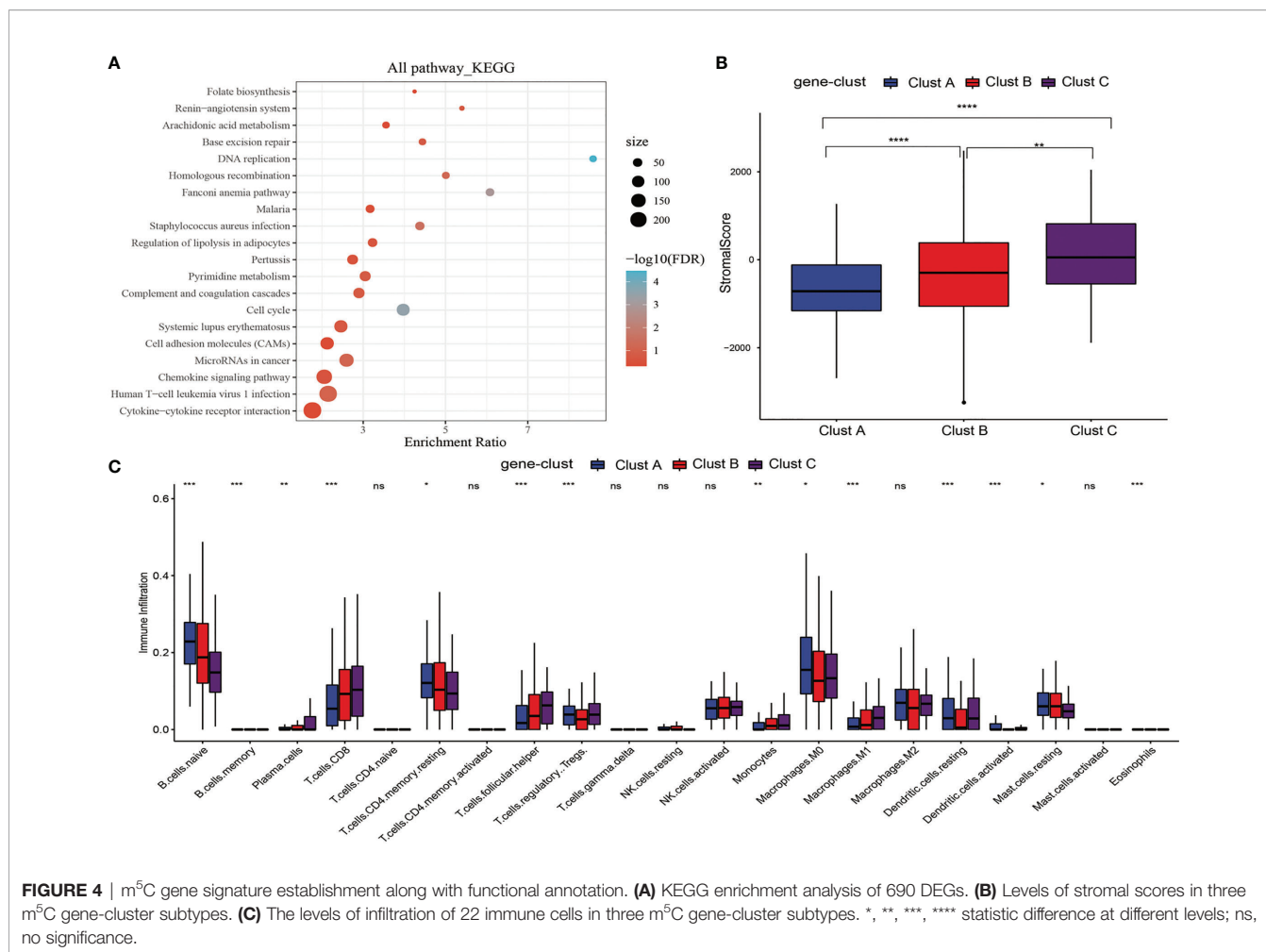
FIGURE 3 | The distribution of clinical features (gender, stage, and age) of samples in the m⁵C-cluster 1–3.

prognosis between high and low m⁵C-score samples. The results suggested that, samples with low m⁵C-scores had better prognosis than those with a high score, regardless of DFS or OS (**Figures 5C, D**). In addition, it was also discovered that there was no difference in the clinical features (such as T, M, and stage) between high and low m⁵C-score samples (**Supplementary Figure 7**). The expression levels of nine m⁵C regulators in the high m⁵C-score group were significantly higher than those in the low score group, and there was significant difference between the two groups (**Supplementary Figure 8**).

Subsequently, this study observed the correlation of the m⁵C-score with TME. First of all, the CIBERSORT method was adopted to evaluate the infiltration level of each immunocyte type in the high and low m⁵C-score TCGA-TPC samples. The results are presented in **Supplementary Figure 9A**. There were significant differences in six cell types between high and low m⁵C-score groups. In addition, this study also calculated the stromal score, immune score, and ESTIMATE score in different samples. As presented in **Supplementary Figure 9B**, in the low m⁵C-score group, the immune score was significantly

higher than that in the high m⁵C-score group, which was consistent with the previous results that show that the low m⁵C-score group had better prognosis than the high score group. Moreover, it was discovered through expression of immune checkpoint genes that there were significant differences in 16 immune checkpoint gene expression levels between high and low m⁵C-score groups (**Supplementary Figure S10**). Based on these findings, low m⁵C-score showed close correlation with immune activation. Furthermore, the m⁵C-score helped to assess m⁵C modification patterns in individual tumors and better assess the TME cell infiltration features of tumors, thus contributing to distinguishing the true or false TME immune infiltration.

