Cellular and Molecular Biology
E-ISSN : 1165-158X / P-ISSN : 0145-5680
www.cellmolbiol.org

The prognostic, immunological and single-cell features of m6A molecules in cervical cancer

Xiaoqiu Zhu[1], Xiaolu Song[2], Qi Chen[3], Ye Peng[2], Fangfang Shi[2*]

[1] Outpatient nursing, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, 3 East Qingchun Road, Hangzhou, Zhejiang, China
[2] Cancer center, Department of Hematology, Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College, Hangzhou, Zhejiang, China
[3] Department of Orthopedics, Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College, Hangzhou, Zhejiang, China

ARTICLE INFO

Original paper

Article history:
Received: May 12, 2023
Accepted: July 06, 2023
Published: September 30, 2023

Keywords:

Cervical cancer, m6A, clusters, bioinformatics, single-cell

ABSTRACT

Cervical cancer (CC) is a growing health concern, emphasizing the need for reliable biomarkers in treatment selection and prognosis assessment. We analyzed gene expression profiles and clinicopathological data from The Cancer Genome Atlas (TCGA) for CC. Using Consensus Cluster Plus, we applied machine learning to cluster the CC cohort. Differential analysis was performed using the edge R package, while weighted correlation network analysis (WGCNA) was conducted using the WGCNA package. Single-sample gene set enrichment analysis (ssGSEA) evaluated immune cell abundance and computed the m6Ascore. Western blot and Q-PCR validated the m6A score in CC. Common copy number variation alterations were observed in the 23 m6A-related genes in CC, and their mutation frequency was summarized in a waterfall chart. Patients were grouped into two clusters, m6AclusterA and m6AclusterB. Improved clinical outcomes were observed in m6AclusterA, while m6AclusterB exhibited higher infiltration of 14 immune cell types. WGCNA analysis generated seven integrated modules, enriched in several biological processes. Prognostic differential genes were used to generate two gene clusters (gene Cluster I and gene Cluster II). Using ssGSEA, the m6Ascore was calculated for each patient. Lower m6Ascore correlated with better clinical outcomes, lower gene mutation frequency, and wild-type status. We investigated the sensitivity of high and low m6Ascore to immunotherapy, visualized through violin and UMAP diagrams showcasing crosstalk among single-cell clusters. The key gene PFKFB4 showed higher expression in CC cell lines and tumor tissues compared to normal cells and tissue. Our study elucidates the role of m6A molecules in predicting prognosis, biological features, and appropriate treatment for CC patients.

Doi: http://dx.doi.org/10.14715/cmb/2023.69.9.13

Introduction

Cervical cancer (CC) is the fourth most common malignancy diagnosed in women worldwide, after breast cancer, colorectal cancer and lung cancer (1). It is also widely recognized as a malignancy caused by the infection of human papillomavirus (1). Despite the increasing rates of early screening and vaccination for CC, the incidence and mortality rates have not decreased significantly. In 2018, approximately 570,000 women developed CC and 311,000 women died of CC (2). The treatment plan of CC is decided according to the clinical stage. Surgery is still the treatment method for early CC. Postoperative adjuvant chemoradiotherapy is decided according to the pathological results. At present, the prognosis of patients with CC is mainly determined by the selection of stage, local tumor size, pathological type and differentiation degree, but they are not enough to accurately estimate the prognosis of patients, and there are no effective biomarkers to predict the prognosis of CC (3). Therefore, it is very important to find new reliable biomarkers for the treatment selection and prognosis assessment of CC.

Undoubtedly, the most important prognostic factors for CC are FIGO stage, pelvic lymph node metastasis, and clinicopathological features of the primary tumor, including tumor size, deep interstitial invasion, Periuterine infiltration, positive surgical margin, and LVSI (4,5). There is increasing interest in biomarkers that can predict response to treatment and survival, driven by the need to individualize treatment and assess outcomes for patients (6). By correlating the aggressive clinical phenotype of tumors with gene expression profiles, a series of molecular markers with potential prognostic value have been identified.

The rapid development of high-throughput sequencing technology enables researchers to obtain tumor genome, transcriptome, and proteome information (7). The Cancer Genome Atlas (TCGA) database effectively integrates the data and information of high-throughput sequencing (8). The data resources are shared by researchers, which is of great value to cancer research (8). TCGA is an important database of cancer genomes jointly managed by the National Cancer Institute (NCI) and the National Human Ge-

* Corresponding author. Email: 497874955@qq.com

nome Research Institute (NHGRH) (8). TCGA database provides genomic, epigenome and transcriptome information for patients with different types of cancer, including gene expression and clinical data for CC. However, how to reliably transform the biological codes contained in big data into results with clinical guiding value is still facing great challenges (9). Nowadays, with the help of various bioinformatics tools and machine learning algorithms, more candidate molecules with potential prognostic value for CC have been reported successively (10–12). However, the prognostic relevance of some biological factors requires further investigation due to the lack of high-throughput data or validation failures in independent centers. Although some biomarkers have been used to predict the clinical outcome of patients with CC, their sensitivity and/or specificity are still unsatisfactory. Therefore, there is an urgent need to find more valuable biomarkers to diagnose, monitor recurrence and assess prognosis (6,13).

The R language is a kind of free software programming language and operating environment that can be used in statistical analysis, drawing and data mining. The R language can carry out some basic statistical tests, build linear or nonlinear models, analyze time series, or classify and cluster data. The source code of R can be freely downloaded for use. The functionality of R can be enhanced by specially written packages. A package is a well-defined collection of R functions, data, and precompiled code. R package can be added with statistical methods, drawing methods, programming interfaces and data output, input and other functions (14,15).

N6-methyladenosine (m6A) is a specific methylation modification of the sixth nitrogen atom (N) on adenine (A) catalyzed by methyl transferase, which can exist in a variety of RNAs, but is most abundant in mRNA of eukaryotic cells (16,17). In recent years, through an in-depth study on the mechanism of m6A methylation, it has been found that the m6A methylation modification process in mRNA is not stable (18), but reversible regulation and action of three related molecules, methyl transferase, demethylase and binding protein (19). Among them, m6A methyl transferases are also called writers (20), m6A demethylases are also called erasers (21), and m6A binding proteins are also called readers (20,21). The m6A methylation of mRNA is similar to the methylation of DNAi, which modifiers the expression products after transcription without changing the gene base sequence (22). Related studies have shown that m6A affects and changes gene expression mainly by intervening and regulating mRNA splicing, translation and stability (23). The mechanism of m6A methylation modification in cervical cancer still needs to be further explored.

Materials and Methods

Data acquisition and processing for CC

The transcriptome expression data and clinical data of CC used in the study were obtained from TCGA database (https://cancergenome.nih.gov/) (24,25). TCGA database contains 309 CC tissues and 3 adjacent tissues. A total of 23 m6A-related genes (FMR1, RBMX, METTL3, HNRNPC, YTHDC1, VIRMA, LRPPRC, YTHDF1, YTHDF2, YTHDF3, ALKBH5, METTL14, METTL16, WTAP, HNRNPA2B1, IGFBP2, YTHDC2, ZC3H13, RBM15, IGFBP1, IGFBP3, RBM15B, and FTO) were acquired from the earlier literature (21,26). GEPIA2 tool was used to explore the relationship of genes with clinical information in TCGA (http://gepia2.cancer-pku.cn/). Single-cell RNA data was downloaded from the GEO (GSE 171894) and processed by the Seurat and the Cell chat R package.

Analysis of differentially expressed genes for CC

The edgeR package works on a table of integer read counts, with rows corresponding to genes and columns corresponding to individual libraries. The count represents the total number of readings aligned to each gene (or other genomic loci). edgeR focuses on differential expression analysis rather than quantification of expression level (27). It is related to the relative changes in expression levels under different conditions, but not directly related to the estimated absolute expression levels. In this study, the edgeR package was used to screen differentially expressed genes (27).

