

# m<sup>5</sup>C regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in papillary thyroid carcinoma

Fei Li (✉ [lifei007@139.com](mailto:lifei007@139.com))

Anhui Medical University <https://orcid.org/0000-0001-6259-8990>

**Qingmei Deng**

Hefei Institutes of Physical Science

**Xiaoxi Pang**

Anhui Medical College: Anhui Medical University

**Shan Huang**

Anhui Medical College: Anhui Medical University

**Jingmiao Zhang**

Anhui Medical College: Anhui Medical University

**Xiaxia Zhu**

Anhui Medical College: Anhui Medical University

**Hong Chen**

Anhui Medical College: Anhui Medical University

**Xiuxia Liu**

Anhui Medical College: Anhui Medical University

---

## Research

**Keywords:** 5-methylcytosine (m<sup>5</sup>C) modification, immune infiltration, subtype, papillary thyroid carcinoma (PTC), tumor microenvironment (TME)

**Posted Date:** September 25th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-82174/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## *Background*

Recently, immune response modulation at epigenetic level is illustrated in studies, but it is still unclear about the possible function of RNA 5-methylcytosine (m<sup>5</sup>C) modification in the cell infiltration within tumor microenvironment (TME).

## *Methods*

In this study, the m<sup>5</sup>C modification patterns from altogether 493 papillary thyroid carcinoma (PTC) samples were assessed completely according to 9 m<sup>5</sup>C regulators. Afterwards, the modification patterns were correlated with the cell infiltration features in TME. The m<sup>5</sup>C modification patterns in tumor samples were quantified through the principal component analysis (PCA) algorithms to establish the m<sup>5</sup>C-score. Moreover, this study mined the signature genes related to the m<sup>5</sup>C-score, constructed the PTC diagnostic model by the support vector machine (SVM) method, and verified its accuracy based on samples in TCGA and GEO databases. The effects of 5 potential drugs based on the PTC m<sup>5</sup>C-score model in PTC cells were investigated through in vitro and in vivo assays (i.e., cell counting kit 8 and xenograft model).

## *Results*

A total of 3 different m<sup>5</sup>C modification patterns were identified, and high differentiation degree was observed in the cell infiltration features within TME under the above 3 identified patterns. It was revealed that, evaluating m<sup>5</sup>C modification patterns in single tumor samples helped to estimate the stromal activity in TME, expression of immune checkpoint genes, and prognosis for patients. Typically, a low m<sup>5</sup>C-score, which was reflected in the activated immunity, predicted the relatively favorable prognostic outcome. Few effective immune infiltration was seen in high m<sup>5</sup>C-score subtype, indicating the dismal patient survival. Finally, this study constructed a diagnostic model using the 10 signature genes highly related to the m<sup>5</sup>C-score, discovered that the model exhibited high diagnostic accuracy for PTC, and screened out 5 potential drugs for PTC based on this m<sup>5</sup>C-score model.

## *Conclusions*

According to findings in the present study, m<sup>5</sup>C modification exerts an important part in forming the TME complexity and diversity. It is valuable to evaluate the m<sup>5</sup>C modification patterns in single tumors, so as to enhance our understanding towards the infiltration characterization in TME and to better guide clinical diagnosis as well as immunotherapies.

## **Background**

Thyroid cancer (THCA), a frequently-occurring endocrine cancer, takes up approximately 1.7% of human cancers[1]. TC can be divided as 4 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, around 30% PTC cases suffer from tumor recurrence[5]. As a result, analyzing the disease features at molecular level is essential.

It is increasingly suggested that, RNA modification at 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 3rd epigenetics layer, such as N1-methyladenosine ( $m^1A$ ), N6-methyladenosine ( $m^6A$ ), together with 5-methylcytosine ( $m^5C$ )[9–13].

Of them,  $m^5C$  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].  $m^5C$  in mRNAs has been extensively studied, and many articles reveal that  $m^5C$  greatly affects mRNAs, tRNAs, and rRNAs[16]. The  $m^5C$  methylation is related to various regulators, such as the  $m^5C$  “readers”, demethylases and methyltransferases. Typically, the methyltransferase “writer” complex enhances RNA methylation at 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  $m^5C$  modification through the degradation of written methylation. The adenosine demethylases, methyltransferases, together with the RNA-binding proteins involved in  $m^5C$  modification are referred to as the  $m^5C$  “erasers” (like TET2),  $m^5C$  “writers” (like NSUN1-7, DNMT1-2, and DNMT3A-3B), as well as  $m^5C$  “readers” (like ALYREF)[17]. More and more studies suggest that  $m^5C$  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, it is still unclear about the typical gene signatures, together with the diagnostic and prognostic significance of  $m^5C$ -related regulators in PTC.

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), as well as new vessels[23]. Notably, the tumor-associated myeloid cells (TAMCs) are constituted by 5

different myeloid subsets, namely, myeloid-derived suppressor cells (MDSCs), tumor-associated macrophages (TAM), 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, 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 help to identify the diverse tumor immune phenotypes and to guide and predict responses to immunotherapies[28]. Further, 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 m<sup>5</sup>C modification, and such relationship can not 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 m<sup>5</sup>C regulators, so as to shed more lights on the TME immunomodulation. The present work combined genome data from 493 TCGA-PTC samples for the comprehensive evaluation of m<sup>5</sup>C modification patterns, and related them to cell infiltration features within TME. Altogether 3 different m<sup>5</sup>C modification patterns were identified, under which the high differentiation degree of TME features were found, indicating the critical part of m<sup>5</sup>C modification in forming individual TME features. On this basis, the scoring system was also established for the quantification of m<sup>5</sup>C modification patterns for individual cases. Finally, this study mined the m<sup>5</sup>C-score related signature genes to construct the PTC diagnostic model using the support vector machine (SVM) method.

## Methods

### Source and preprocessing of PTC data

The work flow chart in the present work is presented in **Supplementary Fig. 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], 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 m<sup>5</sup>C regulators

Altogether 13 regulators were obtained based on TCGA datasets to identify the diverse m<sup>5</sup>C regulators-mediated m<sup>5</sup>C modification patterns. All the 13 genes, except for ALYREF and NSUN1, were with available expression profiles. The remaining 11 m<sup>5</sup>C regulators contained 1 eraser (TET2) and 10 writers (NSUN2-7, DNMT1-2, 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 Fig. 2**). Later, consensus clustering was adopted for identifying the different m<sup>5</sup>C modification patterns according to 9 m<sup>5</sup>C regulators 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 (GSVA) together with functional annotation

For investigating the heterogeneities in biological process among the m<sup>5</sup>C 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 (DEGs) across the different m<sup>5</sup>C phenotypes

For identifying the m<sup>5</sup>C-associated genes, the patients were divided to 3 different m<sup>5</sup>C modification patterns according to 9 m<sup>5</sup>C 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.

## m<sup>5</sup>C gene signature construction

For quantifying m<sup>5</sup>C modification patterns among individual tumors, the scoring system, m<sup>5</sup>C-score, was built based on the m<sup>5</sup>C gene signature, as shown below:

First of all, DEGs obtained based on the diverse m<sup>5</sup>C-clusters were subjected to normalization across all samples, then, those overlapped genes were selected. Afterwards, all cases were divided to 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 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 m<sup>5</sup>C 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, m<sup>5</sup>C-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 m<sup>5</sup>C phenotype-associated gene expression levels.