Last, this study integrated the influences of the m⁵C-score and various immunocyte infiltration levels on the prognosis for PTC patients. From **Figure 6**, it was discovered that resting CD4⁺ memory T cells and CD8⁺ T cells were mainly enriched in low m⁵C-score samples, while activated NK cells and monocytes were mostly enriched in the high m⁵C-score group. Then, the median infiltration level of the above four cell types was used to



divide all samples into high and low immunocyte infiltration level groups. It was discovered that samples with low m⁵C-score and low infiltration level of resting CD4⁺ memory T cells had the best prognosis, while those with high m⁵C-score and low infiltration level of resting CD4⁺ memory T cells had the poorest prognosis. In addition, samples with low m⁵C-score and high CD8⁺ T cell infiltration had the best prognosis, while those with high m⁵C-score and low CD8⁺ T cell infiltration had the poorest prognosis. Furthermore, it was found that samples with low m⁵C-score and high monocyte infiltration had the best prognosis, while those with high m⁵C-score and high monocyte infiltration had the poorest prognosis. According to the prognostic prediction model, we analyzed the correlation between the m⁵C-score and Treg expression in 24 PTC patients. The m⁵C-score showed a negative relationship with CD3+CD4+/CD3+CD8+ ($r = -0.9543$, $p < 0.0001$; **Figure 5E**), but a positive relationship with CD4+CD25+ Tregs percentage ($r = 0.4477$, $p = 0.015$; **Figure 5F**).

Construction and Verification of the m⁵C-Score-Based PTC Diagnostic Model

First of all, this study calculated the correlation of 49 m⁵C phenotype-related genes with the m⁵C-score. Then, 10

signature genes related to the m⁵C-score were screened by the threshold of correlation coefficient >0.4, which were used as the features to construct the SVM classification model.

In order to verify the classification efficiency and accuracy of the model, we used the expression profile data of TCGA tumor samples as the training set. The m⁵C-score was utilized to classify the samples into high and low groups. Then, the expression profile data of these 10 genes were used to construct the SVM classification model to classify the TCGA-TPC samples. It was discovered that, compared with the m⁵C-score classification results, the accuracy reached 98.3%, and the sensitivity was up to 88.9%. The 493 samples were accurately classified, with an area under the ROC curve (AUC) of 0.936 (**Figure 7A**). The above results demonstrated that the classification model constructed based on these 10 signature genes well simulated the classification results of the m⁵C-score. The gene number was substantially reduced, which significantly improved the classification efficiency.

Thereafter, all the 551 TCGA samples (including 493 tumor samples and 58 normal samples) were used as the verification set 1. The abovementioned 10 genes were used as the features to construct the SVM classification model to classify the samples. Surprisingly, it was discovered that the model accurately classified

