

# Landscape Of M<sup>6</sup>a RNA Methylation Regulators and Tumor Microenvironment Cell-Infiltration Characterization In Gastroesophageal Adenocarcinomas

**Duanrui Liu**

Jinan Central Hospital Affiliated to Shandong University

**Jingyu Zhu**

Jinan Central Hospital Affiliated to Shandong University

**Zongming Wang**

Jinan Central Hospital Affiliated to Shandong University

**Yusong Fang**

Jinan Central Hospital Affiliated to Shandong University

**Mingjie Yuan**

Jinan Central Hospital Affiliated to Shandong University

**Xiaoli Ma**

Jinan Central Hospital Affiliated to Shandong University

**Yanfei Jia** (✉ [jiayanfei\\_@126.com](mailto:jiayanfei_@126.com))

jinan central hospital <https://orcid.org/0000-0002-0670-1604>

**Yunshan Wang**

Jinan Central Hospital Affiliated to Shandong University

---

## Research

**Keywords:** m6A, Tumor microenvironment, Gastroesophageal Adenocarcinomas, Prognosis, Immunotherapy

**Posted Date:** February 11th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-180985/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** RNA N6-methyladenosine (m<sup>6</sup>A) modification plays a nonnegligible role in shaping individual tumor microenvironment (TME) characterizations. However, the landscape and relationship of m<sup>6</sup>A modification and TME cell infiltration remain unknown in gastroesophageal adenocarcinomas (GEA).

**Methods:** We systematically examined the TME of GEA focusing on the distinct m<sup>6</sup>A modification patterns from the public databases. Intrinsic patterns of m<sup>6</sup>A modification were evaluated for associations with clinicopathological characteristics, underlying biological pathways, tumor immune cell infiltration, oncological outcomes and treatment responses. We generated a single-cell transcriptome atlas of the GEA sample inhouse to validate the role of the m<sup>6</sup>A modification pattern on TME.

**Results:** We identified and validated the landscape of m<sup>6</sup>A regulators and tumor-infiltrating immune cells in GEA. Then, two distinct m<sup>6</sup>A modification patterns of GEA (cluster1/2 subgroup) were defined, and we correlated two subgroups with TME cell-infiltrating characteristics. Cluster2 subgroup correlates with a poorer prognosis, down-regulated PD-1 expression, higher risk scores and distinct immune cell infiltration. Additionally, PPI and WGCNA network analysis were integrated to identify key module genes closely related to immune infiltration of GEA to find immunotherapy markers. And COL4A1 and COL5A2 in brown module were significantly correlated to the prognosis, PD-1/L1 and CTLA-4 expression of GEA patients. Interestingly, low COL5A2 expression was linked to an enhanced response to anti-PD-1 immunotherapy. Finally, a prognostic risk score was constructed using three m<sup>6</sup>A regulator-associated signatures that represented an independent prognosis factor for GEA. The single-cell transcriptome atlas of GEA sample validated the role of m<sup>6</sup>A modification pattern on TME and revealed that three m<sup>6</sup>A regulators are highly expressed in CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, Tregs and Macrophages.

**Conclusions:** Our work revealed m<sup>6</sup>A RNA methylation regulators are a type of vital participant in the malignant progression and TME diversity of GEA. m<sup>6</sup>A modification patterns of COL5A2 may be the potential biomarker contributes to predicting the response to anti-PD-1 immunotherapy.

## Background

Gastroesophageal adenocarcinomas (GEA) is still a major cause of cancer-related mortality worldwide [1]. Currently, the development of effective targeted therapeutics for GEA patients lags behind other cancers. Despite recent improvements in the multidisciplinary and multimodality treatment, in fact, the overall prognosis for GEA patients remains poor, with a global 5-year survival rate lower than 30% for gastric cancer (GC) and about 19% for esophagus adenocarcinoma [2]. For purpose of the high heterogeneity and complicated disease processes of GEA, there is still a lack of effective prognostic markers in this disease. Therefore, identifying molecular biomarkers and novel potential therapies is critical to predict GEA patients' prognosis and determine the personalized treatment.

Notably, N6-methyladenosine ( $m^6A$ ), the most abundant modification on mRNAs in eukaryotes, is closely related to stem cell differentiation, immune response, embryonic development, and microRNA (miRNA) editing; they also play an essential role in the progression of various cancers [3–7]. The  $m^6A$  methylation levels in tumors mainly depend on the expression of  $m^6A$  methylation regulators.  $m^6A$  is modulated by methyltransferase complex (“writers”), demethylases (“erasers”) and RNA-binding proteins (“readers”), which performs a series of biological functions [8]. The aberrant expression of  $m^6A$  regulators plays a vital regulatory role in tumor progression, prognosis, and radio-resistance. Yongsheng Li et al. [9] have showed the characteristics of  $m^6A$  RNA methylation across 33 types of cancer and speculated that the mechanism of  $m^6A$  RNA modification might be associated with the activation or depression of some oncogenic pathways such as PI3K-AKT-mTOR signaling, KRAS and P53 pathways. However, given the limited knowledge of the role of  $m^6A$  methylation in GEA, studying the precise correlation between  $m^6A$ -related regulator genes and its clinical prognosis is in high demand.

Immunotherapy represented by immunological checkpoint blockade (ICB, PD-1/L1 and CTLA-4) has demonstrated surprising clinical efficacy in a small number of patients with durable responses. In September 2017, the U.S. Food and Drug Administration (FDA) granted accelerated approval for pembrolizumab for the treatment of patients with recurrent, locally advanced or metastatic, gastric adenocarcinoma or GEA, whose tumors expressed PD-L1 and with disease progression on or after 2 or more systemic therapies. Disappointingly, response rates of immune checkpoint inhibitor monotherapy in GEA are around 10%-25% depending on the number of previous lines of chemotherapy and PD-L1 status [10]. Hence, it is important and necessary to understand the complexity of the tumor microenvironment (TME) and identify subclasses of tumor immune microenvironment existing in the patients' tumor to predict and administer corresponding immunotherapy. Notably, several studies have indicated the special relationship between TME infiltrating immune cells and  $m^6A$  modification. For instance, Dali et al. [11] showed that loss of YTHDF1 in classical dendritic cells enhanced the cross-presentation of tumor antigens and the cross-priming of CD8<sup>+</sup> T cells in vivo and YTHDF1 may be a potential therapeutic target in anticancer immunotherapy. Bo Zhang et al. [12] determined three distinct  $m^6A$  modification patterns in gastric cancer and found that the TME cell-infiltrating characteristics under these three patterns were highly consistent with the three immune phenotypes of tumors. Seungwon Yang et al. [13] suggested that  $m^6A$  demethylation by fat mass and obesity-associated protein (FTO) increases melanoma growth and decreases response to anti-PD-1 blockade immunotherapy. Na Li et al. [14] showed that Alkbh5 regulated the composition of tumor-infiltrating Treg and myeloid-derived suppressor cells and sensitized tumors to cancer immunotherapy. However, until now, the role of  $m^6A$  regulators in the malignancy and prognosis of GEA has not been comprehensively clarified. Therefore, research focusing on  $m^6A$  regulators is warranted to elucidate the potential regulatory mechanism of  $m^6A$  methylation in TME, which may reveal the potential mechanism and targets of immunotherapy.

In this study, we systematically evaluated the role of  $m^6A$  modification, and correlated the  $m^6A$  modification with the TME cell-infiltrating characteristics in GEA. Two GEA subtypes (cluster1/2) were

determined via the consensus clustering for m<sup>6</sup>A regulators that stratified the prognosis of patients, different TIICs and PD-1 expression. After WGCNA analysis, low COL5A2 expression was found to be linked to enhanced response to anti-PD-1 immunotherapy. Risk score developed from three m<sup>6</sup>A regulator-based signatures was an independent prognostic indicator of patients with GEA. The m<sup>6</sup>A regulator-based risk signatures were significantly related to the immune cell infiltration levels of patients with GEA. We employ the single-cell RNA sequencing (scRNA-seq) to validate m<sup>6</sup>A regulatory genes in distinct cell populations of TME in GEA. Therefore, this study sought to provide insights into the regulatory mechanisms associated with TME and the strategies for GEA immunotherapy.

## **Materials And Methods**

### **Data processing**

The mRNA (RNA sequencing) Fragments Per Kilobase of transcript per Million Fragments standardized expression data and corresponding clinicopathological features of TCGA-STAD&ESCA cohorts were retrieved for 159 GEA tissues and 39 adjacent non-tumor tissues from The Cancer Genome Atlas (TCGA, <http://cancergenome.nih.gov/>) and 121 GEA tissues from the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>). Patients without prognostic information were excluded from the analysis. The dataset of GSE96669 was obtained using the GPL10558 platform (Illumina HumanHT-12 V4.0 expression beadchip). We utilized the limma package to conduct the normalization process, deleting the normal or repeated samples for subsequent analysis. Then, the clinicopathological parameters for included samples also were download from the TCGA database. The relevant data TCGA and GEO provided is publicly available and open source; hence, approval by a local ethics committee was not required.

### **Evaluation of tumor-infiltrating immune cells (TIICs)**

CIBERSORT algorithm was applied to calculate the fractions of the 22 types of TIICs [15], which is considered better than previous deconvolution methods for the analysis of unknown mixture content and noise. We used this algorithm to statistically estimate the relative proportions of cell subpopulations from complex tissue expression profiles, making it a useful tool to estimate the abundances of special cells in mixed tissue. In this research, we used the R package “CIBERSORT” to estimate the fraction of immune cells of TCGA samples, which followed by quality filtering that tumor samples with  $P < 0.05$  were selected for the following analysis.

### **Generation of immune score, stromal score, and ESTIMATE score**

ESTIMATE algorithm was exploited to infer the fraction of immune and stromal cells in tumor tissues based on gene expression signature, including microarray expression

data sets, new microarray, as well as RNA-seq transcriptome profiles. The R script of the ESTIMATE algorithm was downloaded from the public source website (<https://sourceforge.net/projects/estimateproject/>). Then, we calculated the immune scores, stromal scores and ESTIMATE scores for each sample of the TCGA dataset, respectively. And the higher the respective score, the larger the ratio of the corresponding component in TME. After we got three scores from the ESTIMATE method, we could classify the samples into high- and low-level groups according to the median score, respectively.