Weighted correlation network analysis

The concept of Weighted correlation network analysis (WGCNA) was first proposed in 2005 (28). Based on the similarity of expression profiles among different samples, a group of genes with the same expression pattern were grouped into different modules, which were involved in separate biological functions and/or regulated by a common mechanism, so as to identify co-expressed gene sets (29). Candidate biomarkers or potential therapeutic targets can then be identified based on the correlation between gene sets and between gene sets and clinical phenotypes (30,31). We used the WGCNA package to construct gene co-expression networks and screen hub genes (32). The steps are as follows: (1) Eliminate outlier samples to ensure the reliability of network construction results; (2) The standard scale-free model fitting index R^2=0.9 was used to select the soft threshold (power=5); (3) Transform the adjacency matrix as a measure of topological similarity into topological overlap matrix (TOM), and calculate the corresponding dissimilarity degree (1-TOM); (4) The system cluster map was drawn to determine the module composed of a set of interrelated genes; (5) module eigengenes (MEs) were calculated to evaluate the correlation between each module eigengenes and clinical features; (6) Extraction of hub genes in important modules; (7) The correlation coefficient between gene significance (GS) and module membership (MM) was calculated and the p-value was obtained.

Establishment of m6A-clusters for CC patients

Consensus clustering is an algorithm that can be used to identify the members and number of clusters in a data set, such as microarray gene expression. Consensus clustering is usually used to determine the optimal number of clusters k. Consensus clustering verifies the rationality of clustering based on resampling, and its main purpose is to evaluate the stability of clustering. Consensus clustering was carried out using the ConsensusClusterPlus package to establish m6A-clusters in this study (33).

Single-sample gene set enrichment analysis

The single-sample gene set enrichment analysis (ssGSEA) quantifies the relative abundance of each immune cell type in the tumor microenvironment (TME) of a single sample by normalized enrichment score (NES) (34). The 28 defined immune cell types and corresponding gene tags

were obtained from The Cancer Immunome Atlas (TCIA, https://tcia.at/) (35). ssGSEA algorithm can also score genes in the genome of interest (36). In this study, the value obtained by summarizing m6A-related genes was m6Ascore.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) is to evaluate the distribution trend of gene set S with known phenotype in sequencing data L to be analyzed and arrange in descending order according to the association degree between gene members or gene subsets in S and known phenotype. The contribution of gene members of a gene set S to the known phenotype is judged according to whether the gene members are enriched at the top or bottom of L (37). In this study, we use the GSEA function in the clusterProfiler package for enrichment analysis (38). Reference gene sets were downloaded from the molecular signature database (MsigDB) (39), including classical pathway gene sets from the Kyoto Encyclopedia of Genes and Genomes, and bioprocess gene sets from Gene Ontology (GO) annotations.

Comparative analysis of somatic mutations

We converted somatic mutation information from ANNOVAR annotation files to MAF file format using the AnnovarToMaf function in the maftools package (40). Finally, we obtained the mutation data of CC samples to evaluate the mutation frequency of m6A-related genes and compare the mutation frequency differences among different subgroups.

RT-PCR

Total mRNA was isolated from cultured cells or liver samples using TRIzol reagent (Invitrogen), according to the manufacturer's instructions. 1ug mRNA was reverse transcribed into cDNA using HiScript III All-in-one RT SuperMix Perfect for qPCR according to the manufacturer's protocol. SYBR Green (YEASEN Biotech) was applied to quantify PCR amplification. The RT-PCR Primer as follows: PFKFB4-F:TCCCCACGGGAATTGACAC,PFKFB4-R:GGGCACACCAATCCAGTTCA.

Cervical cancer tissue samples

The research was approved by the Research Ethics Committee of The Second Affiliated Hospital of Fujian Medical University in the study. Total, 14 pairs of cervical cancer tissues and their corresponding adjacent normal tissues were obtained from patients who underwent surgery between March 2020 and May 2021 at the Second Affiliated Hospital of Fujian Medical University, Quanzhou City, China. The 14 paired samples were applied to protein extraction for Western blot detection.

Cell culture and transfection

cervical cancer cell lines C-33a, Si-Ha, CaSki and HUCEC cells were purchased from the Cell Bank of Type Culture Collection of the Chinese Academy of Sciences. All cells were routinely cultured in DMEM (Invitrogen) supplemented with 10% fetal bovine serum (FBS; Life Technologies, New York, USA) in a humidified incubator containing 5% CO2 at 37°C.

Western blot

The cancer tissue and paracancerous tissue samples of cervical cancer patients were removed from the liquid nitrogen tank, fully ground by hand, and centrifuged to take the supernatant, and the protein concentration was measured by total protein concentration determination (BCA method). The following primary antibodies: PFKFB4 polyclonal antibody (Abcam, ab154588, 1:1000) and the GAPDH-specific polyclonal antibody (Abcam; ab8245, 1:1000).

Results

The landscape of CNV frequency and expression of m6A-related genes in CC

There was a total of 23 m6A-related genes (FMR1, RBMX, METTL3, HNRNPC, YTHDC1, VIRMA, LRPPRC, YTHDF1, YTHDF2, YTHDF3, ALKBH5, METTL14, METTL16, WTAP, HNRNPA2B1, IGFBP2, YTHDC2, ZC3H13, RBM15, IGFBP1, IGFBP3, RBM15B, and FTO) included in our study. We carried out an analysis to evaluate the CNV frequency and we found that CNV alteration of the 23 m6A-related genes in CC was common (Figure 1A). Our data revealed FMR1 and

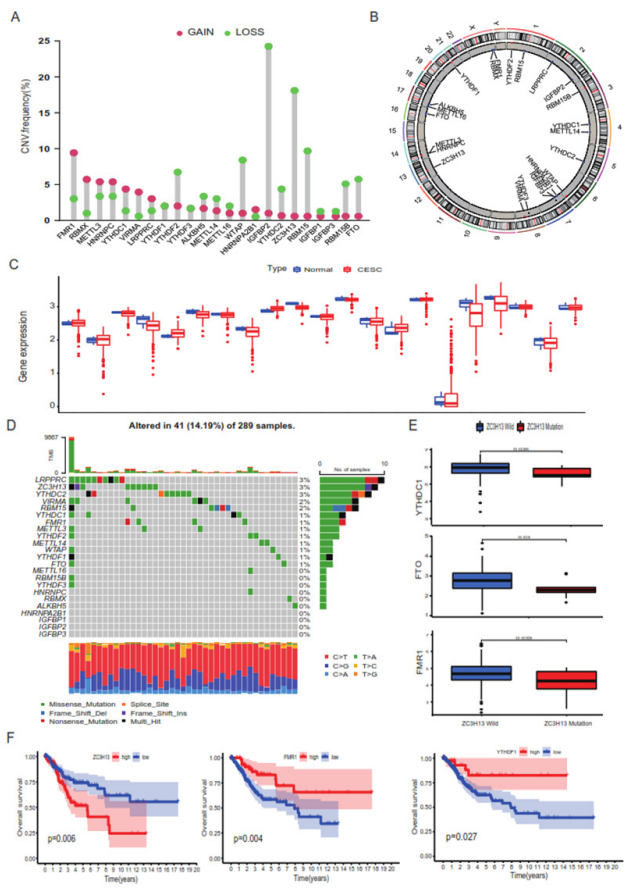

Figure 1. Landscape of CNV frequency and expression of m6A-related genes in CC. **(A)** CNV frequency of gain and loss for 23 m6A-related genes. **(B)** Circus plots displaying distributions of m6A-related genes and corresponding. **(C)** The expression of 21 m6A-related genes between normal and tumor tissues for CC. **(D)** The mutation frequency of the m6A-related genes in waterfall chart. **(E)** The expression differences of three m6A regulators (YTHDC1, FTO, and FMR1) between CC patients with ZC3H13 non-mutant and ZC3H13 mutant. **(F)** The results of survival analysis for ZC3H13, FMR1, and YTHDF1.

