N6-methylandenosine-related gene expression signatures for predicting the overall survival and immune responses of patients with colorectal cancer

Title: N6-methylandenosine-related gene expression signatures for predicting the overall survival and immune responses of patients with colorectal cancer

Lili Yu[1], Lijuan Wang[1], Jing Sun[1], Xuan Zhou[1], Yeting Hu[2], Lidan Hu[3], Yazhou He[4], Chunqing Lin[5], Jie Chen[6], Kefeng Ding[2,7†], Xue Li[1†], Evropi Theodoratou[8,9*], Malcolm Dunlop[8*]

[1] Department of Big Data in Health Science School of Public Health, Center of Clinical Big Data and Analytics of The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[2] Department of Colorectal Surgery and Oncology, Key Laboratory of Cancer Prevention and Intervention, Ministry of Education, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[3] The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China

[4] Department of oncology, West China School of Public Health and West China Fourth Hospital, Sichuan University, Sichuan, China

[5]National Cancer Center, National Clinical Research Center for Cancer, and Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

[6] Center for Global Health, Zhejiang University, Hangzhou, China

[7] Cancer Center, Zhejiang University, Hangzhou, China

[8] Cancer Research UK Edinburgh Centre, Medical Research Council Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK

[9] Centre for Global Health, Usher Institute, University of Edinburgh, Edinburgh, UK

[*]Joint last authors.

[†]Correspondence to:
Xue Li, School of Public Health and the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China, Email: xueli157@zju.edu.cn
Kefeng Ding, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China, Email: dingkefeng@zju.edu.cn

Abstract

N6-methyladenosine (m6A) modification has been demonstrated to exhibit a crucial prognostic effect on colorectal cancer. Nonetheless, potential mechanism of m6A in survival rate and immunotherapeutic response remains unknown. This study aimed to investigate the genes associated with m6A regulators and to develop a risk score for predicting the overall survival (OS) of CRC patients. RNA-seq transcriptomic profiling data of COAD/READ samples and mutation data were obtained from The Cancer Genome Atlas (TCGA). Multivariate Cox regression analysis was conducted to identify the m6A-related gene expression signatures associated with CRC survival. An m6A-related prognostic risk score was developed in TCGA dataset and its predictive performance of CRC survival was further validated in Gene Expression Omnibus (GEO) datasets. It was shown that a risk score comprising 18 m6A-related mRNAs in combination with clinical characteristics yielded C-statistics of 0.85 (95%CI: 0.79-0.91), 0.84 (95%CI: 0.79-0.90) and 0.80 (95%CI: 0.71-0.88) for the prediction of the 1-, 3-, 5-year OS of CRC in TCGA cohort. We further used this risk score as classifier to investigate the molecular characteristics, immune microenvironment, and likelihood of response to immunotherapy between the low- and high-risk groups. The mutations of oncogenes occurred more frequently in the high-risk group and the composition of immune cells in tumour microenvironment (TME) was significantly distinct between the low- and high-risk groups. The low-risk group had a lower microsatellite instability (MSI) score, T-cell exclusion score and dysfunction score, implying that low-risk patients may have a better immunotherapy response than high-risk patients. In summary, a prognostic risk score derived from m6A-related gene expression signatures could serve as a potential prognostic predictor for CRC survival and indicator for predicting immunotherapy response in CRC patients.

Introduction

Colorectal cancer (CRC) is the third most common cancer and a leading cause of cancer mortality worldwide.[1] Although the survival time of CRC patients has been significantly extended by clinical treatment, the 5-year OS of CRC patients is still not ideal, with a rate of approximate xx-68%. Presently, emerging evidence has shown that the discovery and application of molecular biomarkers may provide important clinical implications on the prognosis and treatment of CRC patients.[2]