## **Selection of m<sup>6</sup>A methylation regulators**

A total of 21 m<sup>6</sup>A methylation regulators were extracted from GSE96669 and TCGA database for identifying different m<sup>6</sup>A modification patterns mediated by m<sup>6</sup>A regulators in GEA. Although 21 regulators have been systematically analyzed in gastric cancer,[12] they have not been systematically analyzed in GEA. These 21 m<sup>6</sup>A regulators included 8 writers (CBLL1, METTL3, METTL14, KIAA1429, RBM15, RBM15B, WTAP, ZC3H13), 2 erasers (ALKBH5, FTO) and 11 readers (ELAVL1, FMR1, HNRNPA2B1, HNRNPC, IGF2BP1, LRPPRC, YTHDC1, YTHDC2, YTHDF1, YTHDF2, YTHDF3). Then, the correlation between the expression of these m<sup>6</sup>A RNA methylation regulators and different clinicopathological features were systematically evaluated.

## **Unsupervised clustering of m<sup>6</sup>A methylation regulators**

In order to further investigate the function of m<sup>6</sup>A RNA methylation regulators in GEA, we clustered the GEA patients into different groups by using the R package ConsensusClusterPlus (50 iterations, resample rate of 80%, and Pearson correlation, <http://www.bioconductor.org/>) based on the expression of the 21 m<sup>6</sup>A RNA methylation regulators [16]. The number of clusters and their stability were determined by the consensus clustering algorithm. Principal Components Analysis (PCA) was used with the R package for v3.6.3 to study the gene expression patterns in different GEA groups.

## **Differentially expressed genes (DEGs)**

We used R package “limma” with  $\log_2 | \text{fold-change (FC)} | > 1$  and adjusted P-value  $< 0.05$  to perform differentiation analysis of the gene expression, and DEGs were generated by the comparison between GEA samples vs. adjacent non-cancerous samples in TCGA and GSE96669 datasets. Venn online software (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) was used to identify the overlapping DEGs between tumor and normal samples.

## **Weighted gene co-expression network analysis (WGCNA) of DEGs**

WGCNA is a useful tool to establish the co-expression network between gene pattern and clinical traits using WGCNA package in R based on the RNA-seq data from TCGA database [17]. In the first step, we calculated a similarity matrix using bi-weight mid-correlation, as it is more robust to outliers. After that, a weighted adjacency matrix was defined by raising the co-expression similarity to appropriate soft-

thresholding power. The best power ( $\beta$ -value) was chosen based on the criterion of approximate scale-free topology. Next, we transformed the adjacency into a topological overlap matrix (TOM) and calculated the corresponding dissimilarity to minimize the effects of noise and spurious associations. Hierarchical clustering was used to produce a hierarchical clustering tree and dynamic tree cut method to assign co-expressed genes to each module. Modules were constructed with a minimum module size of 20 genes, and highly similar modules were combined using a dissimilarity threshold of 0.25.

## **Screening significant modules and functional enrichment analysis**

In order to identify the significance of each module, gene significance (GS) was calculated using linear regression by  $\log_{10}$  conversion of the p value between gene expression and clinical features. And module eigengenes (MEs) were defined as the first principal component of each gene module and adopted as the representative of all genes in each module. Next, we calculated the correlation between gene modules and clinical traits by WGCNA package in R and draw a heatmap. After obtaining these, genes in the gene modules from Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed to observe the function of selected significant gene modules using the clusterProfiler package in R. Enriched terms and pathways with adjusting P value  $< 0.05$  were selected.

## **Protein–protein interaction (PPI) network and Hub genes identification**

For clarifying the drivers of inducing carcinogenesis in a more reliable way, PPI analysis was performed necessarily. The Retrieval of Interacting Genes (STRING) database (<http://string-db.org>) online tool was used to evaluate interactive relationships and generate PPI networks among the DEGs in selected gene modules. Interaction score 0.7 served as cutoff value prior to visualization. Then Cytoscape software ([http://cytoscape.org/development\\_team.html](http://cytoscape.org/development_team.html)) was selected to visualize the results of PPI networks. Furthermore, CytoHubba app identifying hub objects from complex interaction in Cytoscape software was used to find top hub genes. Subsequently, top hub genes were selected and ranked by the Maximal Clique Centrality (MCC) method. Afterward, to select key genes that affect prognosis, survival data including living status and survival time was extracted from the TCGA database. Kaplan-Meier survival curves were built to screen for genes significantly associated with prognosis.

## **Construction of m<sup>6</sup>A related genes signature**

Univariate Cox regression analysis of the expression of 21 m<sup>6</sup>A RNA methylation regulators was conducted to determine the candidate genes associated with overall survival (OS). After that, regulators associated with OS in univariate analyses were subsequently selected for the least absolute shrinkage and selection operator (LASSO) Cox regression to construct a m<sup>6</sup>A-related risk signature for clinical prognosis [18]. Finally, three m<sup>6</sup>A RNA methylation regulators with their corresponding coefficients were determined by the minimum mean cross-validated error, choosing the optimal penalty parameter  $\lambda$  related

to the minimum 10-fold cross validation within the training set. The risk score of each patient with GEA in the TCGC cohort was calculated using the following formula:

$$\text{Risk score(RS)} = \sum_{i=1}^N (\text{Coef}_i \times X_i)$$

Where  $X_i$  is the standardized expression value of each selected m<sup>6</sup>A RNA methylation regulator, and  $\text{Coef}_i$  is the corresponding coefficient of the gene. All patients were divided into low- and high-risk groups based on the median value of the risk scores. Survival curves in the high-risk and low-risk groups were estimated using the Kaplan-

Meier method. Additionally, the receiver operating characteristic (ROC) curves and area under the ROC curves (AUC values) were applied to assess sensitivity and specificity. An AUC > 0.5 was considered as a significant diagnostic model.

#### Gene set enrichment analysis (GSEA)

GSEA is a computational method usually used to determine whether a set of basically defined gene sets exhibits statistically significant differences between two biological states. GSEA was provided by the JAVA program with MSigDB v7.1 and downloaded from the website of Broad Institute [19]. According to the median value of RS, the samples were divided into two groups, and "c2.cp.kegg.v7.1.symbols.gmt" gene set enrichment analysis was carried out, with p-value < 0.05 and q-value < 0.05 as indicative of statistical significance. The enrichment pathway was visualized using the R packages "ggplot2" and "clusterProfiler".

## Patients and sample information

We totally collected 16 non-neoplastic and neoplastic samples from GEA patients who underwent surgical treatments in the Gastrointestinal Surgery Department of Jinan Central Hospital Affiliated to Shandong University from 2018 to 2020. Fresh samples from two patients were immediately used for single-cell RNA-seq analysis, and the rest of fresh tumor and non-neoplastic tissues were frozen and stored at -80°C that used for PCR analysis. Clinical characteristics of included patients was shown in supplementary table S1. This research was approved by the Medical Ethics Committee of Jinan Central Hospital Affiliated to Shandong University and the sample acquisition and usage was performed in accordance with the approved guidelines. Informed consent was acquired from each involved patient.

## Quantitative real-time polymerase chain reaction (qRT-PCR)

For evaluating the expression levels of three signature regulators and hub genes, we extracted total RNA from clinical GEA samples by using RNA trizol reagent (CW BIO). According to the instructions of the manufacturer, cDNA synthesis was carried out by using reverse transcription kit (CW BIO). The qRT-PCR analysis was conducted on The LightCycler 480 Real-Time PCR System. Related gene expression levels

were calculated using the  $2^{-\Delta\Delta CT}$  method and the related GAPDH mRNA expression was used as an endogenous control. Primer sequences are presented in supplementary Table S2.

## Tissue dissociation and preparation of single-cell suspensions

Two tumors and One adjacent normal mucosa were processed immediately after being obtained from GEA patients. Fresh samples were then washed with phosphate-buffered saline (PBS, CORNING). Tissues were dissociated into single cells in dissociation solution (Tumor Dissociation Kit, human; Miltenyi Biotec) in 37 °C water bath with shaking for 20 min at 100 rpm. Digestion was terminated with 1× PBS containing 10% FBS(CORNING). Cell suspensions were filtered using a 70 μm filter and then centrifuged at 300g for 5 min at 4°C to pellet dead cells and red blood cells. The cells were washed twice and re-suspended in PBS with 0.04% bovine serum albumin (BSA, Sigma). The overall cell viability was confirmed by trypan blue exclusion and single cell suspensions were counted using a Countstar Rigel S2 Counter.

## Library preparation and sequencing

Single-cell suspensions were loaded to 10× Chromium to capture 18070 single cell according to the manufacturer's instructions of 10× Genomics Chromium Single-Cell 3' kit (V3). The cDNA amplification and library construction steps were conducted via following the standard protocol. Libraries were sequenced on an Illumina NovaSeq 6000 sequencing system (paired-end multiplexing run,150bp) by LC-Bio Technology co.ltd (HangZhou,China) at a minimum depth of 20000 reads per cell.

### Single-cell gene expression quantification, quality control, and cell type determination

The sequencing data from 10x Genomics were aligned and quantified using the CellRanger software package (version 3.1) against the human reference genome (GRCh38.p13). Raw gene expression matrices were imported and processed using the Seurat R package (version 3.1.1). Standard scRNA-seq filtering excludes low-quality cells with > 10 ~ 25% mitochondrial UMIs, we filtered low-quality cells by two measurements: 1) less than 500 expressed genes.2) over 25% UMIs derived from the mitochondrial genome. Gene expression matrices of the remaining 18070 cells were normalized to the total cellular UMI count. using the LogNormalize method of the "Normalization" function of the Seurat software to calculated the expression value of genes; PCA (Principal component analysis) analysis was performed using the normalized expression value. Within all the PCs, the top 10 PCs were used to do clustering and t-SNE analysis; To find clusters, selecting weighted Shared Nearest Neighbor (SNN) graph-based clustering method. Marker genes for each cluster were identified with the Wilcoxon rank-sum test via the FindAllMarkers function in Seurat. This selects markers genes which are expressed in more than 10% of the cells in a cluster and average log (Fold Change) of greater than 0.26.