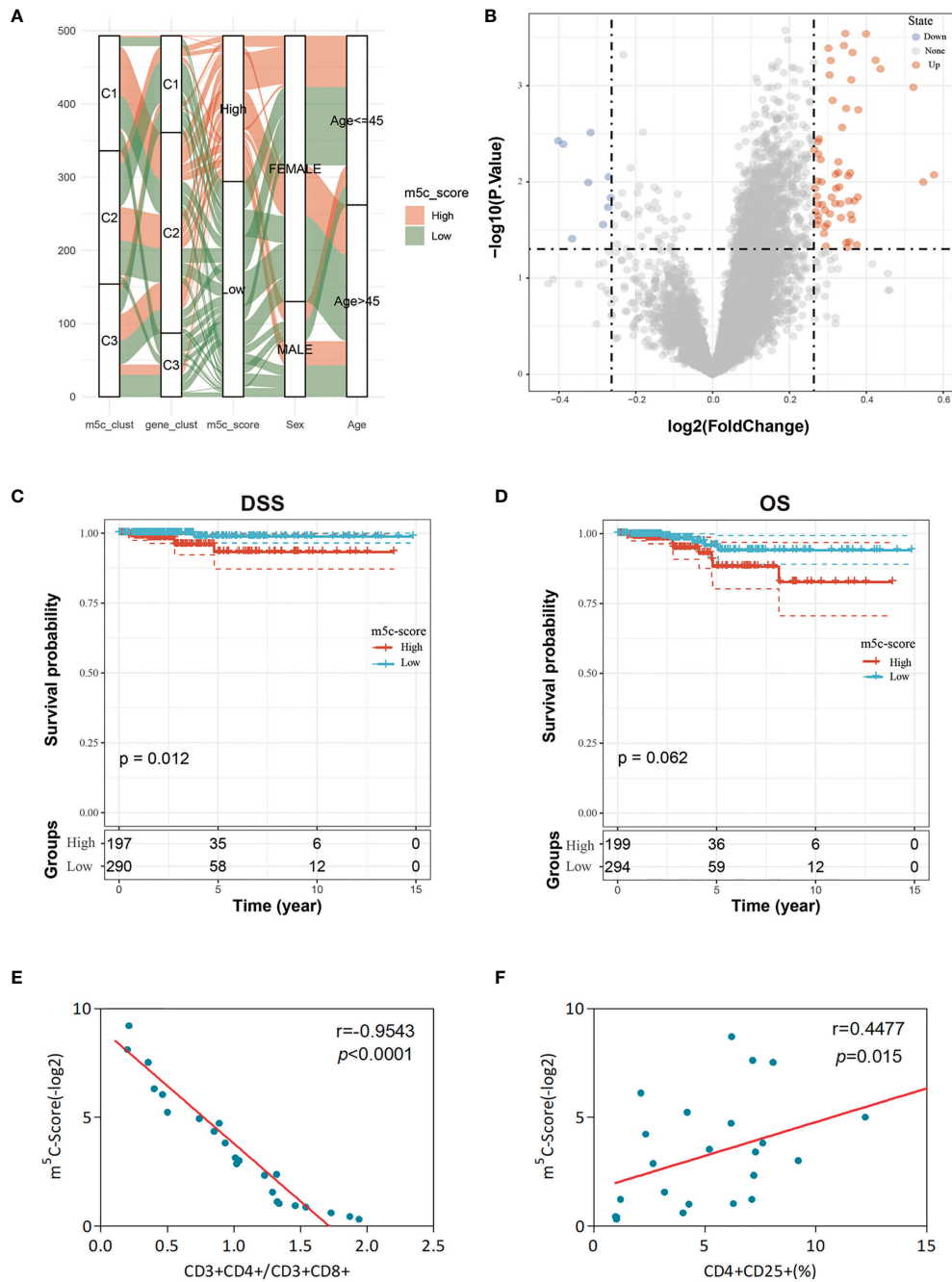
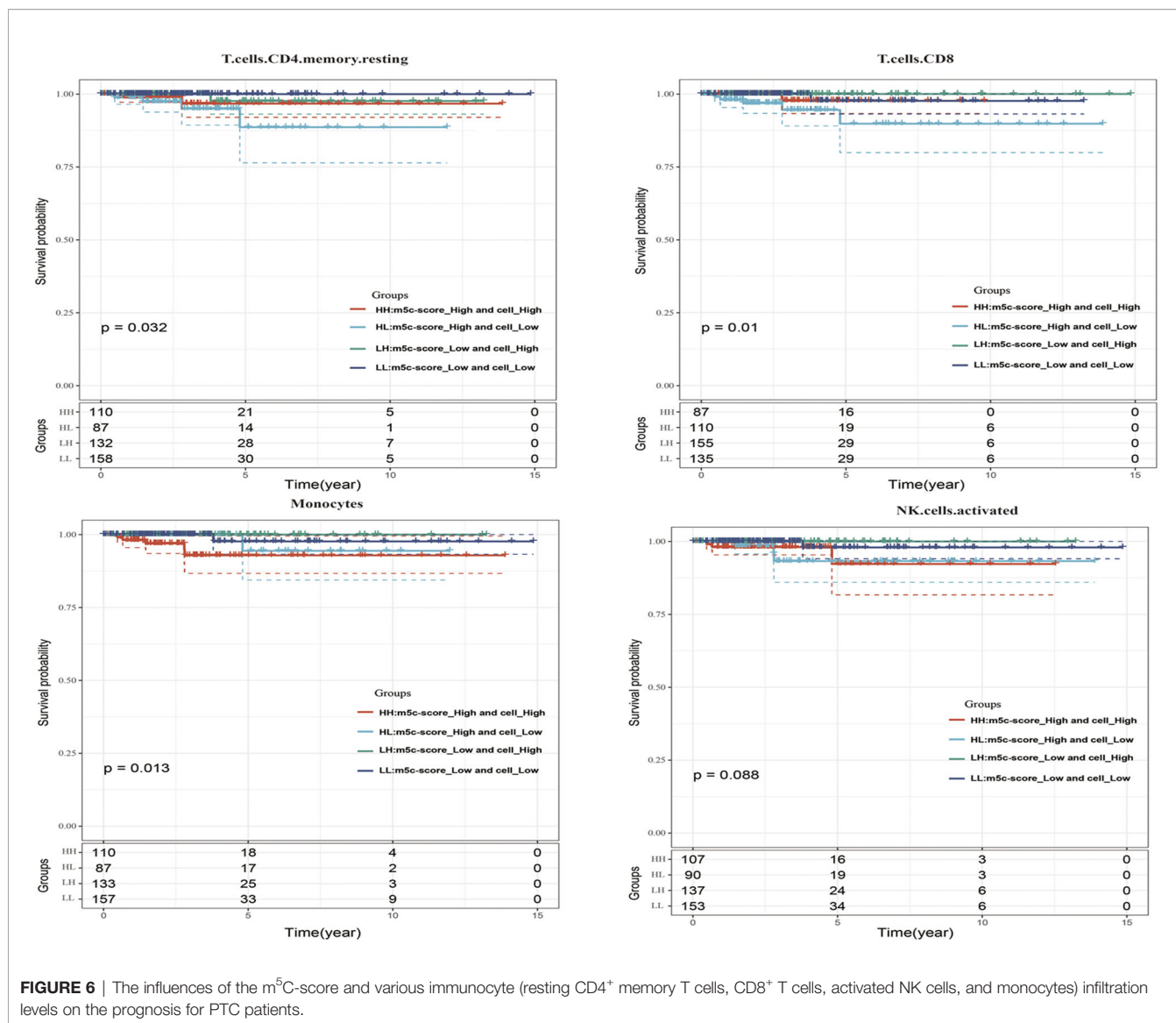