## m<sup>5</sup>C-Score based PTC diagnostic model establishment

First of all, this study mined the signature genes significantly correlated with m<sup>5</sup>C-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 m<sup>5</sup>C regulator expression levels. Thereafter, Kruskal-Wallis test and one-way ANOVA were applied in comparing the heterogeneities among three groups. Based on correlations of m<sup>5</sup>C-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 m<sup>5</sup>C-score by testing the possible threshold values to find the maximal rank statistic. Later, all cases were classified as high or low m<sup>5</sup>C-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 (HR) for the m<sup>5</sup>C regulators as well as the m<sup>5</sup>C phenotype-associated genes. Meanwhile, the receiver operating characteristic (ROC) curve was plotted to assess the sensitivity and specificity of our diagnostic model and the m<sup>5</sup>C-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 9 regulators-mediated m<sup>5</sup>C methylation modification patterns

According to 9 m<sup>5</sup>C regulators with expression profiles in the TCGA-PTC dataset, PTC samples were identified from normal samples (Fig. 1A). Afterwards, the expression profile data of these 9 m<sup>5</sup>C regulators were carried out z-score standardization using the scale function in mosaic package. Then, 3 different m<sup>5</sup>C modification patterns were discovered according to those 9 m<sup>5</sup>C regulators expression patterns (Fig. 1B and C). These 3 patterns were named as m<sup>5</sup>C-cluster 1–3. It was observed from Fig. 1D that, the expression level of these 9 m<sup>5</sup>C regulators showed significant differences among the 3 distinct subtype samples.

Prognostic analysis was also carried out for these 3 major m<sup>5</sup>C modification patterns, which suggested that m<sup>5</sup>C-cluster 2 modification pattern showed survival advantage (Fig. 1E). However, due to the speciality of PTC and the good overall prognosis, there was no significant statistical difference among these 3 subtypes. Besides, average survival time of samples in these 3 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 m<sup>5</sup>C modification patterns

For exploring those biological behaviors in the different m<sup>5</sup>C modification patterns, GSVA was conducted. It was illustrated from Fig. 2A that, m<sup>5</sup>C-cluster 1 significantly associated with the amino acid metabolic pathways, m<sup>5</sup>C-cluster 2 was enriched to the endocrine system, lipid metabolism, and cancer, whereas m<sup>5</sup>C cluster-3 was associated with cell cycle, DNA repair and nucleic acid metabolism.

Further, the distribution of clinical features of samples in the above 3 subtypes was statistically analyzed. The statistical results are displayed in **Supplementary Table 3** and Fig. 3. It was found from the results that, multiple clinical features in the 3 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 3 subtype samples. As shown in Fig. 2B, the stromal score in m<sup>5</sup>C-cluster 2 was the highest, followed by m<sup>5</sup>C cluster-3, while m<sup>5</sup>C 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 3 m<sup>5</sup>C modification patterns (Fig. 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 m<sup>5</sup>C-cluster 1 and m<sup>5</sup>C-cluster 3, whereas excessive resting/activated DCs were found in m<sup>5</sup>C-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 3 subtype samples. As found from Fig. 2D, there were significant differences in the expression of these 34 immune checkpoints among the 3 subtypes. Most immune checkpoint genes were highly expressed in m<sup>5</sup>C-cluster 2, followed by m<sup>5</sup>C-cluster 3, while m<sup>5</sup>C-cluster 1 had the lowest expression, which was consistent with the average survival time of samples in the 3 subtypes.

## **m<sup>5</sup>C gene signature establishment along with functional annotation**

For better investigating the possible biological behaviors of all the m<sup>5</sup>C modification patterns, the limma package was used to determine 690 m<sup>5</sup>C phenotype-associated DEGs (**Supplementary Fig. 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 m<sup>5</sup>C modification in cancer cells themselves and in TME immunomodulation (Fig. 4A). For better validating such regulatory mechanism, the unsupervised clustering analysis was performed using those 690 m<sup>5</sup>C phenotype-associated genes, for the sake of classifying cases to distinct genome subtypes. Similar to clustering analysis of m<sup>5</sup>C modification patterns, 3 different m<sup>5</sup>C modification genome phenotypes were found, which were referred to as m<sup>5</sup>C gene-cluster A-C, separately (**Supplementary Fig. 4**). According to such results, there were 3 different m<sup>5</sup>C methylation modification patterns in PTC. Besides, there were diverse signature genes in the 3 different gene clusters (**Supplementary Fig. 4**). The m<sup>5</sup>C regulators expression levels were significantly different among the 3 m<sup>5</sup>C gene-clusters (**Supplementary Fig. 5**), consistent with the results obtained for m<sup>5</sup>C methylation modification patterns. The expression quantities of these 9 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 m<sup>5</sup>C-associated phenotypes**

First of all, we analyzed the stromal scores of 3 m<sup>5</sup>C gene-cluster subtypes. The results suggested that (Fig. 4B), there were significant differences in the stromal score of 3 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 3 m<sup>5</sup>C gene-cluster subtypes. As observed from Fig. 4C, the distribution of 15 immunocytes in three subtypes showed statistically significant differences. These findings revealed the important role of m<sup>5</sup>C 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 m<sup>5</sup>C methylation modification patterns of individual cases. Due to the m<sup>5</sup>C modification complexity and heterogeneity in individual samples, this study established the scoring system (m<sup>5</sup>C-score) using the phenotype-associated genes for quantifying m<sup>5</sup>C modification patterns in individual PTC cases. Besides, those attribute alterations in individual patients were visualized by the alluvial diagram (Fig. 5A). It was discovered from the figure that, among the 3 m<sup>5</sup>C-cluster subtypes, samples in m<sup>5</sup>C-cluster 2 and m<sup>5</sup>C-cluster 3 subtypes were mostly distributed in the low m<sup>5</sup>C-score score group, while those in high m<sup>5</sup>C-score score group were basically derived from the m<sup>5</sup>C-cluster 1 subtype. In the 3 m<sup>5</sup>C gene-cluster subtypes, the m<sup>5</sup>C-score values of Cluster A and Cluster C samples were lower. Samples aged over 40 years were mostly classified into the low m<sup>5</sup>C-score score group, while females mostly belonged to the high m<sup>5</sup>C-score 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 logFC > log<sub>2</sub>(1.2) and p < 0.05, 67 DEGs were screened, including 58 up-regulated and 9 down-regulated ones (Fig. 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 Fig. 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 m<sup>5</sup>C-score with patient survival, and analyzed the difference in prognosis between high and low m<sup>5</sup>C-score score samples. The results suggested that, samples with low m<sup>5</sup>C-score scores had better prognosis than those with high score, regardless of DFS or OS (Fig. 5C and 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 m<sup>5</sup>C-score samples (**Supplementary Fig. 7**). The expression levels of 9 m<sup>5</sup>C regulators in high m<sup>5</sup>C-score group were significantly higher than those in low score group, and there was significant difference between two groups (**Supplementary Fig. 8**).

Subsequently, this study observed the correlation of m<sup>5</sup>C-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 m<sup>5</sup>C-score TCGA-TPC samples. The results are presented in **Supplementary Fig. 9A**. There were significant

differences in 6 cell types between high and low m<sup>5</sup>C-score groups. In addition, this study also calculated the stromal score, immune Score and ESTIMATE score in different samples. As presented in **Supplementary Fig. 9B**, in the low m<sup>5</sup>C-score group, the immune Score was significantly higher than that in high m<sup>5</sup>C-score group, which was consistent with the previous results that the low m<sup>5</sup>C-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 genes expression levels between high and low m<sup>5</sup>C-score groups (**Supplementary Figure S10**). Based on these findings, low m<sup>5</sup>C-score showed close correlation with immune activation. Further, m<sup>5</sup>C-score helped to assess m<sup>5</sup>C 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.