## Statistical analysis

Data were analyzed using R software (version 3.6.3) and GraphPad Prism (version 6). The Wilcoxon's test was used to compare the expression of m<sup>6</sup>A RNA methylation regulators between cancer and normal tissues. Spearman correlation analysis was performed using "corrplot" package in R. The distributions of age, sex, histological grade and TNM stage between clusters and between risk subgroups were analyzed using Chi-square test. Wilcoxon rank sum or Kruskal–Wallis rank sum test as the significance test depending on the number of clinicopathological features and immunotherapy response for comparison. Survival curves were plotted by using the "survival" package in R. The ROC analysis was performed for the evaluation of the AUC value in the follow-up period with the "survivalROC" package. Log-rank test was used to assess statistical significance. All statistical results with  $p < 0.05$  were regarded to be statistically significant.

## Results

# The Landscape of m<sup>6</sup>A Methylation Regulators and TIICs in GEA

To explore the important biological functions of each m<sup>6</sup>A RNA methylation regulator in tumorigenesis and development, we first compare 21 m<sup>6</sup>A methylation regulators expression in tumor and normal samples. And the result strongly indicates that the expression of most m<sup>6</sup>A RNA methylation regulators was related to the occurrence of GEA (Fig. 1A-D). We speculate that the change of m<sup>6</sup>A RNA methylation regulators may be an intrinsic trait that can characterize individual differences. Correlation analysis was also employed to investigate the relationship between the expression level of m<sup>6</sup>A RNA methylation regulators of GEA. The relationship between the 21 m<sup>6</sup>A RNA methylation regulators is positively correlated, including YTHDF3 and KIAA1429 are most relevant in the TCGA dataset and the strongest positive correlation consisted between METTL3 and RBM15B in GSE96669 (Fig. 1E, F). The analyses presented above suggested that the high heterogeneity of expressional alteration landscape in m<sup>6</sup>A regulators between normal and GEA samples, indicating that the expression imbalance of m<sup>6</sup>A regulators played a crucial role in the GEA occurrence and progression.

Next, the difference between GEA tissues and adjacent tissues in 22 immune cells was analyzed by using the CIBERSORT algorithm in TCGA. Obviously, the proportion of immune cells in GEA tumor tissues was significantly different from that in normal tissues (Supplementary Figure S1A, B). We speculate that the change in the correlation of immune cells may be an internal characteristic that can reflect the external differences. Next, we investigated mutual relationship between 22 immune cells in GEA samples, and the results showed that most of the relationships between the immune cells were a negative correlation, and the Macrophages M2 and the B cells naive are most negative relevant (supplementary Figure S1C). Meanwhile, the positive correlation between NK cells resting and T cells CD4 memory activated is the most significant (supplementary Figure S1C).

# Correlation of TME Components with Clinicopathological Characteristics and m<sup>6</sup>A Methylation Regulators

For determining the relationship between the proportion of immune and stromal components in TME with the clinicopathological characteristics, we analyzed the corresponding clinical information of GEA cases from the TCGA database. Stromal score was positively correlated to TMN-T stages ( $P = 0.009$ ), tumor grades ( $P < 0.01$ ) and tumor stages ( $P = 0.037$ ) (supplementary Figure S2A); immune scores were associated with advanced tumor grades ( $P = 0.021$ ), higher TNM-N level ( $P = 0.042$ ), and females have higher immune scores than males ( $P = 0.045$ ) (supplementary Figure S2B); ESTIMATE score showed the positive correlation with N, T classification of TMN stages, tumor grade and sex ( $P < 0.05$ ) (supplementary Figure S2C). Therefore, these results indicated that the ratio of immune and stromal components was related to the progress of GEA, such as invasion and metastasis. Next, to explore the correlation between the high/low ratio of immune and stromal components in TME with m<sup>6</sup>A regulators, we found that most m<sup>6</sup>A regulators were highly expressed in samples with low immune and stromal scores, which indicate a special connection between TME components and m<sup>6</sup>A regulators (supplementary Figure S2D, E).

## Consensus Clustering for m<sup>6</sup>A RNA Methylation Regulators Correlated with Distinct Survival, Immune Cell Infiltration

As GEA patients have a very poor prognosis, we tried to investigate GEA subtypes based on the expression of m<sup>6</sup>A RNA methylation regulators to explore its possible pathogenesis. We removed 39 normal samples and grouped 159 cancer tissues using the ConsensusClusterPlus package. According to the expression similarity of m<sup>6</sup>A RNA methylation regulators,  $k = 2$  was the most optimum with clustering stability datasets increasing from  $k = 2-9$  (Fig. 2A-C, supplementary Figure S3). Hence, GEA samples from the TCGA dataset were pre-classified into two groups (100 samples in one group labeled as cluster1 and 59 samples in another group labeled as cluster2) through consensus cluster analysis. PCA was performed to elucidate the difference in transcriptional profiles between the Cluster1 and Cluster2 subgroups. And the results showed that there was a clear distinction between these two subgroups and the results of our classification by m<sup>6</sup>A RNA methylation regulators are correct (Fig. 2D). Kaplan–Meier survival analysis for the clustered samples revealed a noticeable decrease in the OS of cluster2 compared with cluster1, suggesting that the 21 methylation regulators could classify the GEA samples at prognostic level (Fig. 2E). Moreover, we discovered that cluster2 have lower PD-1 expression and most TIICs fraction was significantly higher in cluster 1, such as T cells CD8, T cells follicular helper, Monocytes, Macrophages M0 and Mast cells resting ( $P < 0.05$ , Fig. 2F, G). The clustering results revealed that m<sup>6</sup>A modification patterns were strongly affect TME infiltrating immune cells fraction and immunotherapy, which further confirmed our hypothesis.

## Differentially Expressed Genes (DEGs) Screening and WGCNA Analysis

For exploring potential biomarkers associated with genome, m<sup>6</sup>A modification and TME cell-infiltration, we firstly identified 1341 DEGs in the GSE96669 dataset (supplementary table S3) and 6360 DEGs in the TCGA dataset (supplementary table S4) between GEA samples and adjacent normal samples (Fig. 3A, B). The overlapping 492 DEGs were used to further construct co-expression network (Fig. 3C, supplementary table S5). Subsequently, WGCNA was performed to construct a gene co-expression network and correlate gene modules with m<sup>6</sup>A cluster, immune scores and stromal scores, which were selected as trait data of WGCNA. In the case of a scale-free network and topological overlap, a hierarchical clustering tree based on dynamic hybrid cutting is established after the outlier samples are eliminated (Fig. 3D, supplementary Figure S4A). To ensure a scale-free network, we selected  $\beta = 4$  (scale-free  $R^2 = 0.90$ ) as a soft threshold (supplementary Figure S4B). Finally, we identified six gene modules correlated to m<sup>6</sup>A cluster, immune scores and stromal scores, respectively.

The Module trait relationships were estimated by the correlation between modules and phenotypes, which made it easier to identify highly correlated modules and phenotypes. Figure 3E showed that the brown module had significant relevant to immune scores (cor = 0.65, P = 2e-19). Scatter diagrams of gene significance was shown in Fig. 3F, which could more directly observe the significance of genes in brown modules related to immune scores. For the significant modules and key genes, GO and KEGG pathway enrichment analyses were performed. And GO analysis showed that the genes in the brown module are mainly enriched in extracellular matrix organization, extracellular structure organization and so forth (Fig. 4A). KEGG pathway enrichment analysis indicated that the genes in the brown module were mainly associated with protein digestion and absorption, PI3K-Akt signaling pathway and so forth (Fig. 4B). Based on the above analysis, we found that most enrichment pathways are associated with tumor progression.

## Hub Genes Identification and Its Role in Immunotherapy

The PPI network among genes in the brown module (40 nodes and 52 edges) was established by using the STRING database. Based on the MCC scores, the top ten highest-scored genes in the brown module were selected as hub genes for further analysis (Fig. 3G). Furthermore, two hub genes (COL4A1, COL5A2) in the brown module were significantly negatively related to the prognosis of patients with GEA (P < 0.05; Fig. 5A, B). Consistent with the above bioinformatic results, PCR analysis also revealed that COL4A1 and COL5A2 were significantly high expressed in tumor (supplementary Figure S5). Immunotherapies represented by PD-1/L1 and CTLA-4 blockade has undoubtedly emerged a major breakthrough in cancer therapy. Obviously, COL4A1 was significantly related to PD-L1 and CTLA-4 expression (Fig. 5C, D). And COL5A2 had significantly correlation with PD-1/L1 and CTLA-4 expression (Fig. 5E-G). The above results suggested that COL4A1 and COL5A2 may be potential biomarkers for predicting the effect of immunotherapy. Next, in the anti-PD-1 cohort (GSE78220), the significant clinical response to anti-PD-1 immunotherapy in patients with low COL5A2 expression compared to those with high COL5A2 expression was confirmed (Fig. 5H). In summary, our work indicated that m<sup>6</sup>A methylation modification patterns

were significantly correlated with tumor immune phenotypes and response to immunotherapy, and COL5A2 may be a potential biomarker contribute to predicting the response to anti-PD-1 immunotherapy.

## Construction and Validation of Prognostic Signatures for m<sup>6</sup>A RNA Methylation Regulators

For the purpose of investigating the prognostic value of the 21 m<sup>6</sup>A regulators in GEA, univariate Cox regression analysis was performed based on the expression levels of the regulators from TCGA (Fig. 6A). The results indicated that 3 out of 21 tested genes including KIAA1429, HNRNPA2B1 and FMR1 are significantly correlated with OS ( $P < 0.05$ ) and three genes are risky genes with HR  $> 1$ . For ensuring the reliability of results, PCR results also suggested that three regulators were significantly high expressed in tumor samples (supplementary Figure S5). To identify the most powerful prognostic m<sup>6</sup>A RNA methylation regulators, the least absolute shrinkage and selection operator (LASSO) Cox regression analysis to the 3 prognosis-related genes was conducted (Fig. 6B, C) and the coefficient of each independent prognostic gene was shown in supplementary table S6. The LASSO results showed that three regulators (KIAA1429, HNRNPA2B1 and FMR1) were the powerful prognostic factors and a risk signature was constructed. Next, the risk score was estimated according to the coefficients obtained from the LASSO analysis. To explore the prognostic role of the three-gene risk signature, we separated the GEA patients in the TCGA dataset into low and high-risk groups based on the median risk score, the results indicated that the high-risk group has a worse survival in patients with GEA (Fig. 6D). To evaluate the predictive accuracy of the risk score model, we performed the ROC curve and calculated the AUC. The AUC was 0.943 in the 5-year ROC curve for the prognostic model (Fig. 6E). Then, the univariate and multivariate Cox regression analyses results suggested the risk signature is an independent prognostic indicator (Fig. 6F, G). In addition, three m<sup>6</sup>A regulators were significantly higher expressed in cluster2 (Fig. 6H, I). This result was consistent with the fact that cluster 2 had a higher risk score, lower immune cells infiltration and poorer prognosis than cluster1.