FIGURE 5 | Clinical features and transcriptome traits of the m⁵C-associated phenotypes. **(A)** Alluvial diagram showing the changes of m⁵C modification patterns, gender, age, gene cluster, and the m⁵C-score. **(B)** DEGs between high and low m⁵C-score samples. **(C, D)** Differences in DFS **(C)** and OS **(D)** between high and low m⁵C-score samples. **(E)** Relationship between the m⁵C-score value and the score of CD3+CD4+/CD3+CD8+ cells of the peripheral blood samples of 24 GBM patients. The m⁵C-Score value was negatively associated with the ratio of CD3+CD4+/CD3+CD8+ cells. **(F)** Relationship between the m⁵C-score value and the percentage of CD4+CD25+ Tregs in peripheral blood samples of the 24 GBM patients. The m⁵C-score value was positively related with the percentage of CD4+CD25+ Tregs.

TCGA-TPC samples into tumor samples and para-carcinoma tissue samples, with a classification accuracy of 89.7% and a sensitivity of 98.6%. Of the 551 samples in verification set 1, 538 were accurately classified, with an AUC of 94.2% (**Figure 7B**).

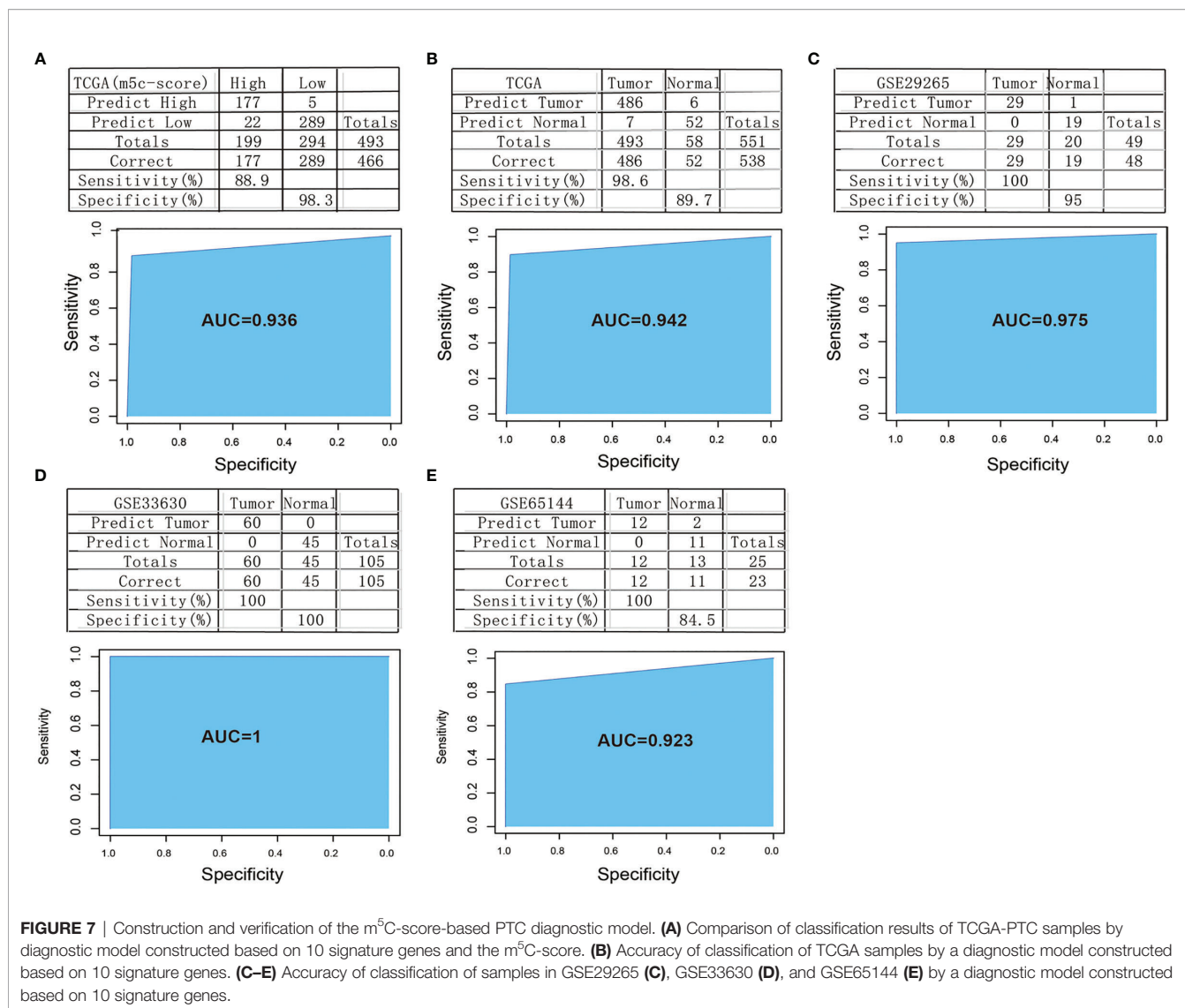
To further verify the model classification efficiency and accuracy, another three sets of microarray data were also downloaded, and the 10 signature genes were used for SVM verification. The GSE29265 dataset was utilized as verification



set 2, which included 49 samples (20 normal samples and 29 tumor samples), with a model classification accuracy of 95%. Of the 49 samples, 48 were accurately classified, the model sensitivity to high and low scores was up to 100%, and the AUC was 97.5% (**Figure 7C**). Meanwhile, the GSE33630 dataset was used as verification set 3, which included 105 samples (45 normal samples and 60 tumor samples). The model classification accuracy reached up to 100%, all the 105 samples were accurately classified, the model sensitivity to high and low scores was 100%, and the AUC was 100 (**Figure 7D**). The GSE65144 dataset was used as verification set 4, which contained 25 samples (13 normal samples and 12 tumor samples). The model classification accuracy was 84.5%, all the 25 samples were accurately classified, the model sensitivity to high and low scores was 100%, and the AUC was 92.3% (**Figure 7E**).