At last, this study integrated the influences of m<sup>5</sup>C-score and various immunocyte infiltration levels on the prognosis for PTC patients. From Fig. 6, it was discovered that, resting CD4<sup>+</sup> memory T cells and CD8<sup>+</sup> T cells were mainly enriched in low m<sup>5</sup>C-score samples, while activated NK cells and Monocytes were mostly enriched in high m<sup>5</sup>C-score group. Then, the median infiltration level of the above 4 cell types was used to divide all samples into high and low immunocyte infiltration level groups. It was discovered that, samples with low m<sup>5</sup>C-score and low infiltration level of resting CD4<sup>+</sup> memory T cells had the best prognosis, while those with high m<sup>5</sup>C-score and low infiltration level of resting CD4<sup>+</sup> memory T cells had the poorest prognosis. In addition, samples with low m<sup>5</sup>C-score and high CD8<sup>+</sup> T cells infiltration had the best prognosis, while those with high m<sup>5</sup>C-score and low CD8<sup>+</sup> T cells infiltration had the poorest prognosis. Further, it was found that samples with low m<sup>5</sup>C-score and high Monocytes infiltration had the best prognosis, while those with high m<sup>5</sup>C-score and high Monocytes infiltration had the poorest prognosis. According to the prognostic prediction model, we analysed the correlation between m<sup>5</sup>C-score and Treg expression in 24 PTC patients. The m<sup>5</sup>C-score showed a negative relationship with CD3 + CD4<sup>+</sup>/CD3 + CD8<sup>+</sup> ( $r = -0.9543$ ,  $p < 0.0001$ ; Fig. 5E), but a positive relationship with CD4 + CD25 + Tregs percentage ( $r = 0.4477$ ,  $p = 0.015$ ; Fig. 5F).

## Construction and verification of the m<sup>5</sup>C-score-based PTC diagnostic model

First of all, this study calculated the correlation of 49 m<sup>5</sup>C phenotype-related genes with m<sup>5</sup>C-score. Then, 10 signature genes related to the m<sup>5</sup>C-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 m<sup>5</sup>C-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 m<sup>5</sup>C-score classification results, the accuracy reached 98.3%, and the sensitivity was up to 88.9%. The 493

samples were accurately classified, with the area under the ROC curve (AUC) of 0.936 (Fig. 7A). The above results demonstrated that, the classification model constructed based on these 10 signature genes well simulated the classification results of  $m^5C$ -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 above-mentioned 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 TCGA-TPC samples into tumor samples and para-carcinoma tissue samples, with the classification accuracy of 89.7% and the sensitivity of 98.6%. 538 of the 551 samples in verification set 1 were accurately classified, with the AUC of 94.2% (Fig. 7B).

To further verify the model classification efficiency and accuracy, another 3 sets of microarray data were also downloaded, and the 10 signature genes were used for SVM verification. The GSE29265 data set was utilized as the verification set 2, which included 49 samples (20 normal samples and 29 tumor samples), with the model classification accuracy of 95%. 48 of the 49 samples were accurately classified, the model sensitivity to high and low scores was up to 100%, and the AUC was 97.5% (Fig. 7C).

Meanwhile, the GSE33630 data set was used as the 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 (Fig. 7D). The GSE65144 data set was used as the 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% (Fig. 7E).

## Potential Drug screening and evaluation for the $m^5C$ -score-based PTC diagnostic model

We firstly used L1000 fireworks display (l1000FWD) tool, and reverse drug screening method for differentially expressed genes in high and low-risk groups of  $m^5C$  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  $m^5C$  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 (Fig. 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 (Fig. 8B).

## Discussion

More and more studies suggest that, the m<sup>5</sup>C modification interacts with different m<sup>5</sup>C 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 m<sup>5</sup>C regulators simultaneously. It is important to identify the different m<sup>5</sup>C 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 9 m<sup>5</sup>C regulators, 3 different m<sup>5</sup>C methylation modification patterns were identified, which showed different TME-infiltrating features. Furthermore, differences in mRNA transcriptome data across different m<sup>5</sup>C modification patterns were suggested to be remarkably related to the biological pathways associated with m<sup>5</sup>C and immunity. Such DEGs were recognized to be the m<sup>5</sup>C-associated signature genes. Consistent with the m<sup>5</sup>C modification phenotype clustering analysis results, 3 genomic subtypes were found using the m<sup>5</sup>C signature genes, and they showed significant correlations with the immune and stromal activation. According to such results, m<sup>5</sup>C modification played an important role in the formation of diverse TME landscapes. As a result, comprehensively assessing m<sup>5</sup>C modification patterns can shed more lights on the features of TME cell infiltration. Due to the differences in individual m<sup>5</sup>C modification patterns, quantifying m<sup>5</sup>C modification patterns in individual tumors is necessary. To this end, the scoring system, namely the m<sup>5</sup>C-score, was constructed in the present work for evaluating m<sup>5</sup>C modification patterns in PTC cases. According to our results, the reliability and robustness of the m<sup>5</sup>C-score to comprehensively assess the m<sup>5</sup>C 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, m<sup>5</sup>C-score might served as the biomarker to independently predict the PTC prognosis. Finally, this study constructed a diagnostic model using the 10 signature genes highly related to the m<sup>5</sup>C-score and discovered that the model exhibited high diagnostic accuracy for PTC.

The m<sup>5</sup>C-score might be adopted clinically for the comprehensive evaluation of m<sup>5</sup>C 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, m<sup>5</sup>C-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, genetic variation. This work elaborated the association of m<sup>5</sup>C-score with the clinicopathological characteristics. Besides, m<sup>5</sup>C-score also served as the 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 m<sup>5</sup>C-score. Noteworthy, some new points were proposed in this study regarding cancer immunotherapy, which was that, it was helpful to target the m<sup>5</sup>C regulators or the m<sup>5</sup>C phenotype-associated genes to alter m<sup>5</sup>C

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 lights on boosting immunotherapy response in patients, recognizing the diverse immune phenotypes of tumor and improving the individualized cancer immunotherapy.

## Conclusions

To sum up, findings in the present study have illustrate the wide regulatory mechanisms of m<sup>5</sup>C methylation modification patterns in the TME. Heterogeneity in m<sup>5</sup>C modification patterns has been identified as the nonnegligible factor, which may induce the TME complexity and heterogeneity. It is important to comprehensively evaluate the m<sup>5</sup>C modification patterns in individual tumors, so as to shed more lights on TME cell-infiltrating features and to guide efficient immunotherapies.

## Declarations

## Conflicts of Interest

The authors declare no conflicts of interest.

## Ethics approval and consent to participate

This study was reviewed and approved by the Institutional Review Board of The Second Affiliated Hospital of Anhui Medical University, and written informed consent was obtained from patients based on the Declaration of Helsinki.

## Funding

This research was supported by Anhui Provincial Natural Science Foundation (2008085QH406).

## Author Contributions

FL and QMD: conceived and designed the experiments. XXP, SH, JMZ, XXZ, HC and XXL: collected the data, and performed the analysis. FL, QMD, and XXP: participated in the discussion of the algorithm. FL, XXP, SH, and JMZ: prepared and edited the manuscript. All authors have read and approved the final manuscript.