Next, we evaluated the relative abundance of 22 TIICs for each patient within two risk groups using CIBERSORT. It was revealed that T cell follicular helper was significantly higher in the high-risk group compared with the low-risk group; however, Monocytes and T cells CD8 were significantly higher in the low-risk group compared with the high-risk group (Fig. 6J). Moreover, to screen for the possible signaling pathways and mechanisms that are significantly altered within high- and low-risk groups, GSEA analysis was performed with data from the TCGA cohort. As shown in Fig. 6K, RNA modification related pathway and cancer associated pathway were more enrichment in high risk group. The pathways enrichment analysis provided evidence of molecular mechanisms affected by the risk signature.

## Correlating Public Datasets to scRNA-seq

To investigate the cell populations and the associated molecular characteristics during the occurrence and development of GEA, we took 2 cases of GEA and 1 case of adjacent normal tissue for single-cell sequencing analysis. After quality control, 18070 single cells were clustered into 13 major clusters.

Cluster-specific genes were used to annotate cell types with classic markers described in the supplementary table S7. Consistent with previous public dataset analysis, CD8<sup>+</sup> T cells were enriched in normal tissue, while CD4<sup>+</sup> T cells, Macrophages and Treg were more concentrated in tumor tissue (Fig. 7A). In addition, it seems that the proportion of CD8<sup>+</sup> T cells and CD4<sup>+</sup> T cells exhibited significant changes, indicating that CD8<sup>+</sup> T cells may differentiate into CD4<sup>+</sup> T cells. To further investigate this ongoing process, we performed trajectory analysis on CD8<sup>+</sup> T cells and CD4<sup>+</sup> T cells. Similar phenomenon was observed in the computational pipeline (Fig. 7B). Meanwhile, we noticed that CD69, T cell activation marker, was downregulated in this differentiation process, while CXCL8 and CXCL13 were upregulated (Fig. 7C). Previous work had reported that CXCL13/CXCR5 signaling axis makes pivotal contributions to the development and progression of human cancers [20]. CXCL8 is mainly secreted by macrophages and contributes to the immunosuppressive microenvironment by inducing PD-L1<sup>+</sup> macrophages in gastric cancer [21]. These results indicated that the immunosuppressive microenvironment is gradually forming in this differentiation process. Then, we found that three regulators (KIAA1429, HNRNPA2B1 and FMR1) were higher expressed in tumor than normal tissue, which verified the reliability of the previous analysis results again (Fig. 7D). Strikingly, three m<sup>6</sup>A regulators originated mostly from T cells such as CD4<sup>+</sup> T cells and Tregs in GEA (Fig. 7E), which suggested that m<sup>6</sup>A modification may participate in sustaining immunosuppressive microenvironment via control T cell differentiation. COL5A2, closely related to the expression of KIAA1429 and FMR1, also proved to be highly expressed in tumor tissues in single-cell analysis results (Fig. 7F, G).

## Discussion

In this present research, we attempted to demonstrate the expression patterns, prognostic values, and effects on the TME of m<sup>6</sup>A regulators in GEA. Differential expression analysis found that the majority of m<sup>6</sup>A RNA methylation regulators were significantly differently expressed between adjacent normal and tumor samples. This should preliminarily suggest that those m<sup>6</sup>A regulators have a distinct role of influencing GEA initiation and could be applied for the subsequent analysis. Compare to normal tissues, GEA is locally infiltrated with higher immune cell subgroups, including B cells naive, B cells memory, Plasma cells, T cells gamma delta, Macrophages M0, Macrophages M1, Dendritic cells resting, Dendritic cells active. Meanwhile, ESTIMATE algorithm-derived immune scores, stromal scores and ESTIMATE scores were applied to facilitate the quantification of the non-tumor components in malignancy [22]. Stromal, immune and ESTIMATE scores for tumor tissue were found to be significantly associated with the clinicopathologic features of the tumor such as age, differentiation grade and TNM stage. Importantly, most m<sup>6</sup>A regulators' expression was significantly associated with immune/stromal scores. Furthermore, we constructed a comprehensive single-cell transcriptome atlas of 18070 cells from 2 samples of GEA and 1 sample of adjacent normal mucosa and revealed m<sup>6</sup>A regulators exert an important role in maintaining aberrant immune infiltration. Therefore, these results demonstrated that aberrant immune infiltration and m<sup>6</sup>A regulators expression in GEA as a tightly regulated process might play important roles in tumor development and this process had clinical importance.

We characterized the effects of distinct m<sup>6</sup>A methylation modification on different GEA subtypes by clustering m<sup>6</sup>A regulators. The two subtypes showed significant differences in patient prognosis, PD-1 expression, immune cell infiltration and RS. This suggests that two subtypes' differences are essential and reflect the heterogeneity of the immune microenvironment of GEA, which is worthy of further study. In order to investigate the expression characteristics of m<sup>6</sup>A methylation regulators in tumor, many studies clustered the tumor samples into different groups using consensus clustering analysis. For instance, Jing Chen et al. [23] identified two clusters of clear cell renal carcinoma with significant differences in OS and tumor stage between them based on the expression pattern of m<sup>6</sup>A RNA methylation regulators by means of consensus clustering. Similarly, Lilan Yi et al. [24] showed that two molecular subtypes were identified by consensus clustering for 15 m<sup>6</sup>A regulators, and two subtypes were distinct in prognosis, PD-L1 expression, immunoscore, and immune cell infiltration. However, so far, the expression of m<sup>6</sup>A regulators has remained elusive for typing research by consensus clustering analysis in GEA. Our research had once again identified the special relationship between m<sup>6</sup>A modification and tumor immune cell infiltration, which has important clinical value.

To elucidate the potential association between the genome, m<sup>6</sup>A modification pattern and TME in GEA, we systematically clustered the co-expressed genes by WGCNA analysis. This approach allowed us to identify gene modules most related to cancer immunological phenotypes. COL4A1 and COL5A2, the two hub genes in the module, in the collagen family can encode pro-alpha collagen chains that are assembled into collagen. They have a relatively high expression in tumor cells and related to the prognosis and clinicopathological factors of patients. Désert et al. [25] reported that elevated expression of COL4A1 was significantly correlated with tumor stage and worse overall survival in patients with hepatocellular carcinoma. Zhang et al. [26] demonstrated the abnormally high expression of COL4A1 in GC, and high expression of COL4A1 was closely correlated with primary tumor size, lymph node metastasis and distant metastasis, with the silencing of COL4A1 significantly inhibiting cell proliferation of GC cells in vitro. Meanwhile, elevated COL4A1 gene expression has been found to be associated with trastuzumab resistance in GC [27]. Several studies had reported that COL5A2 might function as a crucial role in the initiation and progression of tumor by using bioinformatics technologies [28, 29]. More importantly, COL5A2, was correlated with stromal scores in GC, promoted the recruitment of circulating monocytes into the TME and facilitated their differentiation into tumor-associated macrophages [30]. Similarly, in our research, we found that COL4A1 and COL5A2 were significantly related to prognosis of GEA patients and regulate TME infiltration characteristics. Intriguingly, the expression of COL5A2 and COL4A1 was significantly correlated with ICB (PD-1/L1 and CTLA-4) expression. Notably, COL5A2 expression was also linked to enhanced response to anti-PD-1 immunotherapy. Above results suggest that distinct m<sup>6</sup>A modification patterns of COL5A2 may establish totally different TME, which influence immunotherapy response.

Whether m<sup>6</sup>A RNA methylation regulators have prognostic value in cancer is of great significance [31]. We performed univariate and LASSO Cox regression analyses to construct a prognostic related risk signature with three m<sup>6</sup>A RNA methylation regulators, including KIAA1429, HNRNPA2B1 and FMR1, which divided

the GEA patients into low- and high-risk groups. In the m<sup>6</sup>A methyltransferase complex, KIAA1429 acts as a scaffold in bridging the catalytic core components of methyltransferase complex and RNA substrates, which affect the installation of m<sup>6</sup>A at specific locations [32]. Ran Miao et al. [33] found that KIAA1429 could serve as an oncogene in gastric cancer by stabilizing c-Jun mRNA in an m<sup>6</sup>A-independent manner. HNRNPA2B1 is a nuclear reader of the m<sup>6</sup>A mark and has important effects on primary microRNA processing and alternative splicing. Barceló et al. [34] reported that HNRNPA2B1 acts as a regulator of KRAS-dependent tumorigenesis through the critical pancreatic ductal adenocarcinoma cells signaling pathway PI3K/AKT. FMR1 gene and consequently lack of synthesis of FMR protein (FMRP) is associated with fragile X syndrome, and FMRP plays a critical role in chromatin dynamics, RNA binding, mRNA transport, and mRNA translation [35, 36]. Jianhao Li et al. [37] indicated that high expression of KIAA1429 and HNRNPA2B1 were significantly associated with poor prognosis in osteosarcoma, and m<sup>6</sup>A regulators might be involved in osteosarcoma progression through humoral immune response. Francesca Zalfa et al. [38] reported that there was an association between FMRP levels and the invasive phenotype in melanoma. In accordance with previous results, we found the three-gene risk signature, KIAA1429, HNRNPA2B1 and FMR1, showed a good performance for predicting GEA patients' prognosis and immune cells infiltration characteristics. Importantly, our further study firstly revealed that the three m<sup>6</sup>A regulators' expression in T cells and macrophages may play a crucial role in tumor progression. Moreover, COL5A2 expression was significantly related to KIAA1429 and FMR1 expression. Therefore, we speculate that KIAA1429 and FMR1 regulate the number of macrophages and T cells in TME by affecting the m<sup>6</sup>A modification of COL5A2, which in turn affects the clinicopathological characteristics of patients.