RBMX were observed with a general frequency of CNV gain (Figure 1A), while IGFBP2, ZC3H13, RBM15, and WTAP had a widespread frequency of CNV loss (Figure 1A). Further, we displayed the location of the 23 m6A-related genes on chromosomes, as follows: YTHDF2 (chr1), RBM15 (chr1), LRPPRC (chr2), IGFBP2 (chr2), RBM15B (chr3), YTHDC1 (chr4), METTL14 (chr4), YTHDC2 (chr5), WTAP (chr6), HNRNPA2B1 (chr7), IGFBP1 (chr7), IGFBP3 (chr7), YTHDF3 (chr8), VIRMA (chr8), ZC3H13 (chr13), HNRNPC (chr14), METTL3 (chr14), FTO (chr16), METTL16 (chr17), ALKBH5 (chr17), YTHDF1 (chr20), RBMX (chrX), and FMR1 (chrX). Among the 23 m6A-related genes, we found a total of 21 m6A-related genes (METTL3, METTL14, WTAP, ZC3H13, RBM15, RBM15B, YTHDC1, YTHDC2, YTHDF1, YTHDF2, YTHDF3, HNRNPC, FMR1, LRPPRC, HNRNPA2B1, IGFBP1, IGFBP2, IGFBP3, RBMX, FTO, and ALKBH5) differentially expressed between normal and CESC tissues (Figure 1C). We summarized the mutation frequency of the m6A-related genes in waterfall chart (Figure 1D). Among the 289 CC samples, 41 experienced mutations, accounting for 14.19% of the total samples (Figure 1D). LRPPRC (3%), ZC3H13 (3%) and YTHDC2 (3%) showed the highest mutation frequency. VIRMA (2%), RBM15 (2%), YTHDC1 (1%), FMR1 (1%), METTL3 (1%), YTHDF2 (1%), METTL14 (1%), WTAP (1%), YTHDF1 (1%), and FTO (1%) occurred genetic mutations in CC (Figure 1D). Nevertheless, METTL16, RBM15B, YTHDF3, HNRNPC, RBMX, ALKBH5, HNRNPA2B1, IGFBP1, IGFBP2, and IGFBP3 had nonsense mutation (Figure 1D). Subsequently, we compared the expression differences of three m6A regulators (YTHDC1, FTO, and FMR1) between CC patients with ZC3H13 non-mutant and ZC3H13 mutant, and we found that the expression of these three m6A regulators was higher in ZC3H13 non-mutant (Figure 1E). The results of survival analysis showed that CC patients with low expression of ZC3H13 had a better prognosis, while those with low expression of FMR1 and YTHDF1 had a worse prognosis (Figure 1F).

Characteristics of m6A-clusters for CC patients

To further explore the value of m6A-related genes in CC, we used the ConsensusClusterPlus algorithm from R to perform a clustering analysis of m6A-related genes, thus determining the optimal number (k=2) of m6A-clusters and two clusters named m6AclusterA, m6AclusterB (Figure 2A and Supplemental Table S1). We observed better clinical outcomes of CC patients in m6AclusterA, however, the difference in survival between m6AclusterA and m6AclusterB did not reach statistical significance (Figure 2B). The heatmap showed the distribution of expression values of 21 m6A-related genes in different subclusters, ages and tumor node metastasis (TNM) classifications. It was easy to observe that the expression values of IGFBP2 and IGFBP3 were higher in m6AclusterA than that in m6AclusterB (Figure 2C). Interestingly, the expression value of IGFBP1 was low in both m6Aclusters (Figure 2C). Based on ssGSEA analysis, we evaluated the infiltration of 23 immune cells of each CC patients and compared the differences between m6AclusterA and m6AclusterB. A total of 14 immune cell types (Activated CD4 T cell, Activated CD8 T cell, CD56 bright natural killer cell, Gamma delta T cell, Immature B cell, MDSC, Macrophage, Mast cell, Natural killer T cell, Neutrophil, Regulatory T cell, T follicular helper cell, Type I T helper cell) showed significant differences between the two m6Aclusters, and all of them showed higher levels of infiltration in m6AclusterB (Figure 2D). We performed GSVA analysis for the two m6Aclusters, the results suggested that Taurine and Hypotaurine metabolism, glycosaminoglycan biosynthesis keratan sulfate, basal cell carcinoma, and hedgehog signaling pathways were enriched in m6AclusterB (Figure 2E). In addition, complement and coagulation cascades, antigen processing and presentation, autoimmune thyroid disease, graft versus host disease, type I diabetes mellitus, allograft rejection, systemic lupus erythematosus, asthma, intestinal immune network for IGA production, cytosolic DNA sensing pathway, RIG I like receptor signaling pathway, NOD-like receptor signaling pathway, JAK-STAT signaling pathway, Toll-like receptor signaling pathway, Natural killer cell-mediated cytotoxicity, and leishmania infection signaling pathways were enriched in m6AclusterA (Figure 2E).

WGCNA analysis identifying seven modules

We identified genes that were differentially expressed between the two m6Aclusters, and then, to further elucidate the mechanistic differences of these differentially expressed genes, we performed WGCNA analysis for these

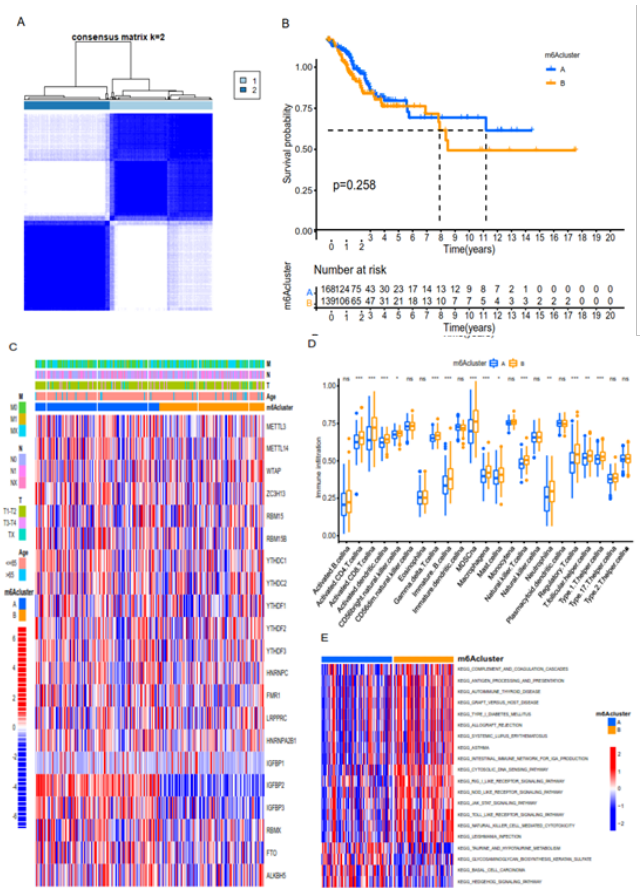

Figure 2. Characteristics of m6A-clusters for CC patients. **(A)** Cluster diagram for subtype analysis of CC samples. The intragroup correlations were the highest and the inter-group correlations were low when k=2. **(B)** Kaplan-Meier survival curve showing survival probability of m6AclusterA and m6AclusterB. **(C)** The heatmap showed the distribution of expression values of 21 m6A-related genes in different subclusters, ages and tumor node metastasis (TNM) classifications. **(D)** The differences between m6AclusterA and m6AclusterB in the infiltration of 23 immune cells. **(E)** GSVA analysis for the two m6Aclusters.

genes. Firstly, the soft threshold was set to five (Figure 3A). High correlation genes were aggregated into modules based on dynamic pruning and clustering, thus clustering these modules and merging modules with a correlation coefficient greater than 0.9 (Figure 3B). A total of seven modules were integrated (Figure 3C). Among the seven modules, the yellow (0.27), blue (0.28), turquoise (0.2), and grey (0.29) modules were positively correlated with m6AclusterA, while the brown (-0.2) and red (-0.14) modules were negatively correlated with m6AclusterA (Figure 3D). As for m6AclusterB, the yellow (-0.27), blue (-0.28), turquoise (-0.2), and grey (-0.29) modules were negatively correlated with m6AclusterB, on the contrary, the brown (0.2) and red (0.14) modules were positively correlated with m6AclusterB (Figure 3D).