## Acknowledgement

We thank the members of technical assistance in Department of Nuclear Medicine, The Second Affiliated Hospital of Anhui Medical University.

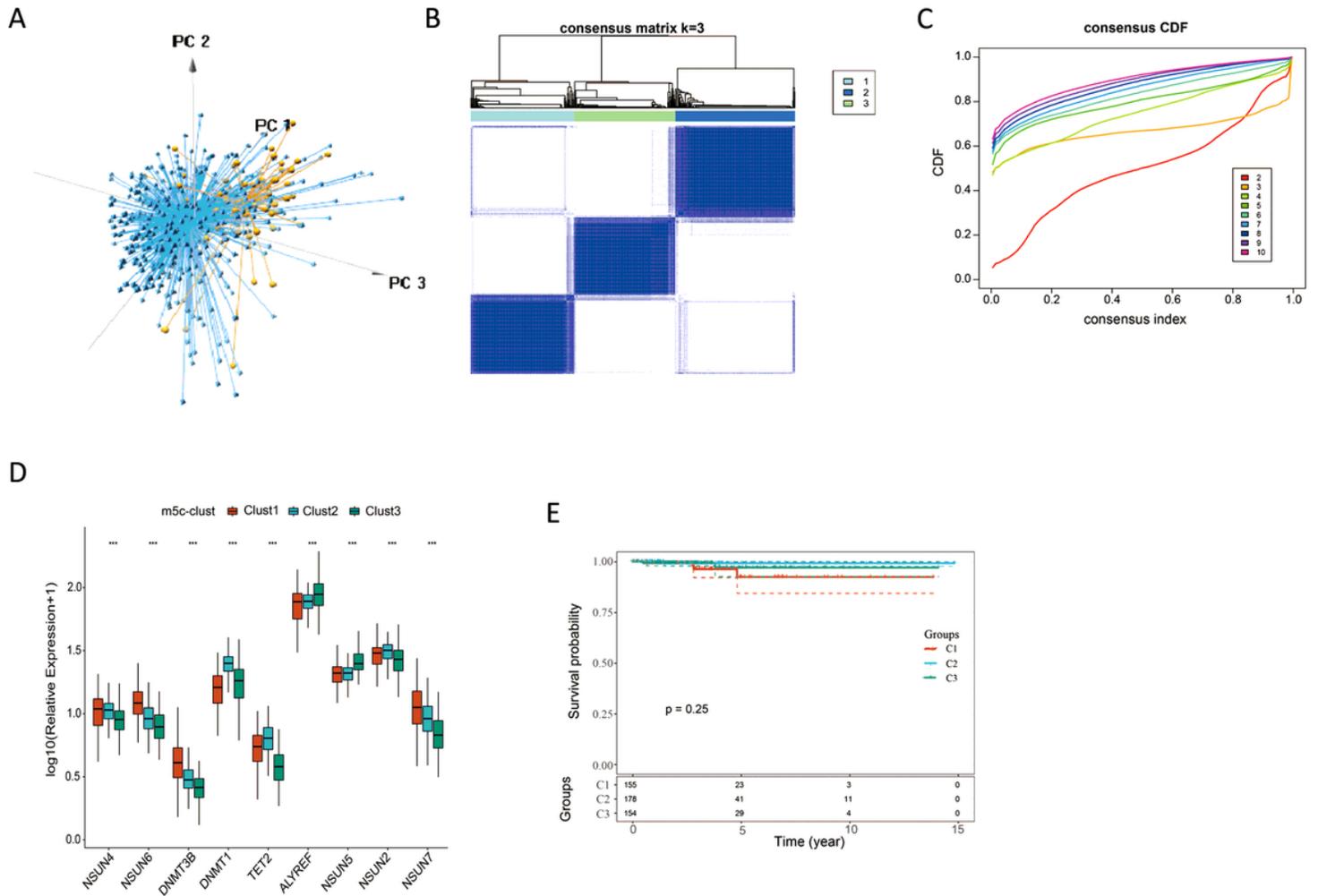
## References

1. Goldenberg D. We cannot ignore the real component of the rise in thyroid cancer incidence. *Cancer*. 2019; 125(14):2362-2363.
2. Qiu J, Zhang W, Zang C, Liu X, Liu F, Ge R, Sun Y, Xia Q. Identification of key genes and miRNAs markers of papillary thyroid cancer. *Biol Res*. 2018; 51(1):45.
3. Chengfeng X, Gengming C, Junjia Z, Yunxia L. MicroRNA signature predicts survival in papillary thyroid carcinoma. *J Cell Biochem*. 2019; 120(10):17050-17058.
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(3):437-442.
5. Luo X, Wu A. Analysis of risk factors for postoperative recurrence of thyroid cancer. *J BUON*. 2019; 24(2):813-818.
6. Nachtergaele S, He C. The emerging biology of RNA post-transcriptional modifications. *RNA Biol*. 2017; 14(2):156-163.
7. Kiss T. Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J*. 2001; 20(14):3617-3622.
8. Zhang LS, Liu C, Ma H, Dai Q, Sun HL, Luo G, Zhang Z, Zhang L, Hu L, Dong X *et al*. Transcriptome-wide Mapping of Internal N(7)-Methylguanosine Methylome in Mammalian mRNA. *Mol Cell*. 2019; 74(6):1304-1316 e1308.
9. Song J, Yi C. Chemical Modifications to RNA: A New Layer of Gene Expression Regulation. *ACS Chem Biol*. 2017; 12(2):316-325.
10. Liu Y, Santi DV. m5C RNA and m5C DNA methyl transferases use different cysteine residues as catalysts. *Proc Natl Acad Sci U S A*. 2000; 97(15):8263-8265.
11. Trecant C, Dlubala A, George P, Pichat P, Ripoché I, Troin Y. Synthesis and biological evaluation of analogues of M6G. *Eur J Med Chem*. 2011; 46(9):4035-4041.
12. Li X, Xiong X, Zhang M, Wang K, Chen Y, Zhou J, Mao Y, Lv J, Yi D, Chen XW *et al*. Base-Resolution Mapping Reveals Distinct m(1)A Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. *Mol Cell*. 2017; 68(5):993-1005 e1009.
13. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M *et al*. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*. 2012; 485(7397):201-206.
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(11):6684-6697.
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(11).

16. Motorin Y, Lyko F, Helm M. 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. *Nucleic Acids Res.* 2010; 38(5):1415-1430.
17. Yang X, Yang Y, Sun BF, Chen YS, Xu JW, Lai WY, Li A, Wang X, Bhattarai DP, Xiao W *et al.* 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. *Cell Res.* 2017; 27(5):606-625.
18. Sibbritt T, Shafik A, Clark SJ, Preiss T. Nucleotide-Level Profiling of m(5)C RNA Methylation. *Methods Mol Biol.* 2016; 1358:269-284.
19. Chen X, Li A, Sun BF, Yang Y, Han YN, Yuan X, Chen RX, Wei WS, Liu Y, Gao CC *et al.* 5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs. *Nat Cell Biol.* 2019; 21(8):978-990.
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(5):861-868.
21. Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, Powderly JD, Carvajal RD, Sosman JA, Atkins MB *et al.* Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med.* 2012; 366(26):2443-2454.
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(6):601-614.
23. Means C, Clayburgh DR, Maloney L, Sauer D, Taylor MH, Shindo ML, Coussens LM, Tsujikawa T. Tumor immune microenvironment characteristics of papillary thyroid carcinoma are associated with histopathological aggressiveness and BRAF mutation status. *Head Neck.* 2019; 41(8):2636-2646.
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(8):1482-1492.
25. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med.* 2013; 19(11):1423-1437.
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.
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(12):e1002194.
28. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, Coussens LM, Gaborilovich DI, Ostrand-Rosenberg S, Hedrick CC *et al.* Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med.* 2018; 24(5):541-550.
29. Fang H, Declerck YA. Targeting the tumor microenvironment: from understanding pathways to effective clinical trials. *Cancer Res.* 2013; 73(16):4965-4977.
30. Chen YT, Shen JY, Chen DP, Wu CF, Guo R, Zhang PP, Lv JW, Li WF, Wang ZX, Chen YP. Identification of cross-talk between m(6)A and 5mC regulators associated with onco-immunogenic features and prognosis across 33 cancer types. *J Hematol Oncol.* 2020; 13(1):22.