The tumor microenvironment plays an essential regulatory role in tumorigenesis, and its heterogeneity can lead to multiple dimensions, including patient prognosis and therapeutic response [39–41]. Here, we generated a single-cell transcriptome atlas of GEA sample to validate the role of m<sup>6</sup>A modification pattern on TME. 13 different cell types were identified in the GEA microenvironment. Notably, CD8<sup>+</sup> T cells mostly originated from normal mucosal tissues, while Macrophages, CD4<sup>+</sup> T cells and Treg cells were enriched in GEA tissues. Therefore, we indicated that the downregulated immunogenicity of cancer cells potentially contributes to the formation of an immunosuppressive microenvironment. Hanjie Li et al. [42] reported that a large population of CD8<sup>+</sup> T cells showing continuous progression from an early effector "transitional" into a dysfunctional T cell state and the intensity of the dysfunctional signature is related to tumor reactivity. Moreover, CD4<sup>+</sup>T cell help in the form of IL-21 may potentially be harnessed to bolster the formation of protective cytolytic CX3CR1<sup>+</sup>CD8<sup>+</sup>T cells and improve control over tumor progression [43]. Our monocle analysis verified again that CD4<sup>+</sup> T cells regulate CD8<sup>+</sup> T cell differentiation in tumor progression of GEA. m<sup>6</sup>A RNA modification controls the differentiation of naive T cells and sustains the suppressive functions of Tregs [5, 44]. Then, our work revealed that three m<sup>6</sup>A regulators are highly expressed in CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, Tregs and Macrophages, which was consistent with previous works. In short, we first discovered that KIAA1429, HNRNPA2B1 and FMR1 regulate T cell differentiation in the GEA microenvironment, which may provide new targets to optimize immunotherapy.

## Conclusions

In summary, this study systematically evaluated prognostic value, role in TME, correlation of immunotherapy, and potential regulatory mechanisms of m<sup>6</sup>A RNA regulators in GEA. A notable result was the identification of three m<sup>6</sup>A RNA methylation regulatory genes risk signature, KIAA1429, HNRNPA2B1 and FMR1, as the essential determinant of GEA patients' prognosis and immune cells infiltration characteristics. The distinct m<sup>6</sup>A modification patterns of COL5A2 were found to be linked to enhanced response to anti-PD-1 immunotherapy. The m<sup>6</sup>A RNA methylation may be involved in regulation of TME via multiple RNA and cancer-related processes such as PI3K/AKT signaling pathway. Our work highlights the importance of CD8<sup>+</sup> T cells differentiation in antitumor immunity and ICB-based immunotherapy for GEA at single-cell level. The information from this study makes important contributions to our understanding of m<sup>6</sup>A RNA regulators and TME of GEA and may help the development of a new generation of immune therapeutics and precision treatment in GEA.

## Abbreviations

m<sup>6</sup>A: RNA N<sup>6</sup>-methyladenosine; TME: Tumor Microenvironment; GEA: Gastroesophageal Adenocarcinomas; TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; DEGs: Differentially Expressed Genes; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; WGCNA: Weighted Gene Co-expression Network Analysis; PPI: Protein–Protein Interaction; TIICs: Tumor-Infiltrating Immune Cells; scRNA-seq: single-cell RNA sequencing.

## Declarations

Ethics approval and consent to participate All experimental protocols were approved by the Ethics Committee of the Jinan Central Hospital (Jinan, China) and performed in accordance with the relevant guidelines and regulations. Written informed consent was obtained from all patients. Consent for publication Not applicable. Availability of data and material The datasets analysed during the current study are available in the GEO repository (<https://www.ncbi.nlm.nih.gov/geo/>) and TCGA repository (<https://portal.gdc.cancer.gov/>). The datasets used and analyzed during the current study are available from the corresponding author on reasonable request. Competing interests No potential conflicts of interest were disclosed. Funding This study was financially supported by the National Natural Science Foundation of China (Nos. 31671468 and 31970728), the Academic Promotion Programm of Shandong First Medical University (No. 2019QL024), the Shandong Provincial Natural Science Foundation of China (Nos. ZR2016HM15, ZR2020LZL005, ZR2020MH050). Authors' Contributions DL contributed to conceptualization, resources, data curation, validation, methodology, writing- original draft. JZ contributed to investigation, writing-review and editing. ZW contributed to data curation, formal analysis, methodology. YF contributed to resources, formal analysis, methodology. MY contributed to software, formal analysis. XM contributed to data curation, visualization. YJ contributed to conceptualization,

supervision, writing-review and editing, funding acquisition and validation. YW contributed to conceptualization, supervision, funding acquisition, validation, visualization.

## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394–424.
2. Cronin KA, Ries LA, Edwards BK. The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute. *Cancer.* 2014;120(Suppl 23):3755–7.
3. Deng X, Su R, Weng H, Huang H, Li Z, Chen J. RNA N(6)-methyladenosine modification in cancers: current status and perspectives. *Cell Res.* 2018;28(5):507–17.
4. Huang H, Weng H, Chen J. m(6)A Modification in Coding and Non-coding RNAs: Roles and Therapeutic Implications in Cancer. *Cancer Cell.* 2020;37(3):270–88.
5. Li HB, Tong J, Zhu S, Batista PJ, Duffy EE, Zhao J, et al. m(6)A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. *Nature.* 2017;548(7667):338–42.
6. Chen T, Hao YJ, Zhang Y, Li MM, Wang M, Han W, et al. m(6)A RNA methylation is regulated by microRNAs and promotes reprogramming to pluripotency. *Cell Stem Cell.* 2015;16(3):289–301.
7. Geula S, Moshitch-Moshkovitz S, Dominissini D, Mansour AA, Kol N, Salmon-Divon M, et al. Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science.* 2015;347(6225):1002–6.
8. Yang Y, Hsu PJ, Chen YS, Yang YG. Dynamic transcriptomic m(6)A decoration: writers, erasers, readers and functions in RNA metabolism. *Cell Res.* 2018;28(6):616–24.
9. Li Y, Xiao J, Bai J, Tian Y, Qu Y, Chen X, et al. Molecular characterization and clinical relevance of m(6)A regulators across 33 cancer types. *Mol Cancer.* 2019;18(1):137.
10. Shitara K, Özgüroğlu M, Bang YJ, Di Bartolomeo M, Mandalà M, Ryu MH, et al. Pembrolizumab versus paclitaxel for previously treated, advanced gastric or gastro-oesophageal junction cancer (KEYNOTE-061): a randomised, open-label, controlled, phase 3 trial. *Lancet.* 2018;392(10142):123–33.
11. Han D, Liu J, Chen C, Dong L, Liu Y, Chang R, et al. Anti-tumour immunity controlled through mRNA m(6)A methylation and YTHDF1 in dendritic cells. *Nature.* 2019;566(7743):270–4.
12. Zhang B, Wu Q, Li B, Wang D, Wang L, Zhou YL. m(6)A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. *Mol Cancer.* 2020;19(1):53.
13. Yang S, Wei J, Cui YH, Park G, Shah P, Deng Y, et al. m(6)A mRNA demethylase FTO regulates melanoma tumorigenicity and response to anti-PD-1 blockade. *Nat Commun.* 2019;10(1):2782.
14. Li N, Kang Y, Wang L, Huff S, Tang R, Hui H, et al. ALKBH5 regulates anti-PD-1 therapy response by modulating lactate and suppressive immune cell accumulation in tumor microenvironment. *Proc*

- Natl Acad Sci U S A. 2020;117(33):20159–70.
15. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–7.
  16. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010;26(12):1572–3.
  17. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
  18. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Softw*. 2011;39(5):1–13.
  19. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
  20. Hussain M, Adah D, Tariq M, Lu Y, Zhang J, Liu J. CXCL13/CXCR5 signaling axis in cancer. *Life Sci*. 2019;227:175–86.
  21. Lin C, He H, Liu H, Li R, Chen Y, Qi Y, et al. Tumour-associated macrophages-derived CXCL8 determines immune evasion through autonomous PD-L1 expression in gastric cancer. *Gut*. 2019;68(10):1764–73.
  22. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612.
  23. Chen J, Yu K, Zhong G, Shen W. Identification of a m(6)A RNA methylation regulators-based signature for predicting the prognosis of clear cell renal carcinoma. *Cancer Cell Int*. 2020;20:157.
  24. Yi L, Wu G, Guo L, Zou X, Huang P. Comprehensive Analysis of the PD-L1 and Immune Infiltrates of m(6)A RNA Methylation Regulators in Head and Neck Squamous Cell Carcinoma. *Mol Ther Nucleic Acids*. 2020;21:299–314.
  25. Désert R, Mebarki S, Desille M, Sicard M, Lavergne E, Renaud S, et al. "Fibrous nests" in human hepatocellular carcinoma express a Wnt-induced gene signature associated with poor clinical outcome. *Int J Biochem Cell Biol*. 2016;81(Pt A):195–207.
  26. Zhang QN, Zhu HL, Xia MT, Liao J, Huang XT, Xiao JW, et al. A panel of collagen genes are associated with prognosis of patients with gastric cancer and regulated by microRNA-29c-3p: an integrated bioinformatics analysis and experimental validation. *Cancer Manag Res*. 2019;11:4757–72.
  27. Huang R, Gu W, Sun B, Gao L. Identification of COL4A1 as a potential gene conferring trastuzumab resistance in gastric cancer based on bioinformatics analysis. *Mol Med Rep*. 2018;17(5):6387–96.
  28. Jiang Y, He J, Guo Y, Tao H, Pu F, Li Y. Identification of genes related to low-grade glioma progression and prognosis based on integrated transcriptome analysis. *J Cell Biochem*. 2020;121(5–6):3099–111.