Enrichment pathway analysis of m6Aclusters

After WGCNA analysis, we screened out genes that were significantly correlated with m6Aclusters to explore their biological processes. The genes correlated with m6AclusterA were enriched in the developmental process, multicellular organismal process, response to stimulus, localization, metabolic process, biological regulation, signaling, immune system process, negative regulation of the biological process, cellular process, regulation of the biological process, positive regulation of the biological process, locomotion, growth, reproductive process, biomineralization, detoxification, and biological process involved in interspecies interaction between organisms (Figure 4A). The genes correlated with m6AclusterB were enriched in the immune system process, regulation of the biological process, cellular process, response to stimulus, positive regulation of the biological process, locomotion, signaling, negative regulation of the biological process, the biological process involved in interspecies interaction between organisms, multicellular organismal process, localization, biological regulation, metabolic process, developmental process, and viral process (Figure 4C). Hence, the genes correlated with the two m6Aclusters were both enriched in the developmental process, multicellular organismal process, response to stimulus, localization, metabolic process, biological regulation, signaling, immune system process, negative regulation of the biological process, cellular process, regulation of the biological process, positive regulation of the biological process, locomotion, and biological process involved in interspecies interaction between organisms. A network diagram of interactions between different gene sets for m6AclusterA (Figure 4B) and m6AclusterB (Figure 4D) were shown in Figure 4B and Figure 4D.

Characteristics of gene Clusters for CC patients

We identified genes that were differentially expressed between the two m6Aclusters (Supplemental Table S2 and Table S3). Utilizing Univariate Cox analysis, the differential genes with prognostic value were filtrated, which were identified as prognostic genes. Further, we used the ConsensusClusterPlus algorithm from R to perform a clus-

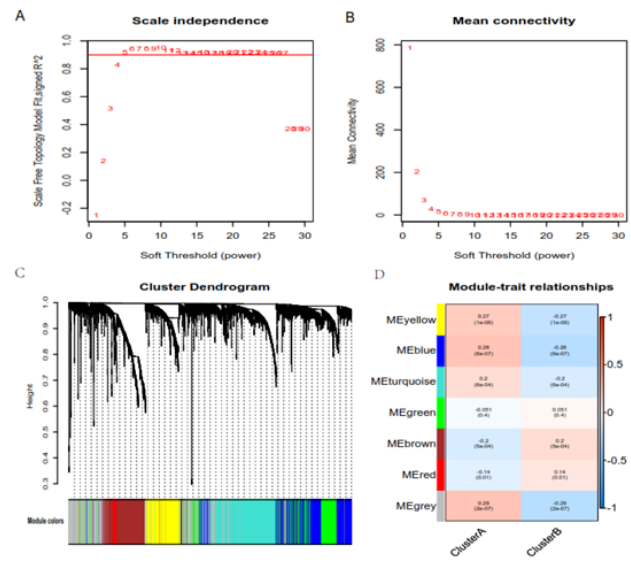

Figure 3. WGCNA analysis integrating seven modules correlated with m6Aclusters. **(A)** Scale independence. The soft Threshold (power) represents the weight, and the ordinate represents the square value of the correlation coefficient between connection degree k and p(k). **(B)** Mean connectivity. Soft Threshold (power) represents the weight, and the ordinate represents the average connection. It is generally required that the power when the square value of the correlation coefficient between k and p(k) reaches 0.9 for the first time is taken as the β value, which can be seen as β=5. **(C)** Systematic cluster tree of genes and gene modules generated by dynamic clipping method. Different colors represent different genetic modules. **(D)** Heatmap of the correlation between module eigengenes and m6Aclusters.

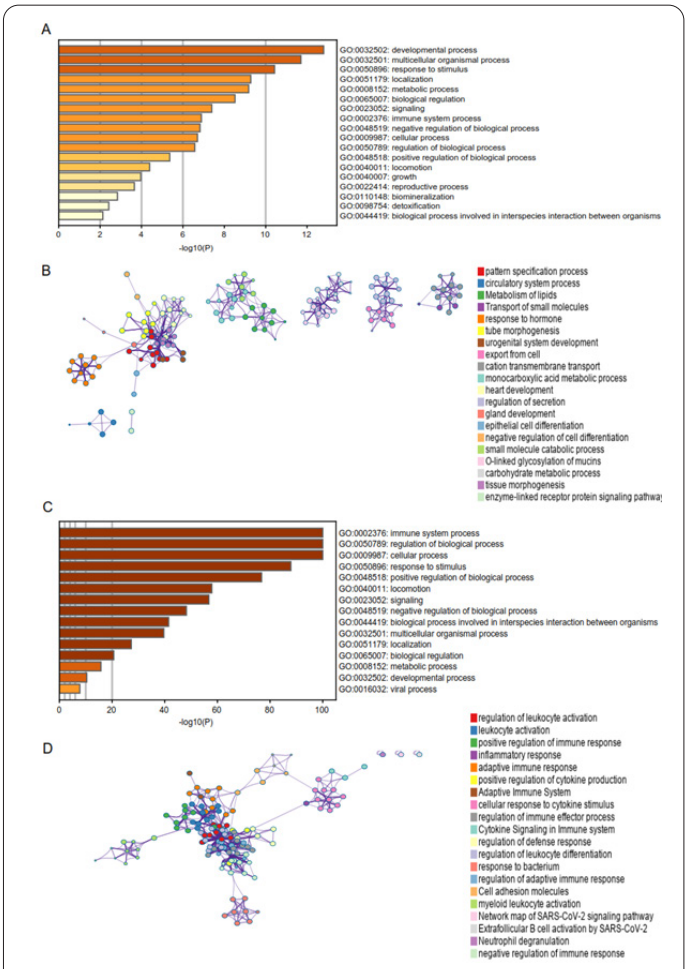

Figure 4. Enrichment pathway analysis of m6Aclusters. **(A)** The biological processes for m6AclusterA. The color represents the p-value. **(B)** A network diagram of interactions between different gene sets for m6AclusterA. **(C)** The biological processes for m6AclusterB. The color represents the p-value. **(D)** A network diagram of interactions between different gene sets for m6AclusterB.

tering analysis of prognostic genes, thus determining the optimal number (k=2) of geneClusters and two clusters named geneClusterI, geneClusterII (Figure 5A). We observed better clinical outcomes of CESC patients in geneClusterII (Figure 5B). The heatmap of gene expression values in the two geneClusters was shown in Figure 5C. As we could see, the expression values of these genes were higher in geneClusterII than that in geneClusterI (Figure 5C). We compared the expression values of m6A-related genes between geneClusterI and geneClusterII. Our data showed that the expression values of METTL3, YTHDF1, YTHDF3, HNRNPC, LRPPRC, IGFBP1, and IGFBP2 were higher in geneClusterI than that in geneClusterII (Figure 5D). Utilizing the ssGSEA algorithm, the m6Ascore for each CC patient was calculated. According to the median m6Ascore, CC patients were divided into a high m6Ascore group and a low m6Ascore group. The better clinical outcomes of CC patients with low m6Ascore were observed (Figure 5E). The m6Ascore was also positively correlated with the majority of immune cells (Figure 5F). The Sankey diagram showed that most CC patients in m6AclusterA were in the low m6Ascore group, and the prognosis was poor. However, most patients in m6AclusterB were divided into the high m6Ascore group, and the prognosis was good (Figure 5G).