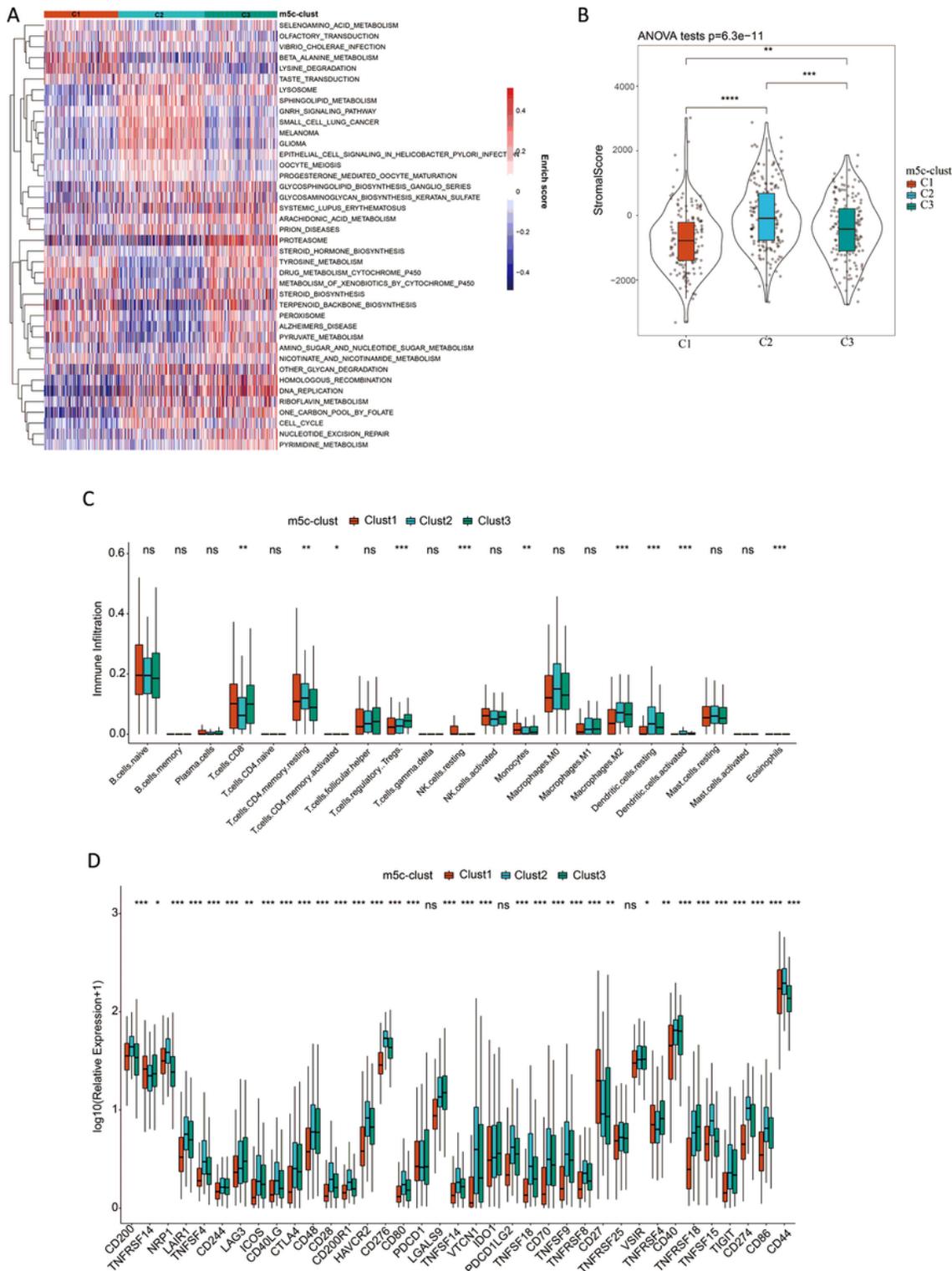
31. Luchtel RA, Bhagat T, Pradhan K, Jacobs WR, Jr., Levine M, Verma A, Shenoy N. High-dose ascorbic acid synergizes with anti-PD1 in a lymphoma mouse model. *Proc Natl Acad Sci U S A*. 2020; 117(3):1666-1677.
32. Tomas G, Tarabichi M, Gacquer D, Hebrant A, Dom G, Dumont JE, Keutgen X, Fahey TJ, 3rd, Maenhaut C, Detours V. A general method to derive robust organ-specific gene expression-based differentiation indices: application to thyroid cancer diagnostic. *Oncogene*. 2012; 31(41):4490-4498.
33. von Roemeling CA, Marlow LA, Pinkerton AB, Crist A, Miller J, Tun HW, Smallridge RC, Copland JA. Aberrant lipid metabolism in anaplastic thyroid carcinoma reveals stearyl CoA desaturase 1 as a novel therapeutic target. *J Clin Endocrinol Metab*. 2015; 100(5):E697-709.
34. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I *et al*. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016; 44(8):e71.
35. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010; 26(12):1572-1573.
36. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013; 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(W1):W199-W205.
38. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol*. 2018; 1711:243-259.
39. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep*. 2017; 18(1):248-262.
40. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015; 43(7):e47.
41. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, Nordgren H, Farmer P, Praz V, Haibe-Kains B *et al*. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst*. 2006; 98(4):262-272.
42. Qian C, Cao X. Dendritic cells in the regulation of immunity and inflammation. *Semin Immunol*. 2018; 35:3-11.