29. Hao S, Lv J, Yang Q, Wang A, Li Z, Guo Y, et al. Identification of Key Genes and Circular RNAs in Human Gastric Cancer. *Med Sci Monit.* 2019;25:2488–504.
30. Wei S, Lu J, Lou J, Shi C, Mo S, Shao Y, et al. Gastric Cancer Tumor Microenvironment Characterization Reveals Stromal-Related Gene Signatures Associated With Macrophage Infiltration. *Front Genet.* 2020;11:663.
31. Pan Y, Ma P, Liu Y, Li W, Shu Y. Multiple functions of m(6)A RNA methylation in cancer. *J Hematol Oncol.* 2018;11(1):48.
32. Knuckles P, Lence T, Haussmann IU, Jacob D, Kreim N, Carl SH, et al. Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA-binding factor Rbm15/Spenito to the m(6)A machinery component Wtap/FI(2)d. *Genes Dev.* 2018;32(5–6):415–29.
33. Miao R, Dai CC, Mei L, Xu J, Sun SW, Xing YL, et al. KIAA1429 regulates cell proliferation by targeting c-Jun messenger RNA directly in gastric cancer. *J Cell Physiol.* 2020.
34. Barceló C, Etchin J, Mansour MR, Sanda T, Ginesta MM, Sanchez-Arévalo Lobo VJ, et al. Ribonucleoprotein HNRNPA2B1 interacts with and regulates oncogenic KRAS in pancreatic ductal adenocarcinoma cells. *Gastroenterology.* 2014;147(4):882 – 92.e8.
35. Alpatov R, Lesch BJ, Nakamoto-Kinoshita M, Blanco A, Chen S, Stützer A, et al. A chromatin-dependent role of the fragile X mental retardation protein FMRP in the DNA damage response. *Cell.* 2014;157(4):869–81.
36. Taha M, Haghighi F, Stefanski A, Nakhaei-Rad S, Kazemineh Jasemi N, Al Kabbani M, et al. Novel FMRP interaction networks linked to cellular stress. *The FEBS journal.* 2020.
37. Li J, Rao B, Yang J, Liu L, Huang M, Liu X, et al. Dysregulated m6A-Related Regulators Are Associated With Tumor Metastasis and Poor Prognosis in Osteosarcoma. *Front Oncol.* 2020;10:769.
38. Zalfa F, Panasiti V, Carotti S, Zingariello M, Perrone G, Sancillo L, et al. The fragile X mental retardation protein regulates tumor invasiveness-related pathways in melanoma cells. *Cell Death Dis.* 2017;8(11):e3169.
39. Fridman WH, Pagès F, Sautès-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer.* 2012;12(4):298–306.
40. Hui L, Chen Y. Tumor microenvironment: Sanctuary of the devil. *Cancer Lett.* 2015;368(1):7–13.
41. Whiteside TL. The tumor microenvironment and its role in promoting tumor growth. *Oncogene.* 2008;27(45):5904–12.
42. Li H, van der Leun AM, Yofe I, Lubling Y, Gelbard-Solodkin D, van Akkooi ACJ, et al. Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. *Cell.* 2019;176(4):775 – 89.e18.
43. Zander R, Schauder D, Xin G, Nguyen C, Wu X, Zajac A, et al. CD4(+) T Cell Help Is Required for the Formation of a Cytolytic CD8(+) T Cell Subset that Protects against Chronic Infection and Cancer. *Immunity.* 2019;51(6):1028–42.e4.

## Figures

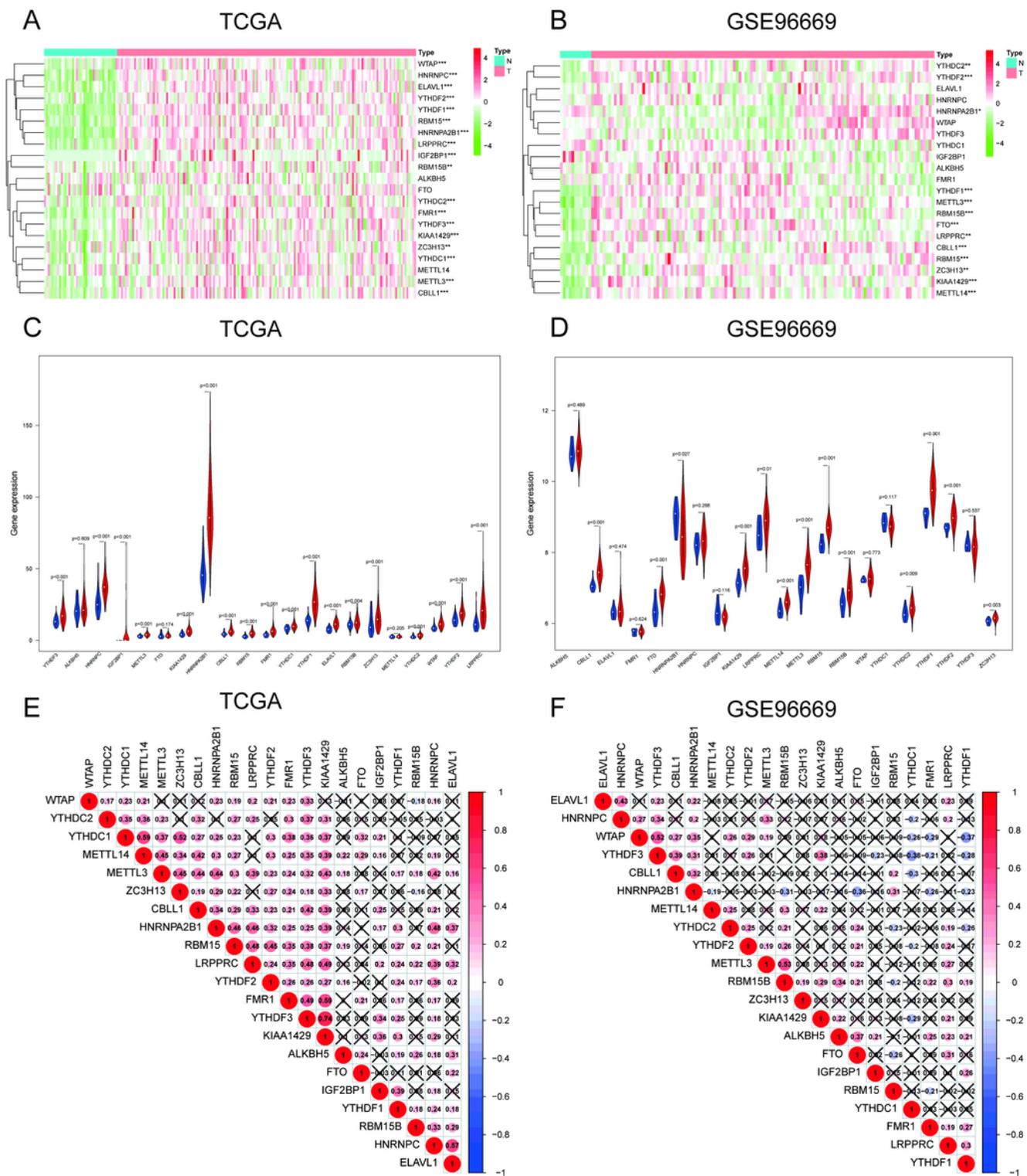
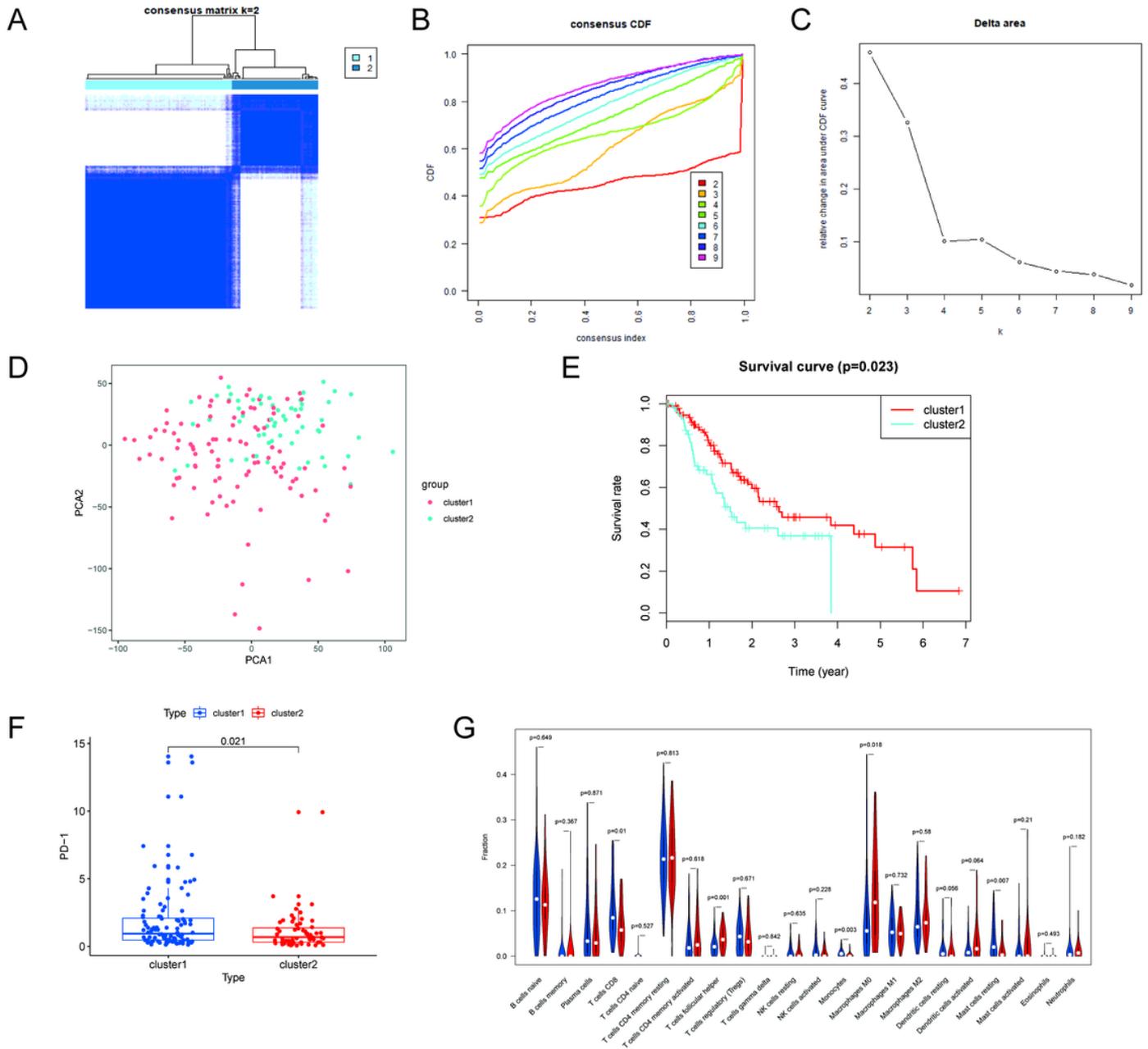


Figure 1

The landscape of m6A RNA methylation regulators in GEA. (A, B) Heatmaps of expression levels of 21 m6A RNA methylation regulators (normal sample vs. tumor sample) from TCGA database (A) and GSE96669 database (B). (C, D) Violin diagrams visualizing 21 m6A RNA methylation regulators in GEA (assume blue is normal and red is gastric cancer) corresponding to (A, B). (E, F) Spearman correlation analysis of the 21 m6A regulators in GEA samples from the TCGA database (E) and GSE96669 database (F). \* $P < 0.05$ ; \*\* $P < 0.01$ ; and \*\*\* $P < 0.001$ ; N, normal sample; T, tumor sample



**Figure 2**

Differential overall survival, TIICs and risk-signature regulators expression of GEA in the cluster 1/2 subgroups. (A) Consensus clustering matrix for k = 2; (B) consensus clustering cumulative distribution function (CDF) for k = 2–9; (C) relative change in area under CDF curve for k = 2–9; (D) principal

component analysis of the total RNA expression profile in the TCGA dataset. GEA in the cluster1 subgroup are marked with red, and the cluster2 subgroup is marked with blue. (E)Kaplan–Meier overall survival (OS) curves for patients in the cluster1/2 subgroup. (F)The differential expression of PD-1 between cluster1 and cluster2 (G) Vioplot visualizing differentially expressed immune cells between cluster 1 and cluster 2 (assume blue is cluster 1 and red is cluster 2). GEA: Gastroesophageal Adenocarcinomas.