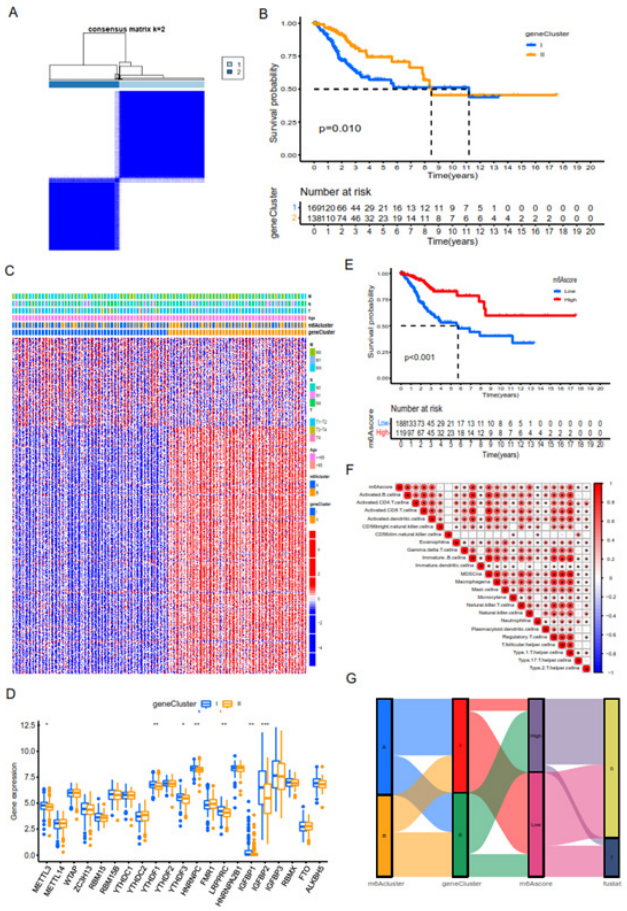

Figure 5. Characteristics of geneClusters for CC patients. **(A)** Cluster diagram for subtype analysis of CC samples. The intragroup correlations were the highest and the inter-group correlations were low when k=2. **(B)** Kaplan-Meier survival curve showing survival probability of geneClustersI and geneClustersII. **(C)** The heatmap of gene expression values in the two geneClusters. **(D)** The expression values of m6A-related genes between geneClusterI and geneClusterII. **(E)** Kaplan-Meier survival curve showing survival probability of m6Ascore. **(F)** Correlation heat map between m6Ascore and immune cells. **(G)** The Sankey diagram shows the relationship among m6Acluster, geneCluster, m6Ascore, and fustat.

Genomic mutation analysis for m6Ascore

The mutational data in the waterfall plot from our study showed that CC patients with low m6Ascore had a lower frequency of gene mutations and tended to be wild-type, while CC patients with high m6Ascore had the opposite frequency (Figure 6A-B). TTN, PIK3CA, KMT2C, MUC16, KMT2D, FLG, SYNE1, EP300, FBXW7, DMD, LRP1B, RYR2, DST, USH2A, ADGRV1, MUC17, HUWE1, LRP2, PCLO, and SYNE2 were the top 20 genes with highest mutation rates (Figure 6A). TTN, PIK-3CA, KMT2C, MUC16, KMT2D, FLG, SYNE1, EP300, FBXW7, DMD, LRP1B, RYR2, DST, USH2A, ADGRV1, MUC17, HUWE1, LRP2, PCLO, and SYNE2 (Figure 6B). Thus, the top 20 genes with mutation rates in the two subgroups were the same (Figure A-B). Further, we explored the prognostic differences of high and low m6Ascore in different clinical factors. The survival curves showed that m6Ascore was a protection factor both for patients with T1-T2 (Figure 6C) and T3-T4 (Figure 6D). In addition, we investigated the sensitivity of high and low m6Ascore to immunotherapy (Figure 6E-I).

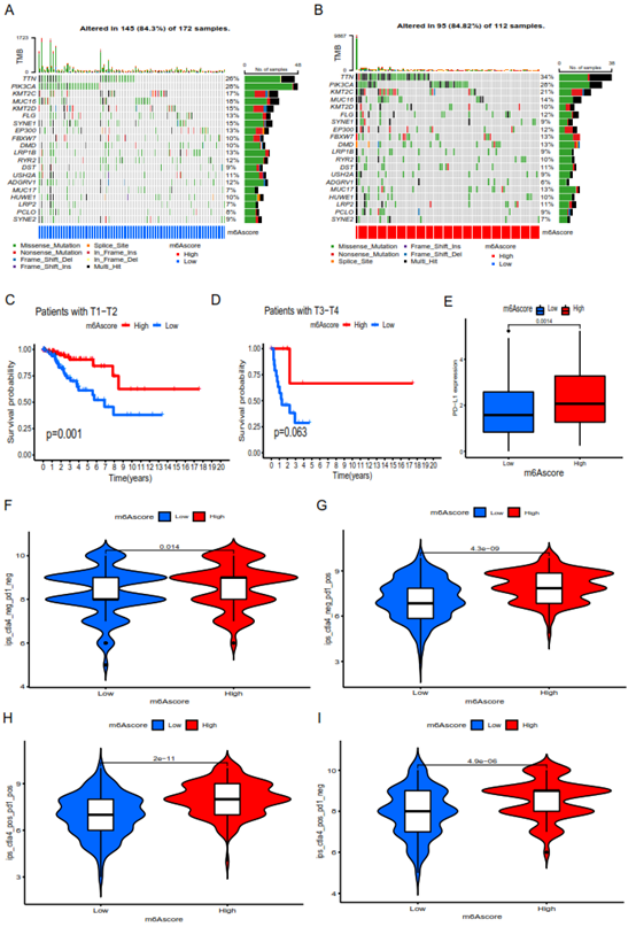

Figure 6. Genomic mutation analysis for m6Ascore. **(A)** Gene mutation frequency in low-m6Ascore. **(B)** Gene mutation frequency in high-m6Ascore. **(C)** The survival curves for patients with T1-T2. **(D)** The survival curves for patients with T3-T4. **(E-I)** The sensitivity of high and low m6Ascore to immunotherapy.

Integrated analysis of Single-cell and m6Ascore

Single-cell sequencing data belonging to CC were obtained, and 15 clusters based on 13447 cells were identified by Seurat QC and other standard processes for subsequent analysis (Figure 7A). The clusters included cluster0 (1629 cells), cluster1 (1562 cells), cluster2 (1376 cells), cluster3 (1311 cells), cluster4 (1168 cells), cluster5 (1027 cells), cluster6 (999 cells), cluster7 (975 cells), cluster8 (858 cells), cluster9 (771 cells), cluster10 (678 cells), cluster11 (495 cells), cluster12 (378 cells), cluster13 (166 cells), cluster12 (54 cells). We then identified 12 cell types with the help of singleR tool and manual annotation (Figure 7B). The 12 cell types were named as Conventional T (2307 cells), Regulatory T (1562 cells), Monocytes (3402 cells), CD8$^+$ T (1311 cells), NK (1027 cells), Cancer (999 cells), NKT (975 cells), B (771 cells), Monocyte-derived macrophages (495 cells), Epithelial (378 cells), Plasmacytoid dendritic (166 cells), and Endothelial (54 cells). In addition, we also scored m6A-related genes with the Seurat R package to cell types. The violin (Figure 7C) and UMAP (Figure 7D) diagrams showed that m6Ascore was different in each cell type.

Crosstalk between m6Ascore and single cell

Utilizing the CellChat algorithm, the impact of m6A-related genes on cell–cell communication was investigated by dividing Epithelial and Tres cells into m6A_high, m6A_median, and m6A_low subgroups (Figure 8A-B). Individual signaling pathways or ligand-receptor-mediated cellular interactions are shown in Figure 8C-F and Supplemental Table S4.