Potential Drug Screening and Evaluation for the m⁵C-Score-Based PTC Diagnostic Model

We firstly used the L1000 fireworks display (L1000FWD) tool, and a reverse drug screening method for differentially expressed genes in high- and low-risk groups of the m⁵C-score and obtained small molecules (drugs, **Supplementary Table 4**). In the interaction database between CMAP drug and gene expression, we analyzed 67 drugs that may interact with genes with different changes in the risk model constructed by the m⁵C-score, and selected 55 small molecules (drugs, **Supplementary Table 5**). We compared the potential drug overlap between L1000 and CMAP annotation, and found that there were five overlapping small molecules (S8), namely cephaeline, emetine, anisomycin, ouabain, and thapsigargin. CCK8 was used to detect the effect of five potential drugs on the growth and metabolic activity of PTC tumor cells. It



was found that compared with the control group, the five drugs could inhibit the growth of thyroid cancer cells in different degrees (**Figure 8A**). Consistent with this, results of subcutaneous transplantation model also showed that intraperitoneal injection of these five drugs could significantly inhibit the growth of tumor, respectively (**Figure 8B**).

DISCUSSION

More and more studies suggest that the m⁵C modification interacts with different m⁵C regulators to play a vital part in anticancer efficacy, inflammation, and intrinsic immunity. A majority of articles have focused on the individual TME cell type or individual regulator, yet no study has completely identified the TME infiltration features mediated by several m⁵C regulators simultaneously. It is important to identify the different m⁵C modification patterns within the TME-infiltrating

cells, so as to display the anticancer immune response in TME and to guide the efficient immunotherapies.

In this study, on the basis of those nine m⁵C regulators, three different m⁵C methylation modification patterns were identified, which showed different TME-infiltrating features. Furthermore, differences in mRNA transcriptome data across different m⁵C modification patterns were suggested to be remarkably related to the biological pathways associated with m⁵C and immunity. Such DEGs were recognized to be the m⁵C-associated signature genes. Consistent with the m⁵C modification phenotype clustering analysis results, three genomic subtypes were found using the m⁵C signature genes, and they showed significant correlations with the immune and stromal activation. According to such results, m⁵C modification played an important role in the formation of diverse TME landscapes. As a result, comprehensively assessing m⁵C modification patterns can shed more light on the features of TME cell infiltration. Due to the differences in individual m⁵C modification patterns, quantifying

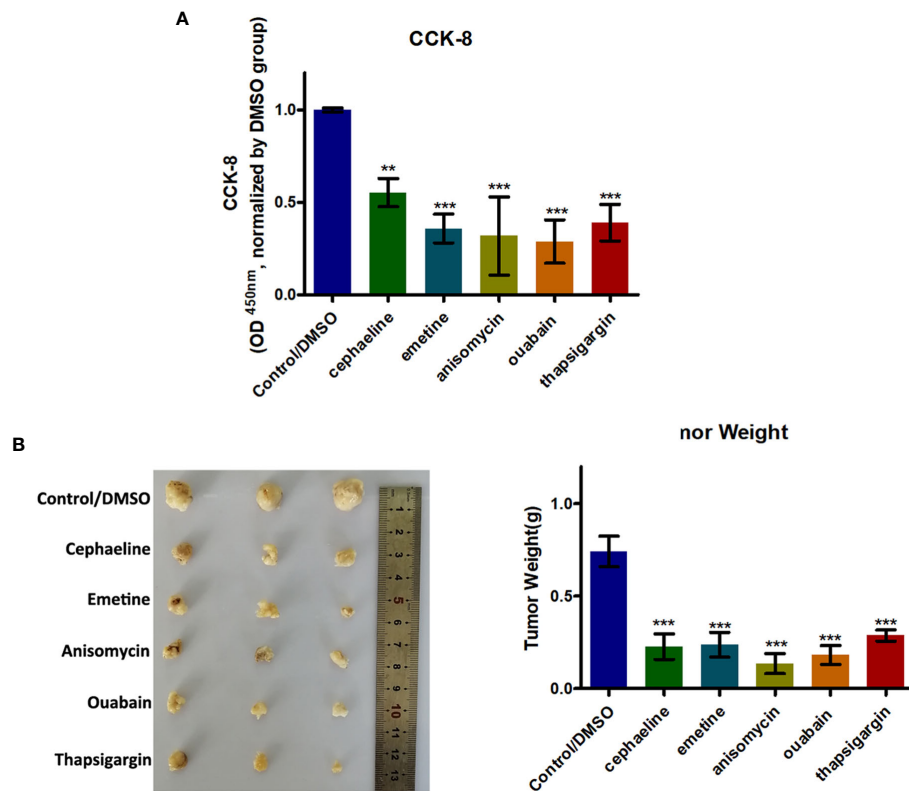


FIGURE 8 | Five potential drugs based on the PTC m⁵C-score model could impair growth of PTC cells. **(A)** Histogram showing the viability of PTC cells with or without five potential drugs (cephaeline, emetine, anisomycin, ouabain, and thapsigargin) for 48 h at 20 μM. **(B)** BALB/c mice were subcutaneously injected with PTC cells. After 5 days, the nude mice were treated with cephaeline, emetine, anisomycin, ouabain, or thapsigargin (100 mg/kg daily, intraperitoneal injection). Tumor weights were measured after 6 weeks (*n* = 5 mice/group). ***statistical significance.