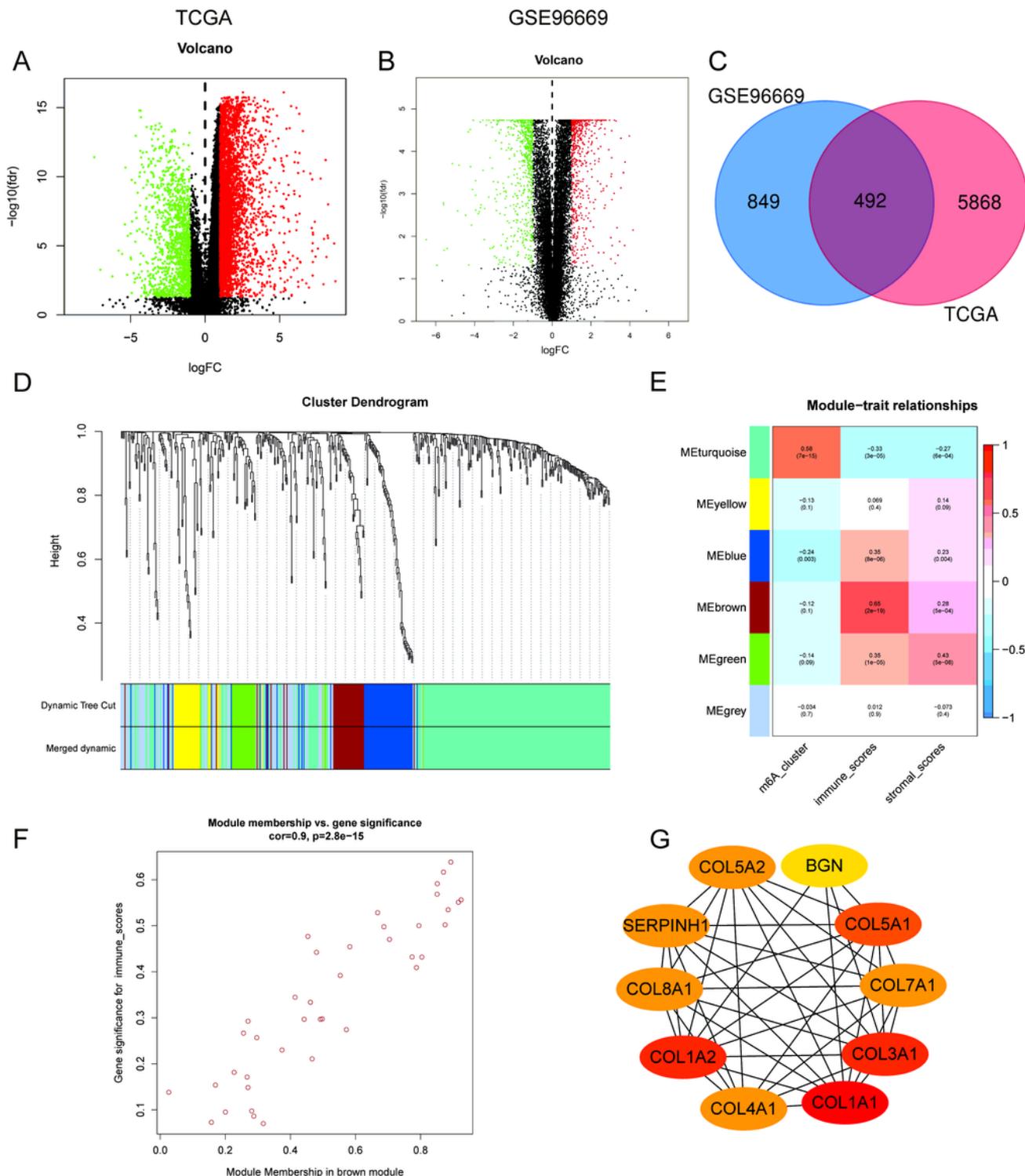
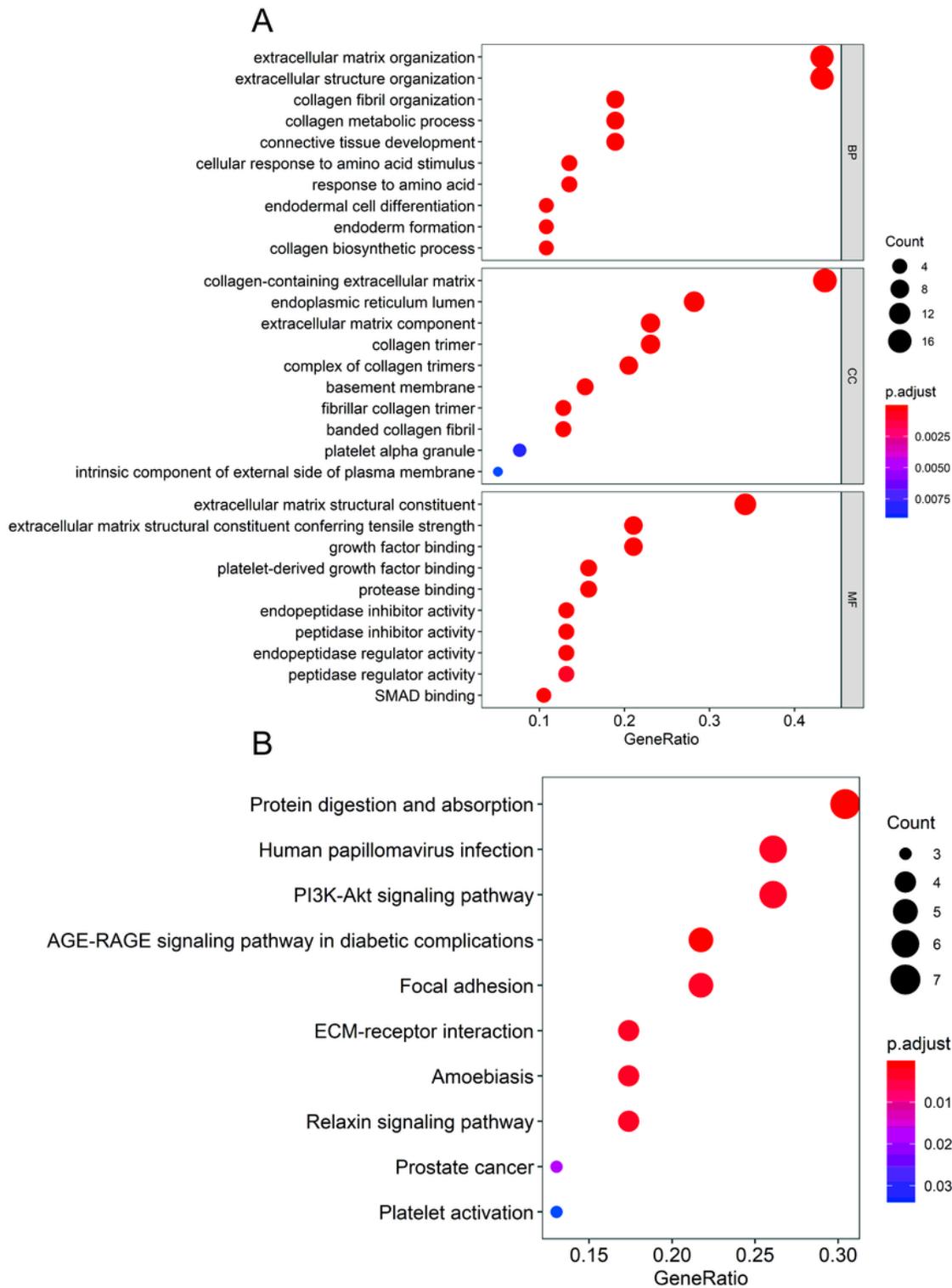