## Figures



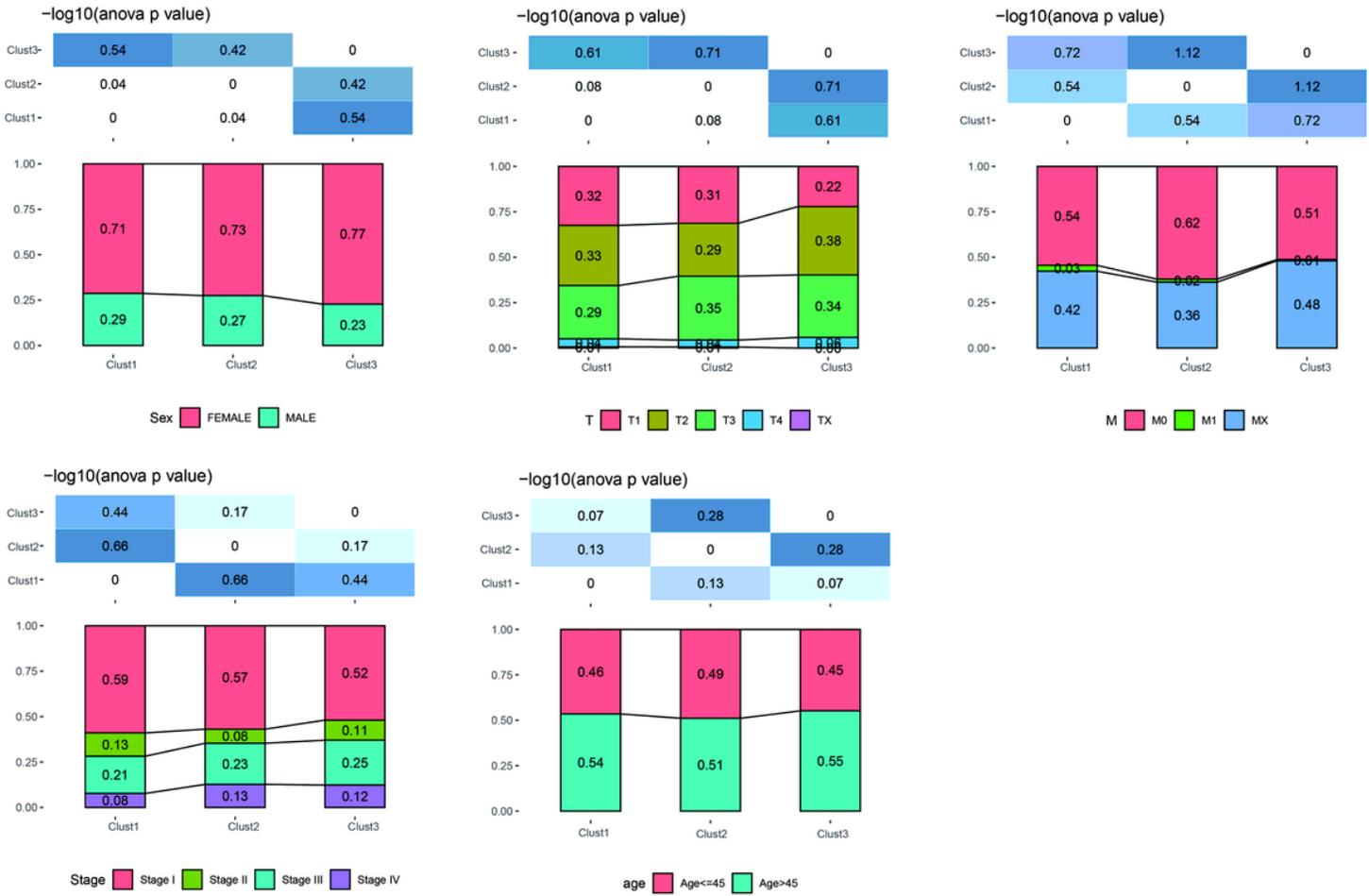
**Figure 1**

The 9 regulators-mediated m5C methylation modification patterns. (A) PCA analysis based on 9 m5C regulators. (B-C) Consensus clustering was adopted for identifying the different m5C modification patterns. (D) The expression levels of 9 m5C regulators in different m5C modification patterns. (E) Survival analysis of the three subtypes.



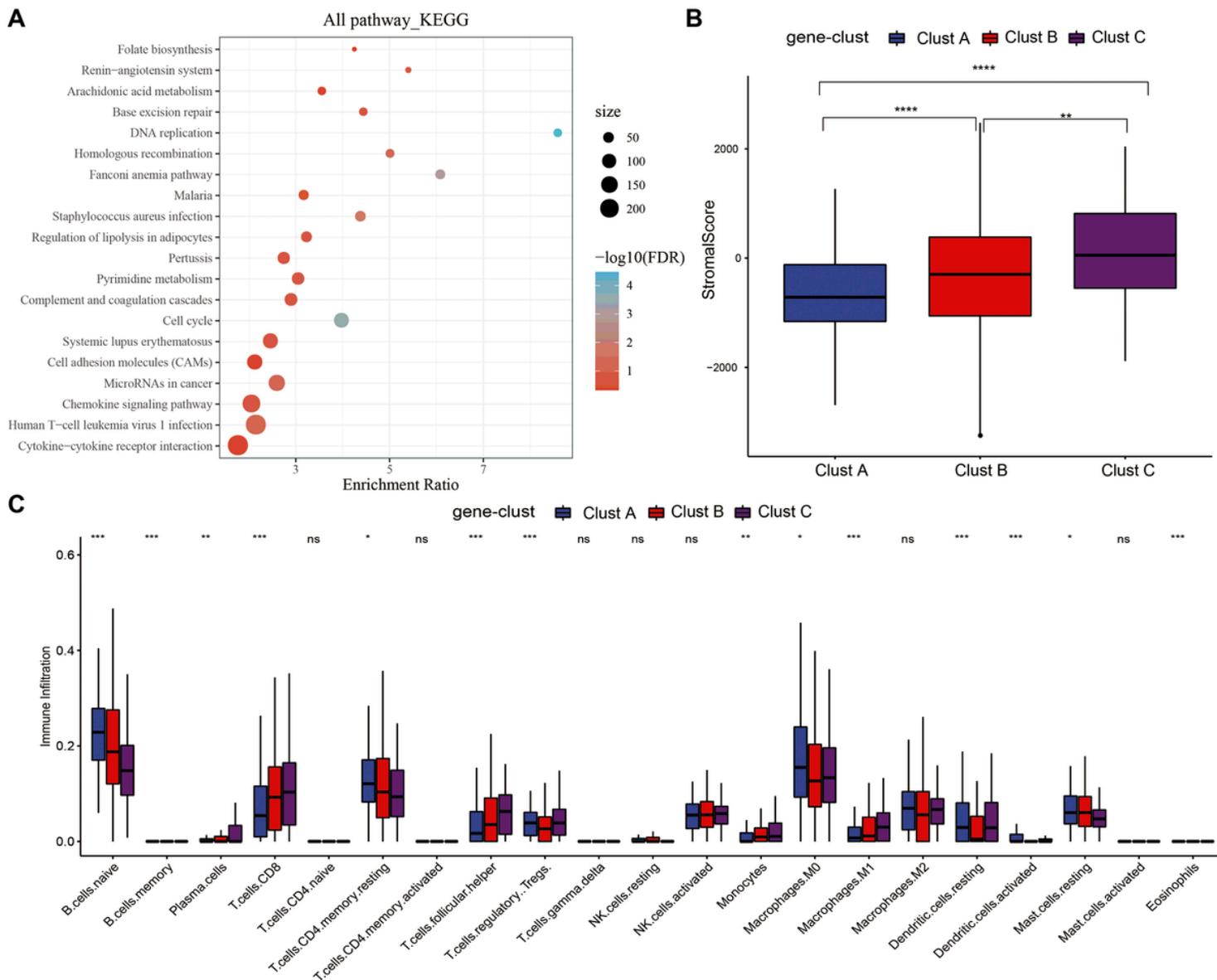
**Figure 2**

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.