m⁵C modification patterns in individual tumors is necessary. To this end, the scoring system, namely, the m⁵C-score, was constructed in the present work for evaluating m⁵C modification patterns in PTC cases. Our results show the reliability and robustness of the m⁵C-score to comprehensively assess the m⁵C modification patterns of individual tumors, and it might be used to better examine TME infiltration patterns (namely, the immune phenotypes of tumor). Integrative analysis further revealed that the m⁵C-score might serve as a biomarker to independently predict the PTC prognosis. Finally, this study constructed a diagnostic model using the 10 signature genes highly related to the m⁵C-score and discovered that the model exhibited high diagnostic accuracy for PTC.

The m⁵C-score might be adopted clinically for the comprehensive evaluation of m⁵C methylation modification patterns together with related TME cell infiltration characteristics for individual patients, thus contributing to determining the tumor immune phenotypes and guiding efficient clinical practice. Furthermore, the m⁵C-score might also be adopted to assess the clinicopathological characteristics of patients, like molecular subtypes, histological subtypes, tumor mutation burden, tumor inflammation stage, tumor differentiation degree, clinical stages, and genetic variation.

This work elaborated the association of the m⁵C-score with the clinicopathological characteristics. Besides, the m⁵C-score also served as a biomarker to independently predict patient survival. The adjuvant chemotherapy efficacy and clinical anti-PD-1/PD-L1 immunotherapy response of patients were also predicted *via* the established m⁵C-score. Noteworthy, some new points were proposed in this study regarding cancer immunotherapy, which was that it was helpful to target the m⁵C regulators or the m⁵C phenotype-associated genes to alter m⁵C modification patterns, and to reverse the negative TME cell infiltration features, so as to develop new drug combinations and new immunotherapeutics. Results in this study shed new light on boosting immunotherapy response in patients, recognizing the diverse immune phenotypes of tumor and improving the individualized cancer immunotherapy.

To sum up, findings in the present study have illustrated the wide regulatory mechanisms of m⁵C methylation modification patterns in the TME. Heterogeneity in m⁵C modification patterns has been identified as a nonnegligible factor, which may induce TME complexity and heterogeneity. It is important to comprehensively evaluate the m⁵C modification patterns in individual tumors, so as to shed more light on TME cell-infiltrating features and to guide efficient immunotherapies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Review Board of The Second Affiliated Hospital of Anhui Medical University. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

FL and QD conceived and designed the experiments. XP, SH, JZ, XZ, HC, and XL collected the data and performed the analysis. FL, QD, and XP participated in the discussion of the algorithm. FL, XP, SH, and JZ prepared and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by Clinical Research Cultivation Program of The Second Affiliated Hospital of Anhui Medical University (2020LCZD14) and Anhui Provincial Natural Science Foundation (2008085QH406).

ACKNOWLEDGMENTS

We thank the members of the Department of Nuclear Medicine, The Second Affiliated Hospital of Anhui Medical University for technical assistance. This research was supported by Clinical Research Cultivation Program of The Second Affiliated Hospital of Anhui Medical University (2020LCZD14).

REFERENCES

1. Goldenberg D. We Cannot Ignore the Real Component of the Rise in Thyroid Cancer Incidence. *Cancer* (2019) 125:2362–3. doi: 10.1002/cncr.32123
2. Qiu J, Zhang W, Zang C, Liu X, Liu F, Ge R, et al. Identification of Key Genes and miRNAs Markers of Papillary Thyroid Cancer. *Biol Res* (2018) 51:45. doi: 10.1186/s40659-018-0188-1
3. Chengfeng X, Gengming C, Junjia Z, Yunxia L. MicroRNA Signature Predicts Survival in Papillary Thyroid Carcinoma. *J Cell Biochem* (2019) 120:17050–8. doi: 10.1002/jcb.28966
4. Higashino M, Ayani Y, Terada T, Kurisu Y, Hirose Y, Kawata R. Clinical Features of Poorly Differentiated Thyroid Papillary Carcinoma. *Auris Nasus Larynx* (2019) 46:437–42. doi: 10.1016/j.anl.2018.10.001
5. Luo X, Wu A. Analysis of Risk Factors for Postoperative Recurrence of Thyroid Cancer. *J BUON* (2019) 24:813–8.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.729887/full#supplementary-material>

Supplementary Figure 1 | The work flow chart in the present work.

Supplementary Figure 2 | Expression of 11 m⁵C regulators in normal samples and PTC samples in TCGA database.

Supplementary Figure 3 | DEGs between different m⁵C modification patterns. **(A)** DEGs between m⁵C-cluster 1 and m⁵C-cluster 2. **(B)** DEGs between m⁵C-cluster 2 and m⁵C-cluster 3. **(C)** DEGs between m⁵C-cluster 1 and m⁵C-cluster 3.