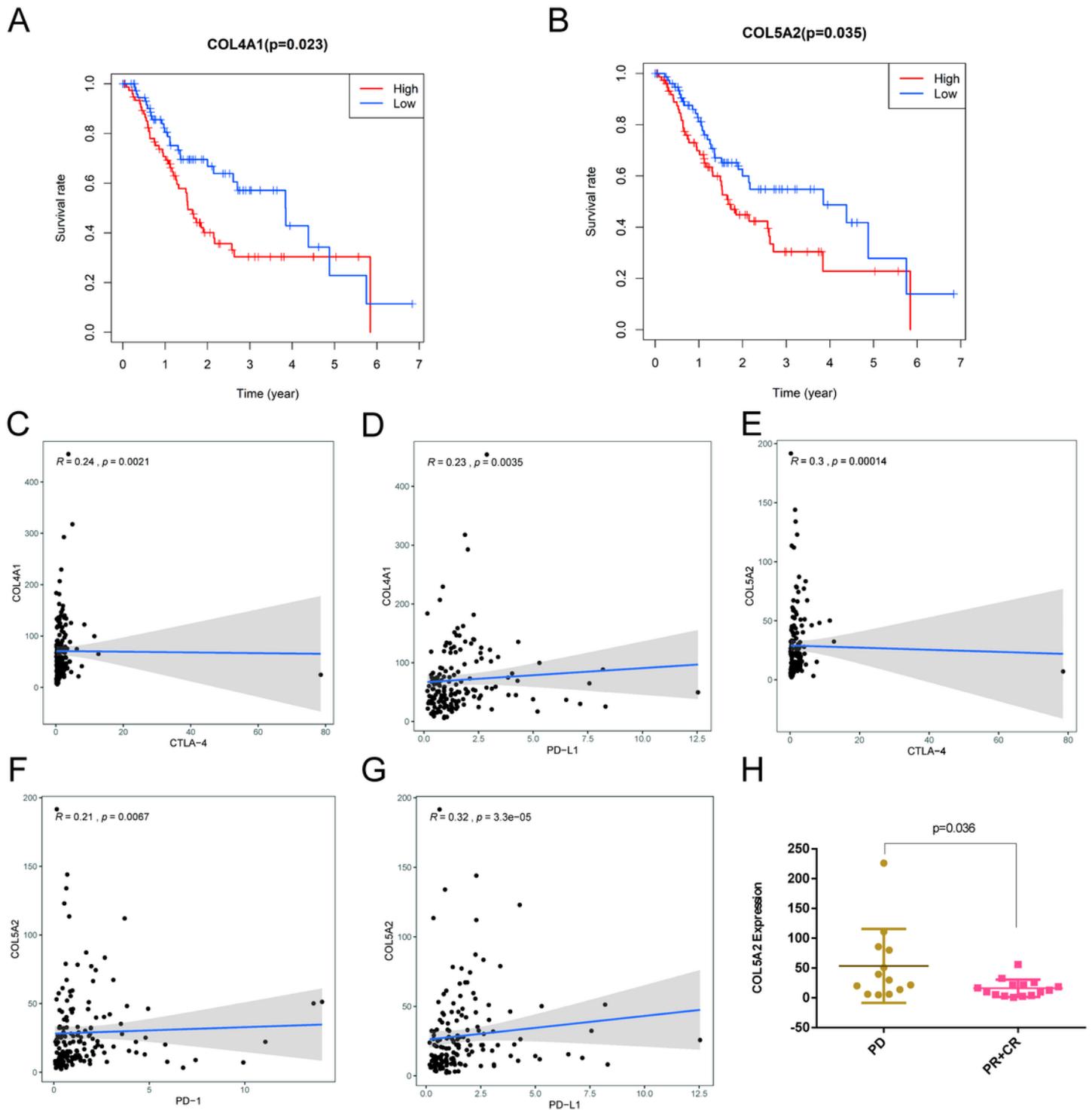
Figure 3

Identification of DEGs and Construction of weighted co-expression network with WGCNA. (A, B) Volcano map of DEGs in TCGA and GSE9669, respectively; green indicates downregulated genes, and red indicates upregulated genes. (C) Venn diagrams of differentially expressed genes of the TCGA datasets and GSE9669 dataset. (D) The cluster dendrogram of genes of GEA patients. Each branch in the figure represents one gene, and every color below represents one co-expression module. (E) Correlation between the gene module and clinical characteristics, including m6A cluster, immune scores and stromal scores. The correlation coefficient in each cell represented the correlation between the gene module and the clinical characteristics, which decreased in size from red to blue. (F) Scatter diagram for module membership vs. gene significance in the brown module. (G) Identification of the hub genes from the PPI network in the brown module using maximal clique centrality (MCC) algorithm. GEA: Gastroesophageal Adenocarcinomas. PPI: protein-protein interaction



**Figure 4**

Functional enrichment analysis. (A) GO functional enrichment analysis of the genes in the brown module. (B) KEGG pathway functional enrichment analysis of the genes in the brown module. GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes.



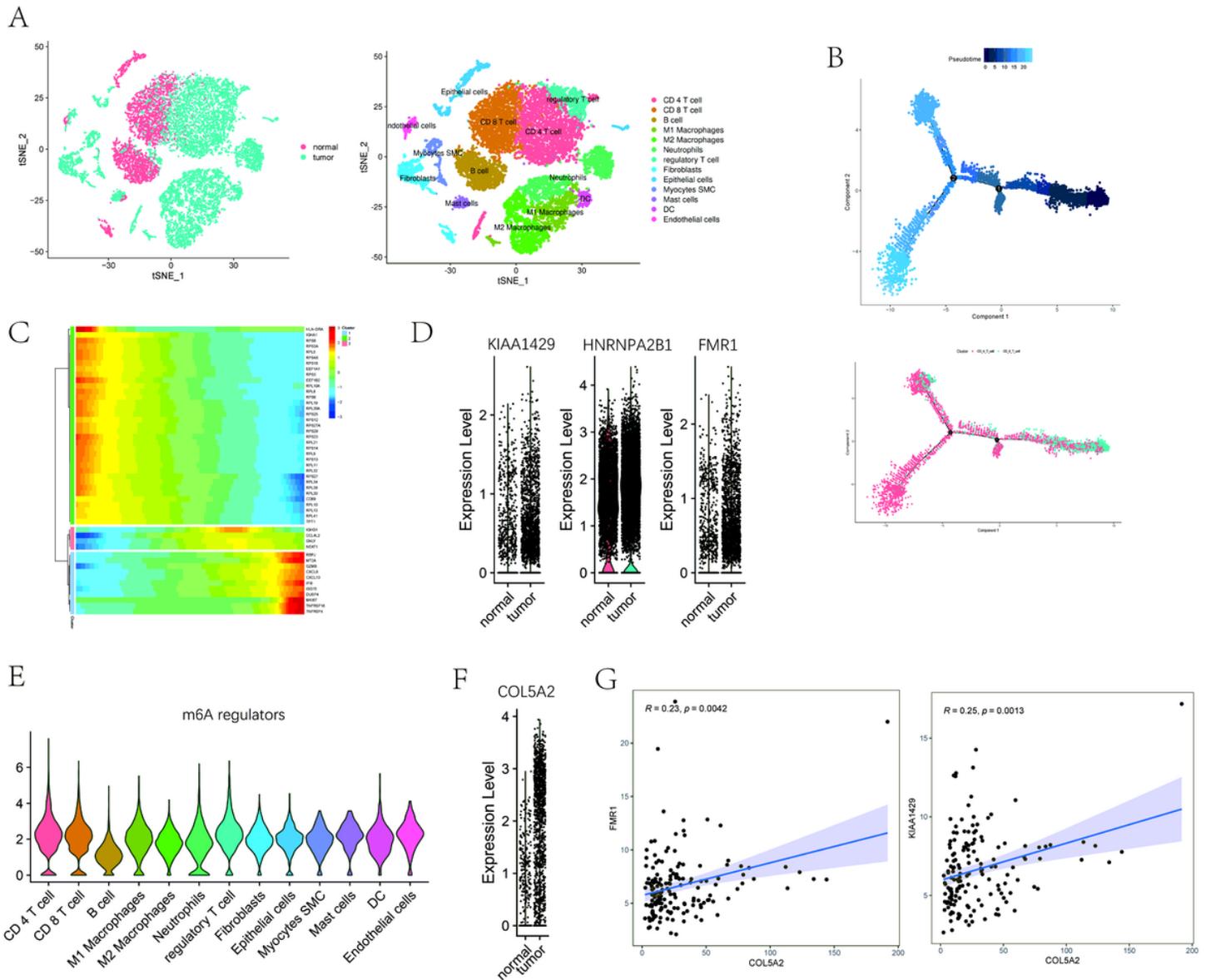
**Figure 5**

Survival analysis of hub genes and their role in immunotherapy. (A-B) Comparison of prognosis between low and high expression level of three hub genes (COL4A1, COL5A2) using Kaplan–Meier curves. (C-G) Correlation analysis of three hub genes with immunological check-point (PD-1/L1 and CTLA-4) expression. (H) Distribution of COL5A2 expression in distinct anti-PD-L1 clinical response groups.

# Image not available with this version

## Figure 6

Risk signature with three m6A RNA methylation regulators. (A). The process of building the signature containing 21 m6A RNA methylation regulators. The hazard ratios (HR), 95% confidence intervals (CI) calculated by univariate Cox regression. (B-C) The coefficients calculated by multivariate Cox regression using LASSO is shown. (D) Kaplan–Meier overall survival (OS) curves for patients in the TCGA datasets are assigned to high and low risk groups based on the risk score. (E) ROC curves showed the predictive efficiency of the risk signature. (F). Univariate Cox regression analyses of the correlation between clinicopathological features and overall survival of patients in the TCGA datasets. (G) Multivariate Cox regression analyses of the correlation between clinicopathological features and overall survival of patients in the TCGA datasets. (H) Risk score by cluster1/2 in TCGA cohort. (I) The differential expression of three risk-signature regulators (KIAA1429, HNRNPA2B1 and FMR1) between cluster1 and cluster2. (J) 22 different immune cells' abundance inferred by CIBERSORT for different risk groups. (K) Enrichment plots from gene set enrichment analysis (GSEA) between high- and low- risk score groups.



**Figure 7**

Identifying infiltrated cell types in GEA and non-malignant tissues. (A) t Stochastic neighbor Embedding (tSNE) plots show sample origin and cell types for the 18070 cells. (B) Trajectory of differentiation from CD8+ T cell into CD4+ T cells predicted by a monocle. (C) Heatmap show upregulated or downregulated top 50 genes in the differentiation process (D) Violin plot shows differential expression level of KIAA1429, HNRNPA2B1 and FMR1 between tumor and normal samples. (E) Violin plot shows the total expression level of three m6A regulators (KIAA1429, HNRNPA2B1 and FMR1) across major cell types. (F) Violin plot shows the differential expression level of COL5A2 between tumor and normal samples. (G) Correlation between COL5A2 level and the expression of KIAA1429 and FMR1 in GEA samples of the TCGA database.

## Supplementary Files

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

- [FigureS1.tif](#)
- [FigureS2.tif](#)
- [FigureS3.tif](#)
- [FigureS4.tif](#)
- [FigureS5.tif](#)
- [TableS.docx](#)
- [TableS3.xls](#)
- [TableS4.xls](#)