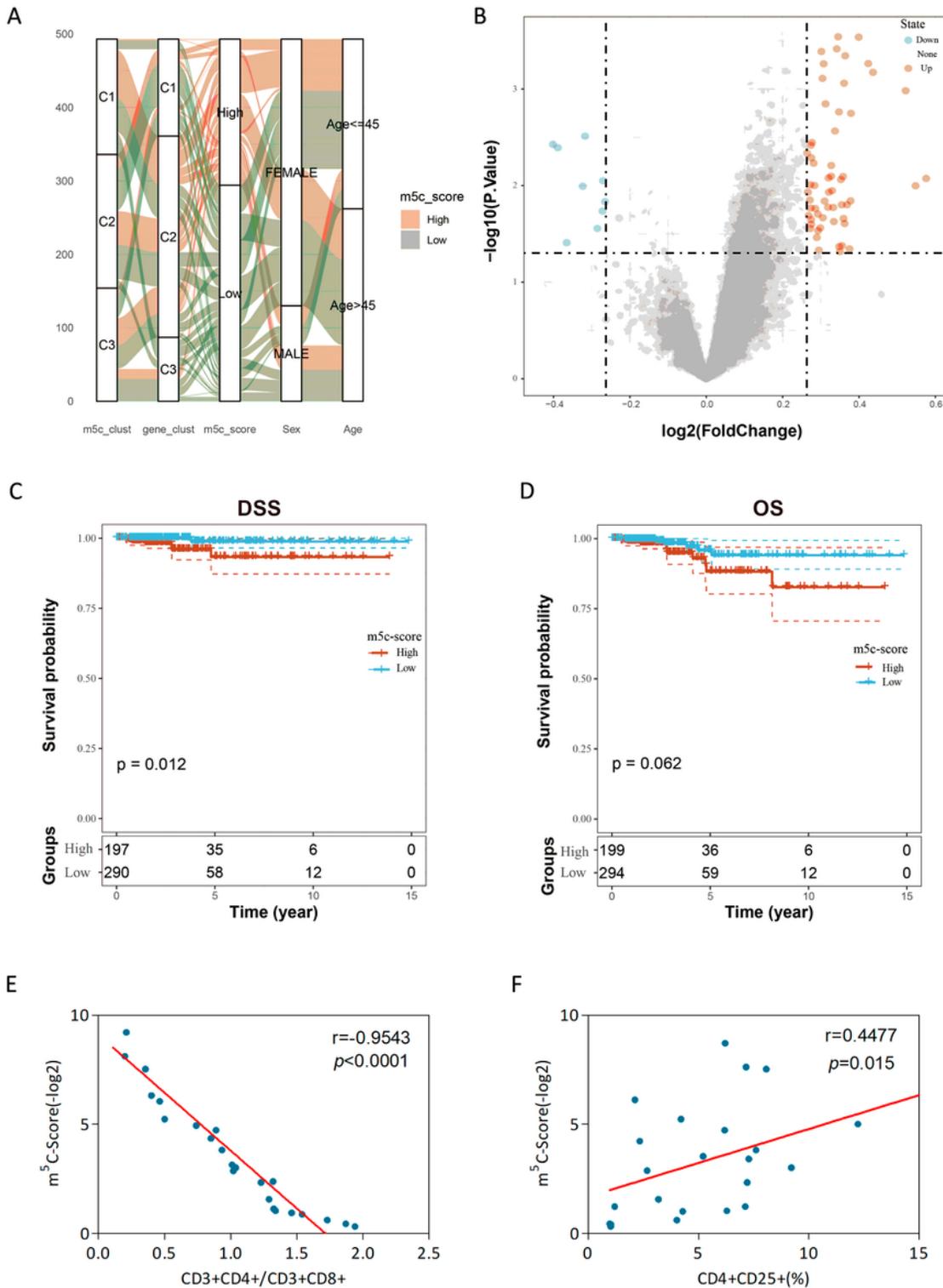
**Figure 3**

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



**Figure 4**

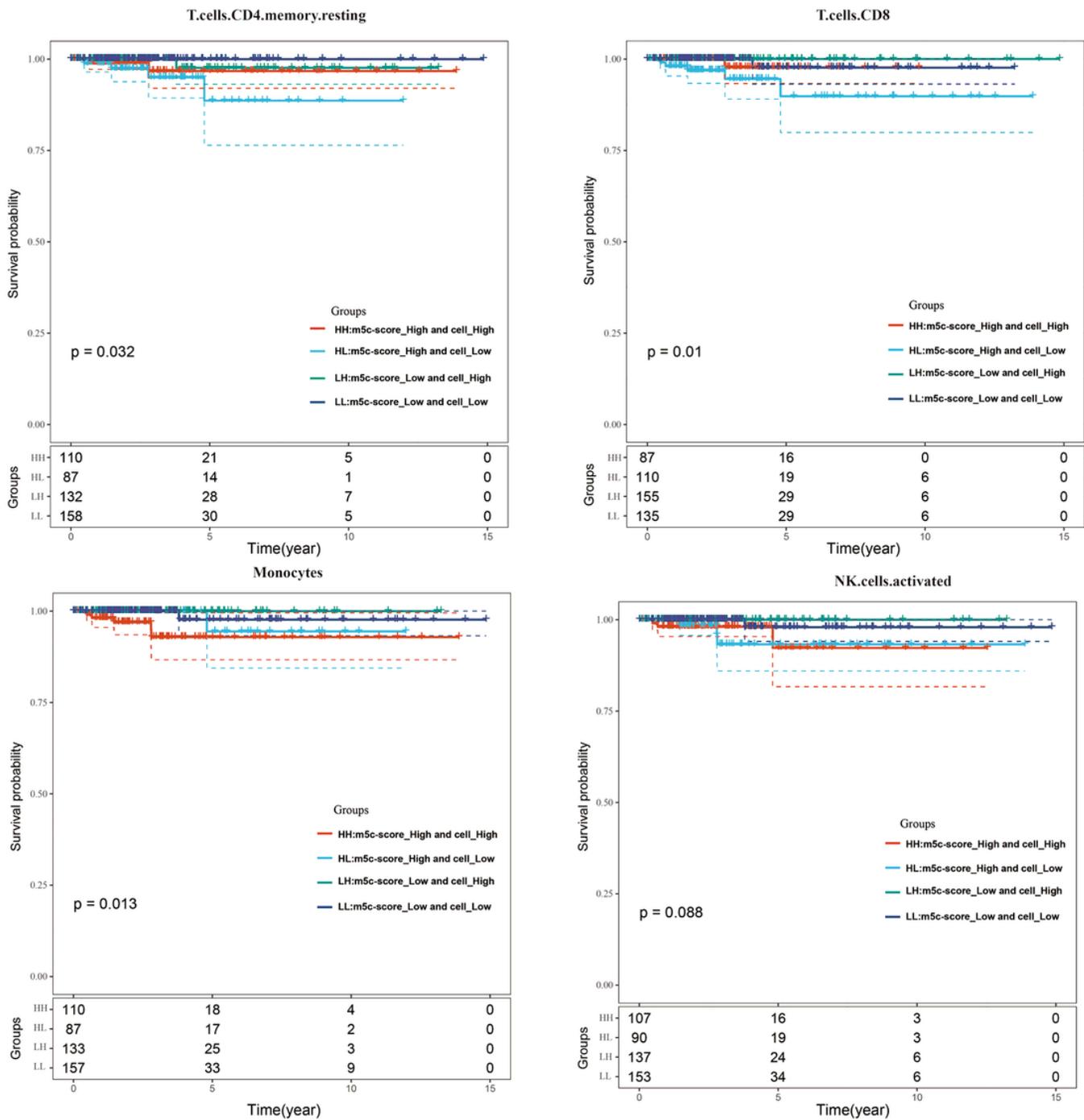
m5C gene signature establishment along with functional annotation. (A) KEGG enrichment analysis of 690 DEGs. (B) Levels of stromal scores in 3 m5C gene-cluster subtypes. (C) The levels of infiltration of 22 immune cells in 3 m5C gene-cluster subtypes.