PFKFB4 Expression

Given the significant prognostic values of the m6A score in CESC, PFKFB4 with the biggest HR values was regarded as the key gene by using the cox-regression analysis for CESC (Figure 9A). Then, we found that PFKFB4 had a significant difference in KM curves between high PFKFB4 and low PFKFB4 groups in the GPEIA2 (Figure

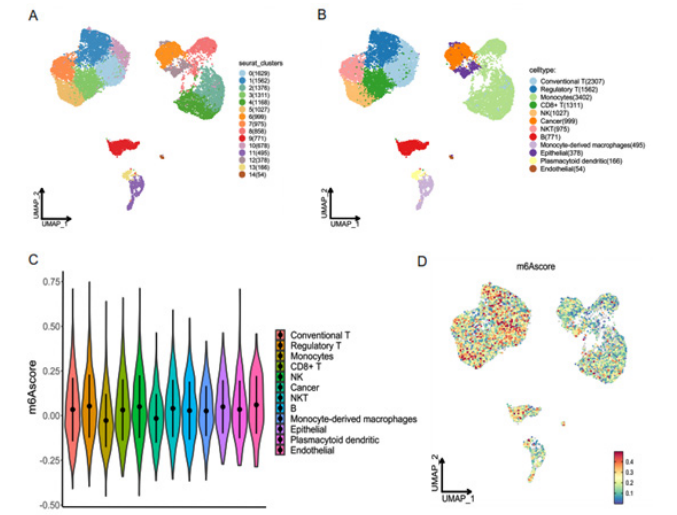

Figure 7. Integrated analysis of Single-cell and m6Ascore. **(A)** Single-cell sequencing data belonging to CC identifying 15 clusters. **(B)** 12 cell types were identified using the singleR tool and manual annotation. **(C)** The violin diagram shows the level of m6Ascore in each cell type. **(D)** UMAP diagram showing the level of m6Ascore in each cell type.

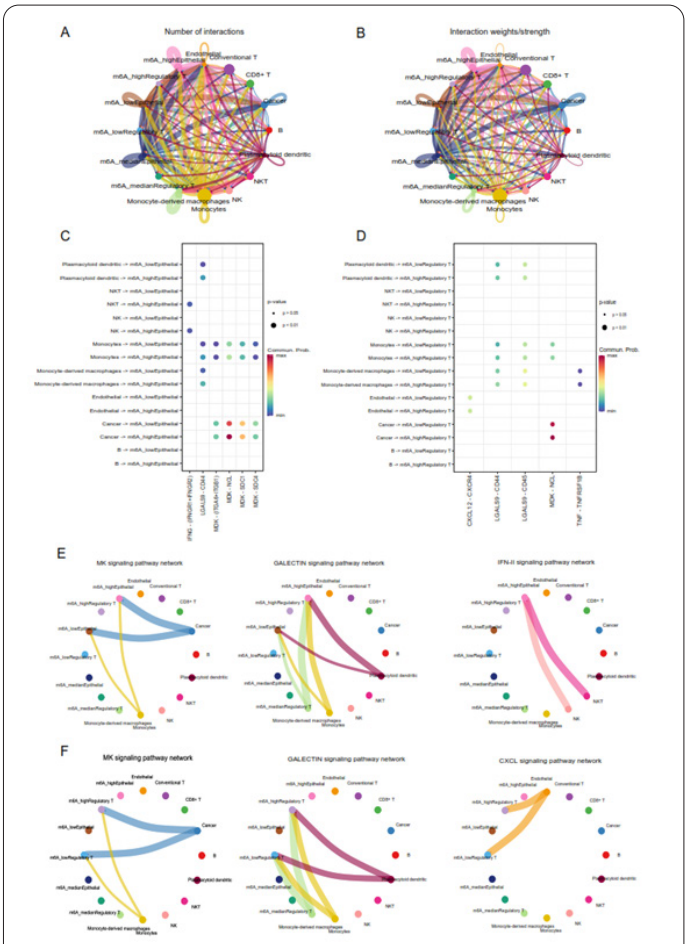

Figure 8. Crosstalk between m6Ascore and single cell. **(A)** Number of interactions for Cell–cell interaction in cell clusters analyzed by Cell Chat. **(B)** Interaction weights/strength for Cell–cell interaction in cell clusters analyzed by Cell Chat. **(C-D)** Connection probability of main signaling pathways in cell clusters. **(E-F)** Cell–cell interaction in cell clusters analyzed by cellphonedb.

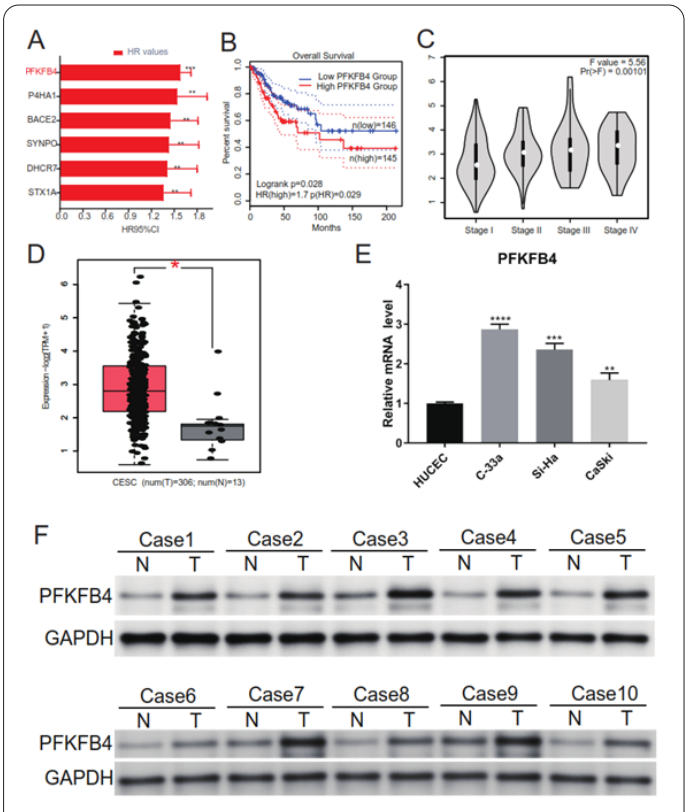

Figure 9. Validation of the key gene of m6A score in public and in-house data.

9B) and a significant relationship with TNM (Figure 9C). PFKFB4 expression was higher in tumor tissues than that in the normal tissue in TCGA-CESC (Figure 9D). In validation, we examined the mRNA level of PFKFB4 in cervical cancer cell lines compared with HUCEC cells, and we found that the mRNA expression of PFKFB4 was significantly increased in cervical cancer cell lines compared with HUCEC cells (Figure. 9E). We also examined the expression of PFKFB4 at the protein level by Western blot. As shown in Figure 9F, PFKFB4 expression was dramatically increased in most of cervical cancer tissues when compared to their normal adjacent tissues.

Discussion

Despite the continuous improvement of medical technology, cancer is still an incurable disease. At present, the main treatment methods for malignant tumors are surgery, radiotherapy, chemotherapy and targeting (41). However, after the initial treatment, patients will still have a recurrence, metastasis, and then lead to the failure of treatment (42). Prognostic factors, such as progression-free time and survival, are critical for tumor treatment. The survival time of tumor patients is often related to the malignant degree and clinical stage of the tumor (43). Accurately judging the survival time of tumor patients can indirectly judge the degree of malignancy of the tumor, and develop effective treatment strategies based on this (44). For patients with better prognoses, follow-up can be strengthened to reduce unnecessary chemoradiotherapy or targeted therapy and immunotherapy, while for patients with poor prognoses, active complementary therapy should be given to finally benefit the patients (44). It is of great significance to predict the survival time after the first diagnosis. The survival time of cancer patients largely depends on the degree of malignancy of cancer cells (45). With the deepening understanding of disease, the molecular characteristics can more accurately predict the degree of malignancy and disease progression of tumor, which is helpful for intervention, reducing recurrence and enhancing the ability of disease management (46). According to the degree of tumor malignancy, the selection of appropriate treatment is very important for tumor management and prevention of recurrence and can also avoid the occurrence of overtreatment. Therefore, it is of great significance to accurately predict the survival time of cancer patients.