Supplementary Figure 4 | 3 different m⁵C modification genome phenotypes were classified by unsupervised cluster analysis.

Supplementary Figure 5 | The expression levels of 9 m⁵C regulators in 3 m⁵C gene-cluster subtypes.

Supplementary Figure 6 | Enrichment analysis of DEGs between high and low m⁵C-score samples.

Supplementary Figure 7 | The distribution of clinical features (gender, stage and age) of high and low m⁵C-score samples.

Supplementary Figure 8 | Expression of 9 m⁵C regulators in high and low m⁵C-score samples.

Supplementary Figure 9 | The correlation of m⁵C-score with TME. **(A)** The levels of infiltration of 21 immune cells in high and low m⁵C-score samples. **(B)** Levels of stromal scores, immune scores and ESTIMATE scores in high and low m⁵C-score samples.

Supplementary Figure 10 | The expression levels of 16 immune checkpoints in high and low m⁵C-score samples.

Supplementary Table 1 | Differentially expressed genes screened and analyzed by the R package limma.

Supplementary Table 2 | Results of single factor survival analysis for the 49 genes.

Supplementary Table 3 | Distribution of clinical features of the three subtypes.

Supplementary Table 4 | Potential drugs for PTC screened and analyzed by I1000FWD tool.

Supplementary Table 5 | Potential drugs for PTC screened and analyzed by CMAP tool.

6. Nachtergaele S, He C. The Emerging Biology of RNA Post-Transcriptional Modifications. *RNA Biol* (2017) 14:156–63. doi: 10.1080/15476286.2016.1267096
7. Kiss T. Small Nucleolar RNA-Guided Post-Transcriptional Modification of Cellular RNAs. *EMBO J* (2001) 20:3617–22. doi: 10.1093/emboj/20.14.3617
8. Zhang LS, Liu C, Ma H, Dai Q, Sun HL, Luo G, et al. Transcriptome-Wide Mapping of Internal N(7)-Methylguanosine Methylome in Mammalian mRNA. *Mol Cell* (2019) 74:1304–16.e1308. doi: 10.1016/j.molcel.2018.10.020
9. Song J, Yi C. Chemical Modifications to RNA: A New Layer of Gene Expression Regulation. *ACS Chem Biol* (2017) 12:316–25. doi: 10.1021/acscchembio.6b00960
10. Liu Y, Santi DV. M5c RNA and M5c DNA Methyl Transferases Use Different Cysteine Residues as Catalysts. *Proc Natl Acad Sci USA* (2000) 97:8263–5. doi: 10.1073/pnas.97.15.8263
11. Trecent C, Dlubala A, George P, Pichat P, Ripoche I, Troin Y. Synthesis and Biological Evaluation of Analogues of M6G. *Eur J Med Chem* (2011) 46:4035–41. doi: 10.1016/j.ejmech.2011.05.076