**Figure 5**

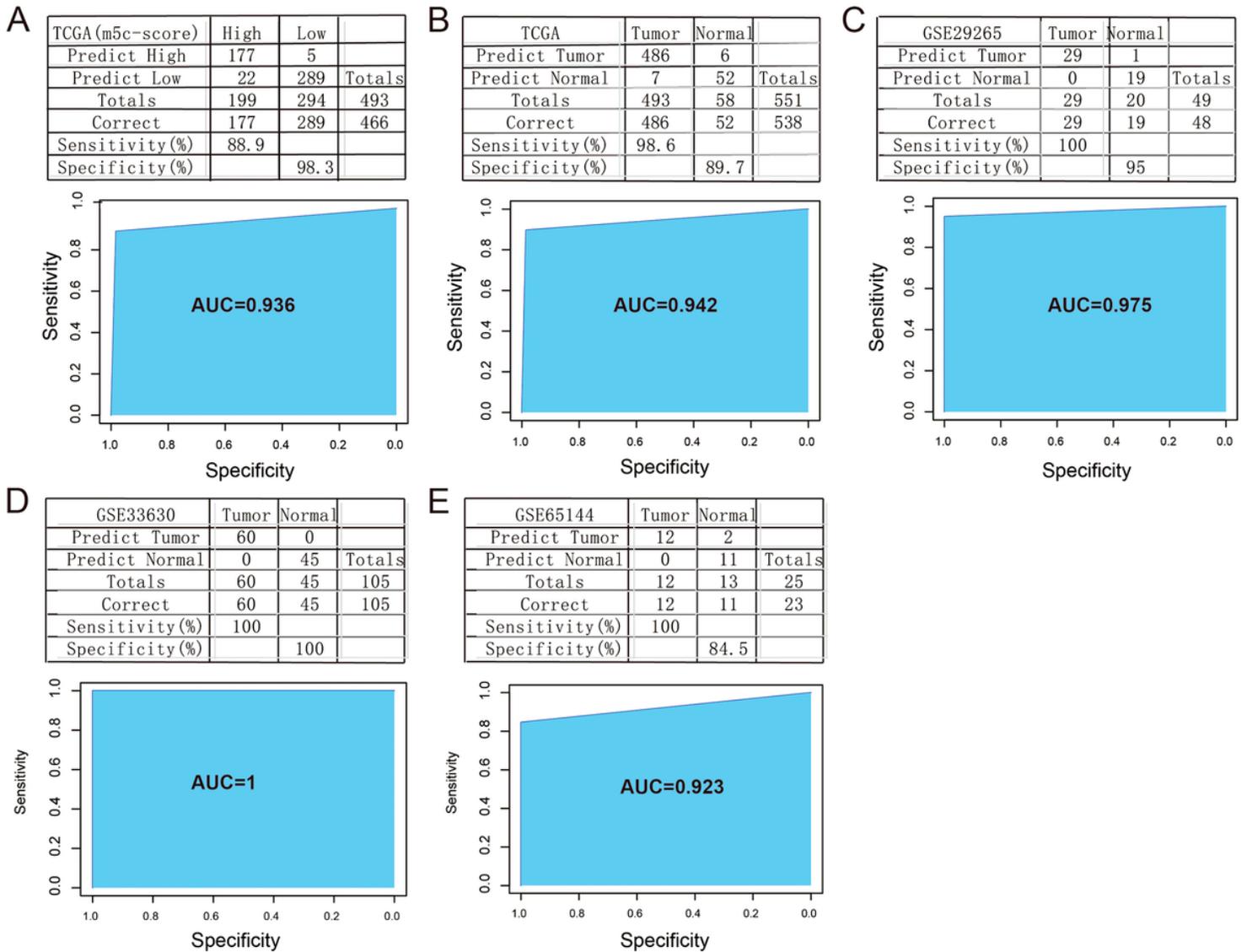
Clinical features and transcriptome traits of the m5C-associated phenotypes. (A) Alluvial diagram showing the changes of m5C modification patterns, gender, age, gene cluster and m5C-score. (B) DEGs between high and low m5C-score samples. (C, D) Differences in DFS (C) and OS (D) between high and low m5C-score 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.



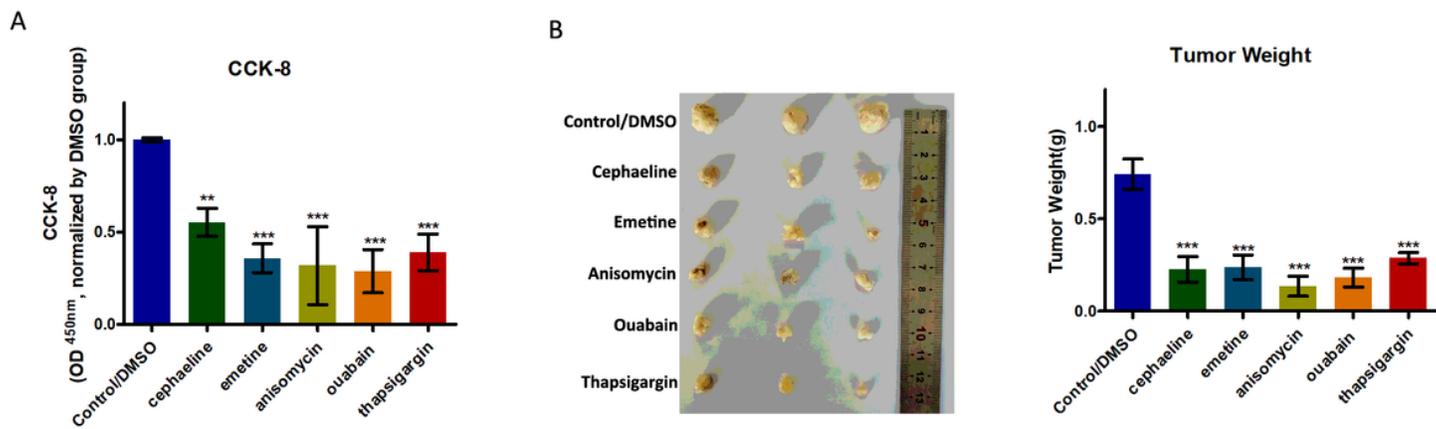
**Figure 6**

The influences of m5C-score and various immunocyte (resting CD4+ memory T cells, CD8+ T cells, activated NK cells, Monocytes) infiltration levels on the prognosis for PTC patients.



**Figure 7**

Construction and verification of the m5C-score-based PTC diagnostic model. (A) Comparison of classification results of TCGA-PTC samples by diagnostic model constructed based on 10 signature genes and m5C-Score. (B) Accuracy of classification of TCGA samples by diagnostic model constructed based on 10 signature genes. (C-E) Accuracy of classification of samples in GSE29265 (C), GSE33630 (D) and GSE65144 (E) by diagnostic model constructed based on 10 signature genes.



**Figure 8**

5 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 5 potential drugs (cephaeline, emetine, anisomycin, ouabain and thapsigargin) for 48 hours at 20  $\mu$ M. (B) BALB/c mice were subcutaneously injected with PTC cells. After five 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).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement12.pdf](#)
- [supplement13.xls](#)
- [supplement14.xls](#)
- [supplement15.xls](#)
- [supplement16.xls](#)
- [supplement17.xls](#)