At present, clinical features are the main prediction parameters for the survival of cancer patients, including pathological classification, stage, tumor size, lymph node metastasis, etc. Clinical features have the characteristics of long-term accumulation and are easy to extract (47). However, the clinical features are mostly the surface features of tumor cells, which is not enough to distinguish the internal heterogeneity of tumor cells (48). The same pathological type and stage often have different survival times. The reason for this is that in addition to the treatment methods, the heterogeneity of the tumor itself is also found. The molecular characteristics of tumors reflect the internal mechanism of tumor cells and are an important basis for distinguishing the intrinsic heterogeneity of tumor cells (49,50). Therefore, molecular features are more suitable to be used as indicators to construct prediction models for the survival of cancer patients. TCGA is currently recognized as the leading cancer research database (8). With the development of genomics, transcriptomics, proteomics, Genome-wide Association Study (GWAS), TCGA and other projects (24,51) the layers of molecular biology and genomics of diseases including cancer have been gradually unveiled (52). The relationship between molecular and genetic changes and human diseases has been intensively studied, and the research results in this field have been applied to various clinical fields (53). In the field of cancer, compared with clinical indicators, the information content of molecular features is huge. It is of great significance to effectively mine the useful information of these molecular features for clinical research and basic research (54). In conclusion, molecular features can better reflect the nature and intrinsic heterogeneity of tumor cells than clinical features and are better features for predicting the prognosis and survival of tumor patients.

With the advent of the era of big data, the size of the original sample data is very large. To find the rules behind these data by using the traditional single algorithm, the calculation amount will be very huge, and the operation process is very complex, the cost cannot be underestimated, and the reliability of the obtained results cannot guarantee integrity and accuracy (55). Compared with traditional statistics, machine learning has the ability to quickly extract and analyze a large number of different types of data (56). With the continuous mining of massive information, machine learning can achieve better prediction results than traditional statistical methods in the case of data with high latitude and multicollinearity (57). When dealing with large samples of data, machine learning has non-negligible advantages (58). The molecular information of cancer is also very complex data (59). The molecular information of cancer can be roughly divided into gene mutation, chromatin variation, copy number variation, DNA modification variation, gene expression and so on (60,61). Among them, there are more than 20,000 human genes, with countless types and numbers of mutations and complex methylation changes (62,63). To sum up, the molecular information of cancer is a huge and complex multidimensional information, which needs to be analyzed by using the advantages of machine learning.

The classification of cancer is closely related to the diagnosis, treatment and prognosis of cancer. Through the classification of cancer, we can provide personalized medicine for patients belonging to different cancer subtypes. The classification of cancer also plays an important role in the design and selection of anticancer drugs (64). The origin of tumor cells can influence but not completely determine the cell classification. The results of cancer classification based on molecular biology and pathological classification based on organs and tissues converge, but there are still some differences (53). Cancer is a disease caused by gene mutations or changes. Studying the classification of tumors from the perspective of molecular biology, such as gene expression, epigenetics, and gene mutations, will have a more fundamental understanding of tumors.

At present, the global high incidence of CC is still facing huge challenges, still has the risk of recurrence or metastasis after the initial cure, but in addition to the clinical information of clinical stage and pathological type, tumor inherent characteristic of biological information whether the prognosis of patients with CC can help its treatment decisions, it remains to be further exploration, to be sure is accurate to predict the prognosis of CC patients. It is of

great significance to improve the treatment and management level of CC patients. Common clinical indicators, to accurately predict the prognosis of patients with cervical cancer survival condition to provide enough useful information, therefore, we carried out this study.

Abbreviations

CC: cervical cancer; TCGA: The Cancer Genome Atlas; WGCNA: weighted correlation network analysis; ssGSEA: single-sample gene set enrichment analysis; NCI: National Cancer Institute; NHGRH: National Human Genome Research Institute; m6A: N6-methyladenosine
TOM: topological overlap matrix; ME: module Eigen gene; GS: gene significance; MM: module membership; TME: tumor microenvironment; NES: enrichment score; TCIA: The Cancer Immunome Atlas; GSEA: Gene Set Enrichment Analysis; MsigDB: molecular signature database; GO: Gene Ontology

Acknowledgments

We gratefully acknowledge the contributions of all individuals and organizations involved in this study.

Interest conflicts

The authors declare that they have no conflicts of interest.

Authors' contribution

CH, XS, and QC contributed to the conception and design of the study, as well as the analysis and interpretation of the data. CH played a key role in writing the manuscript, while XS and QC provided critical input during the manuscript review process. YP and FS were involved in the data analysis, interpretation, and revision of the manuscript. YW provided guidance on the study design, data interpretation, and manuscript review. SQ assisted in data collection, analysis, and interpretation, and contributed to the preparation of the manuscript. Collectively, all authors made significant contributions to the study and manuscript preparation.

Funding

This study was supported by grants from the National Natural Science Foundation of China (NSFC) Youth Program (81901596), Natural Science Foundation of Zhejiang Province (LGF22H080009), Health Bureau Foundation of Zhejiang Province (2020KY015, 2022KY017, 2022KY044 and 2023KY479), Chinese Traditional Medicine Administration Bureau Foundation of Zhejiang Province (2022ZQ012 and 2023ZL269), and Education Foundation of Zhejiang Province (Y202146099).

Data Availability

The RNA sequencing profiles are able to be gained from The Cancer Genome Atlas (TCGA) (https://toil.xenahubs.net). Further inquiries can be directed to the corresponding author.

References

1. Small W, Bacon MA, Bajaj A, Chuang LT, Fisher BJ, Harkenrider MM, et al. Cervical cancer: A global health crisis. Cancer 2017;123:2404–12.
2. Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, et al. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. Lancet Glob Heal 2020;8:e191–203.
3. Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. Lancet 2019;393:169–82.
4. Narayan K, Fisher R, Bernshaw D. Significance of tumor volume and corpus uteri invasion in cervical cancer patients treated by radiotherapy. Int J Gynecol Cancer 2006;16:623–30.
5. Moore DH. Cervical cancer. Obstet Gynecol 2006;107:1152–61.
6. Gadducci A, Guerrieri ME, Greco C. Tissue biomarkers as prognostic variables of cervical cancer. Crit Rev Oncol Hematol 2013;86:104–29.
7. Rego SM, Snyder MP. High throughput sequencing and assessing disease risk. Cold Spring Harb Perspect Med 2019;9:a026849.
8. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. Wspolczesna Onkol 2015;1A:A68–77.
9. Goodwin S, McPherson JD, McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. Nat Rev Genet 2016;17:333–51.
10. Mao X, Qin X, Li L, Zhou J, Zhou M, Li X, et al. A 15-long noncoding RNA signature to improve prognosis prediction of cervical squamous cell carcinoma. Gynecol Oncol 2018;149:181–7.
11. Liang B, Li Y, Wang T. A three miRNAs signature predicts survival in cervical cancer using bioinformatics analysis. Sci Rep 2017;7:1–8.
12. Li X, Tian R, Gao H, Yang Y, Williams BRG, Gantier MP, et al. Identification of a histone family gene signature for predicting the prognosis of cervical cancer patients. Sci Rep 2017;7:1–13.
13. Lee S, Sahasrabuddhe V V., Mendoza-Cervantes D, Zhao R, Duggan MA. Tissue-based Immunohistochemical Biomarker Expression in Malignant Glandular Lesions of the Uterine Cervix: A Systematic Review. Int J Gynecol Pathol 2018;37:128–40.
14. Mente S, Kuhn M. The Use of the R Language for Medicinal Chemistry Applications. Curr Top Med Chem 2020;12:1957–64.
15. Chan BKC. Data analysis using R programming. Adv Exp Med Biol 2018;1082:47–122.
16. Delaunay S, Frye M. RNA modifications regulating cell fate in cancer. Nat Cell Biol 2019;21:552–9.
17. Huang H, Weng H, Chen J. The Biogenesis and Precise Control of RNA m6A Methylation. Trends Genet 2020;36:44–52.
18. Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, et al. N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol 2011;7:885–7.
19. Pinello N, Sun S, Wong JJL. Aberrant expression of enzymes regulating m 6 A mRNA methylation: implication in cancer. Cancer Biol Med 2018;15:323–34.
20. Lan Q, Liu PY, Haase J, Bell JL, Hüttelmaier S, Liu T. The critical role of RNA m6A methylation in cancer. Cancer Res 2019;79:1285–92.
21. Chen XY, Zhang J, Zhu JS. The role of m6A RNA methylation in human cancer. Mol Cancer 2019;18:1–9.
22. Zhao W, Qi X, Liu L, Liu Z, Ma S, Wu J. Epigenetic Regulation of m6A Modifications in Human Cancer. Mol Ther - Nucleic Acids 2020;19:405–12.
23. Lee Y, Choe J, Park OH, Kim YK. Molecular Mechanisms Driving mRNA Degradation by m6A Modification. Trends Genet 2020;36:177–88.
24. Wang Z, Jensen MA, Zenklusen JC. A practical guide to The Cancer Genome Atlas (TCGA). Methods Mol Biol 2016;1418:111–41.
25. Fattahi F, Kiani J, Khosravi M, Vafaei S, Mohammadi A, Madjd Z, et al. Enrichment of Up-regulated and Down-regulated Gene Clusters Using Gene Ontology, miRNAs and lncRNAs in Colorectal Cancer. Comb Chem High Throughput Screen 2019;22:534–45.
26. Jin Y, Wang Z, He D, Zhu Y, Hu X, Gong L, et al. Analysis of