12. Li X, Xiong X, Zhang M, Wang K, Chen Y, Zhou J, et al. Base-Resolution Mapping Reveals Distinct M(1)A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. *Mol Cell* (2017) 68:993–1005.e1009. doi: 10.1016/j.molcel.2017.10.019
13. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, et al. Topology of the Human and Mouse M6a RNA Methylomes Revealed by M6a-Seq. *Nature* (2012) 485:201–6. doi: 10.1038/nature11112
14. Liu RJ, Long T, Li J, Li H, Wang ED. Structural Basis for Substrate Binding and Catalytic Mechanism of a Human RNA:m5C Methyltransferase Nsun6. *Nucleic Acids Res* (2017) 45:6684–97. doi: 10.1093/nar/gkx473
15. Jacob R, Zander S, Gutschner T. The Dark Side of the Epitranscriptome: Chemical Modifications in Long Non-Coding RNAs. *Int J Mol Sci* (2017) 18:2387. doi: 10.3390/ijms18112387
16. Motorin Y, Lyko F, Helm M. 5-Methylcytosine in RNA: Detection, Enzymatic Formation and Biological Functions. *Nucleic Acids Res* (2010) 38:1415–30. doi: 10.1093/nar/gkp1117
17. Yang X, Yang Y, Sun BF, Chen YS, Xu JW, Lai WY, et al. 5-Methylcytosine Promotes mRNA Export - NSUN2 as the Methyltransferase and ALYREF as an M(5)C Reader. *Cell Res* (2017) 27:606–25. doi: 10.1038/cr.2017.55
18. Sibbritt T, Shafik A, Clark SJ, Preiss T. Nucleotide-Level Profiling of M(5)C RNA Methylation. *Methods Mol Biol* (2016) 1358:269–84. doi: 10.1007/978-1-4939-3067-8_16
19. Chen X, Li A, Sun BF, Yang Y, Han YN, Yuan X, et al. 5-Methylcytosine Promotes Pathogenesis of Bladder Cancer Through Stabilizing mRNAs. *Nat Cell Biol* (2019) 21:978–90. doi: 10.1038/s41556-019-0361-y
20. Zhong S, Li C, Han X, Li X, Yang YG, Wang H. Idarubicin Stimulates Cell Cycle and TET2-Dependent Oxidation of DNA 5-Methylcytosine in Cancer Cells. *Chem Res Toxicol* (2019) 32:861–8. doi: 10.1021/acs.chemrestox.9b00012
21. Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, et al. Safety, Activity, and Immune Correlates of Anti-PD-1 Antibody in Cancer. *N Engl J Med* (2012) 366:2443–54. doi: 10.1056/NEJMoa1200690
22. Bergdorf K, Ferguson DC, Mehrad M, Ely K, Stricker T, Weiss VL. Papillary Thyroid Carcinoma Behavior: Clues in the Tumor Microenvironment. *Endocr Relat Cancer* (2019) 26:601–14. doi: 10.1530/ERC-19-0074
23. Means C, Clayburgh DR, Maloney L, Sauer D, Taylor MH, Shindo ML, et al. Tumor Immune Microenvironment Characteristics of Papillary Thyroid Carcinoma Are Associated With Histopathological Aggressiveness and BRAF Mutation Status. *Head Neck* (2019) 41:2636–46. doi: 10.1002/hed.25740
24. Pitt JM, Marabelle A, Eggermont A, Soria JC, Kroemer G, Zitvogel L. Targeting the Tumor Microenvironment: Removing Obstruction to Anticancer Immune Responses and Immunotherapy. *Ann Oncol* (2016) 27:1482–92. doi: 10.1093/annonc/mdw168
25. Quail DF, Joyce JA. Microenvironmental Regulation of Tumor Progression and Metastasis. *Nat Med* (2013) 19:1423–37. doi: 10.1038/nm.3394
26. Song J, Deng Z, Su J, Yuan D, Liu J, Zhu J. Patterns of Immune Infiltration in HNC and Their Clinical Implications: A Gene Expression-Based Study. *Front Oncol* (2019) 9:1285. doi: 10.3389/fonc.2019.01285
27. Ali HR, Chlon L, Pharoah PD, Markowitz F, Caldas C. Patterns of Immune Infiltration in Breast Cancer and Their Clinical Implications: A Gene-Expression-Based Retrospective Study. *PLoS Med* (2016) 13:e1002194. doi: 10.1371/journal.pmed.1002194
28. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the Tumor Immune Microenvironment (TIME) for Effective Therapy. *Nat Med* (2018) 24:541–50. doi: 10.1038/s41591-018-0014-x
29. Fang H, Declerck YA. Targeting the Tumor Microenvironment: From Understanding Pathways to Effective Clinical Trials. *Cancer Res* (2013) 73:4965–77. doi: 10.1158/0008-5472.CAN-13-0661
30. Chen YT, Shen JY, Chen DP, Wu CF, Guo R, Zhang PP, et al. Identification of Cross-Talk Between M(6)A and 5mC Regulators Associated With Immunogenic Features and Prognosis Across 33 Cancer Types. *J Hematol Oncol* (2020) 13:22. doi: 10.1186/s13045-020-00854-w
31. Luchtel RA, Bhagat T, Pradhan K, Jacobs WRJr., Levine M, Verma A, et al. High-Dose Ascorbic Acid Synergizes With Anti-PD1 in a Lymphoma Mouse Model. *Proc Natl Acad Sci USA* (2020) 117:1666–77. doi: 10.1073/pnas.1908158117
32. Tomas G, Tarabichi M, Gacquer D, Hebrant A, Dom G, Dumont JE, et al. A General Method to Derive Robust Organ-Specific Gene Expression-Based Differentiation Indices: Application to Thyroid Cancer Diagnostic. *Oncogene* (2012) 31:4490–8. doi: 10.1038/onc.2011.626
33. von Roemeling CA, Marlow LA, Pinkerton AB, Crist A, Miller J, Tun HW, et al. Aberrant Lipid Metabolism in Anaplastic Thyroid Carcinoma Reveals Stearoyl CoA Desaturase 1 as a Novel Therapeutic Target. *J Clin Endocrinol Metab* (2015) 100:E697–709. doi: 10.1210/jc.2014-2764
34. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAAbiolinks: An R/Bioconductor Package for Integrative Analysis of TCGA Data. *Nucleic Acids Res* (2016) 44:e71. doi: 10.1093/nar/gkv1507
35. Wilkerson MD, Hayes DN. ConsensusClusterPlus: A Class Discovery Tool With Confidence Assessments and Item Tracking. *Bioinformatics* (2010) 26:1572–3. doi: 10.1093/bioinformatics/btq170
36. Hanzelmann S, Castelo R, Guinney J. GSEA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinformatics* (2013) 14:7. doi: 10.1186/1471-2105-14-7
37. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: Gene Set Analysis Toolkit With Revamped UIs and APIs. *Nucleic Acids Res* (2019) 47:W199–205. doi: 10.1093/nar/gkz401
38. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells With CIBERSORT. *Methods Mol Biol* (2018) 1711:243–59. doi: 10.1007/978-1-4939-7493-1_12
39. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-Cancer Immunogenomic Analyses Reveal Genotype-Immuno-phenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep* (2017) 18:248–62. doi: 10.1016/j.celrep.2016.12.019
40. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res* (2015) 43:e47. doi: 10.1093/nar/gkv007
41. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, et al. Gene Expression Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade to Improve Prognosis. *J Natl Cancer Inst* (2006) 98:262–72. doi: 10.1093/jnci/djj052
42. Qian C, Cao X. Dendritic Cells in the Regulation of Immunity and Inflammation. *Semin Immunol* (2018) 35:3–11. doi: 10.1016/j.smim.2017.12.002

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Li, Deng, Pang, Huang, Zhang, Zhu, Chen and Liu. 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.