m6A-Related Signatures in the Tumor Immune Microenvironment and Identification of Clinical Prognostic Regulators in Adrenocortical Carcinoma. Front Immunol 2021;12:637933.
27. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2009;26:139–40.
28. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 2005;4.
29. Giulietti M, Occhipinti G, Principato G, Piva F. Weighted gene co-expression network analysis reveals key genes involved in pancreatic ductal adenocarcinoma development. Cell Oncol 2016;39:379–88.
30. Ye Y, Guo J, Xiao P, Ning J, Zhang R, Liu P, et al. Macrophages-induced long noncoding RNA H19 up-regulation triggers and activates the miR-193b/MAPK1 axis and promotes cell aggressiveness in hepatocellular carcinoma. Cancer Lett 2020;469:310–22.
31. Wu H, Chen S, Yu J, Li Y, Zhang X yan, Yang L, et al. Single-cell Transcriptome Analyses Reveal Molecular Signals to Intrinsic and Acquired Paclitaxel Resistance in Esophageal Squamous Cancer Cells. Cancer Lett 2018;420:156–67.
32. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:1–13.
33. Seiler M, Huang CC, Szalma S, Bhanot G. ConsensusCluster: A software tool for unsupervised cluster discovery in numerical data. Omi A J Integr Biol 2010;14:109–13.
34. Hänzelmann S, Castelo R, Guinney J. GSVA: Gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics 2013;14:1–15.
35. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep 2017;18:248–62.
36. Yi M, Nissley D V., McCormick F, Stephens RM. ssGSEA score-based Ras dependency indexes derived from gene expression data reveal potential Ras addiction mechanisms with possible clinical implications. Sci Rep 2020;10:1–16.
37. 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;102:15545–50.
38. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R package for comparing biological themes among gene clusters. Omi A J Integr Biol 2012;16:284–7.
39. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011;27:1739–40.
40. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. Genome Res 2018;28:1747–56.
41. Meir H, Kenter G, Burggraaf J, Kroep J, Welters M, Melief C, et al. The Need for Improvement of the Treatment of Advanced and Metastatic Cervical Cancer, the Rationale for Combined Chemo-Immunotherapy. Anticancer Agents Med Chem 2014;14:190–203.
42. Huang H, Feng YL, Wan T, Zhang YN, Cao XP, Huang YW, et al. Effectiveness of Sequential Chemoradiation vs Concurrent Chemoradiation or Radiation Alone in Adjuvant Treatment after Hysterectomy for Cervical Cancer: The STARS Phase 3 Randomized Clinical Trial. JAMA Oncol 2021;7:361–9.
43. Knoth J, Pötter R, Jürgenliemk-Schulz IM, Haie-Meder C, Fokdal L, Sturdza A, et al. Clinical and imaging findings in cervical cancer and their impact on FIGO and TNM staging – An analysis from the EMBRACE study. Gynecol Oncol 2020;159:136–41.
44. Yankeelov TE, Quaranta V, Evans KJ, Rericha EC. Toward a science of tumor forecasting for clinical oncology. Cancer Res 2015;75:918–23.
45. Paliouras AR, Monteverde T, Garofalo M. Oncogene-induced regulation of microRNA expression: Implications for cancer initiation, progression and therapy. Cancer Lett 2018;421:152–60.
46. Ben-Hamo R, Jacob Berger A, Gavert N, Miller M, Pines G, Oren R, et al. Predicting and affecting response to cancer therapy based on pathway-level biomarkers. Nat Commun 2020;11:3296.
47. Ryzhov A, Corbex M, Piñeros M, Barchuk A, Andreasyan D, Djanklich S, et al. Comparison of breast cancer and cervical cancer stage distributions in ten newly independent states of the former Soviet Union: a population-based study. Lancet Oncol 2021;22:361–9.
48. Fang J, Zhang H, Jin S. Epigenetics and cervical cancer: From pathogenesis to therapy. Tumor Biol 2014;35:5083–93.
49. TCGAN. Integrated genomic and molecular characterization of cervical cancer The Cancer Genome Atlas Research Network *. Nature 2017;543:378–84.
50. Vafaei S, Saeednejad Zanjani L, Habibi Shams Z, Naseri M, Fattahi F, Gheytanchi E, et al. Low expression of Talin1 is associated with advanced pathological features in colorectal cancer patients. Sci Rep 2020;10:1–18.
51. Bakhoum MF, Esmaeli B. Molecular characteristics of uveal melanoma: Insights from the cancer genome atlas (TCGA) project. Cancers (Basel) 2019;11:1061.
52. Nair M, Sandhu SS, Sharma AK. Cancer molecular markers: A guide to cancer detection and management. Semin. Cancer Biol., vol. 52, Elsevier; 2018, p. 39–55.
53. Yau C, , Toshinori Hinoue DMW, Taylor AM, Cherniack AD, Wiznerowicz M, Sanchez-Vega F, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell 2018;173:291–304.
54. Matthew H. Bailey, Collin Tokheim, Eduard Porta-Pardo, ..., Gordon B. Mills, Rachel Karchin, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell 2018;173:371–85.
55. Schlick T, Portillo-Ledesma S. Biomolecular modeling thrives in the age of technology. Nat Comput Sci 2021;1:321–31.
56. Kantarjian H, Yu PP. Artificial intelligence, big data, and cancer. JAMA Oncol 2015;1:573–4.
57. Mortazavi BJ, Downing NS, Bucholz EM, Dharmarajan K, Manhapra A, Li SX, et al. Analysis of Machine Learning Techniques for Heart Failure Readmissions. Circ Cardiovasc Qual Outcomes 2016;9:629–40.
58. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: A review. Biotechnol Adv 2021;49:107739.
59. Sahlolbei M, Fattahi F, Vafaei S, Rajabzadeh R, Shiralipour A, Madjd Z, et al. Relationship Between Low Expressions of tRNA-Derived Fragments with Metastatic Behavior of Colorectal Cancer. J Gastrointest Cancer 2022;53:862–9.
60. Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, et al. The chromatin accessibility landscape of primary human cancers. Science (80-) 2018;362:eaav1898.
61. Nagy M, Radakovich N, Nazha A. Machine Learning in Oncology: What Should Clinicians Know? JCO Clin Cancer Informatics 2020;4:799–810.
62. McCullough A. Comprehensive molecular portraits of human breast tumours. vol. 2013. 2013.
63. Zhang H, Luo YB, Wu W, Zhang L, Wang Z, Dai Z, et al. The molecular feature of macrophages in tumor immune microenvironment of glioma patients. Comput Struct Biotechnol J 2021;19:4603–18.
64. Berman J. Modern classification of neoplasms: Reconciling dif-

ferences between morphologic and molecular approaches. BMC Cancer 2005;5:1–12.