N^6-methyladenosine (m6A) is one of the most prominent and abundant forms of internal RNA modification affecting RNA stability and translational efficiency.[3, 4] This modification is a dynamic reversible process in mammalian cells regulated by methyltransferases, demethylases, and binding proteins, which are also known as "writers", "erasers", and "readers".[5] In-depth understanding of these regulators would help reveal the role and mechanism of m6A modification in post-transcriptional regulation. To date, accumulating evidence demonstrated that dysregulated m6A methylation modification is correlated with disorders of multiple biological processes including dysregulate cell death and proliferation, tumour malignant progression, and immunomodulatory abnormality,[6] thus could be closely associated with a variety of human diseases, in particular cancer.[7] For instance, a recent study reported that METTL3, one of m6A regulators, directly induced m6A-GLUT1-mTORC1 axis to promote CRC development.[8] Additionally, it is shown that the dysregulated expression of *YTHDF2* can restrain cell proliferation by reducing the mRNA stability of EGFR in liver cancer.[9] However, the specific role of m6A regulators in the dysregulation of mRNAs, and how m6A modifications contribute to CRC prognosis remains unclear.

The tumour microenvironment (TME), which is where the tumour is located and which is composed of various cancer cells, stromal cells, and distinct recruited cells (infiltrating immune cells, bone marrow-derived cells), plays a crucial role in tumour progression and affects the clinical benefit from novel strategies of immunological checkpoint blockade (ICB).[10, 11] Emerging studies have made efforts to understand the heterogeneity and complexity of the TME to improve immunotherapy strategies by comprehensive analysis of m6A regulators.[12]

Predicting the immunotherapy response of CRC patients based on multiple m6A-related biomarkers has the potential to develop a personalised treatment strategy and therefore to increase the success of ICB.[13-15]

In this study, we sought to elucidate the m6A related mRNAs signatures for predicting the overall survival (OS) and immune responses of CRC patients using transcriptome data from The Cancer Genome Atlas (TCGA)[16] and Gene Expression Omnibus (GEO)[17, 18] datasets. We focused on the m6A-related genes and developed a multivariate Cox prediction model for the OS of CRC patients and examined its prognostic ability in immunotherapy response. We additionally explored the candidate drugs targeting these m6A-related gene signatures using the publicly available Genomics of Drug Sensitivity in Cancer (GDSC) database for predicting drug sensitivity[19]. Findings from this study are helpful to predict the prognosis of CRC and develop personalized CRC treatment strategies.

Materials and Methods

Study population and datasets

A study sample of 644 CRC patients from the TCGA was used as a training dataset. RNA-seq [Fragments Per Kilobase of transcript per Million mapped reads (FPKM normalized)] were acquired from Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/) using the R package "TCGAbiolinks", which was specifically developed for integrative analysis with Genetic Data Commons (GDC) data[20]. Then FPKM values were transformed into transcripts per kilobase million (TPM) values. The corresponding clinicopathological information and somatic mutation data of CRC patients were obtained from the cBioPortal database (https://portal.gdc.cancer.gov/). Two study samples (GSE39582, N=566; GSE17536, N=177) from the GEO database were used as validation datasets, and their normalized microarray gene expression data and clinicopathological data were obtained online (https://www.ncbi.nlm.nih.gov/geo/). Those RNA probe sets were re-annotated using the Ensemble database (http://www.ensembl.org). CRC patients with missing OS values or OS < 30 days were excluded in order to reduce statistical bias in this analysis.

Identification of m6A-related prognostic genes

The expression matrices of 21 m6A regulators were retrieved from the TCGA, including the expression data of eight writers (*METTL3, METTL14, METTL16, RBMX2, RBM15B, WTAP, KIAA1429*, and *ZC3H13*), two erasers (*FTO* and *ALKBH5*), and eleven readers (*YTHDC1, YTHDC2, YTHDF1, YTHDF2, YTHDF3, IGF2BP1, EMR1, LRPPRC, HNRNPA2B1, HNRNPC,* and *ELAVL1*). Based on the RNA-seq data, Pearson's correlation analysis was firstly implemented to identify m6A-related genes, using the criteria of |Pearson R| >0.3 and p <0.001. Univariable and multivariable Cox regression models (false discovery rate, FDR<0.05) and the least absolute shrinkage and selection operator (LASSO) Cox regression (with the penalty parameter estimated by 5-fold cross-validation) were conducted subsequently to select the m6A-related prognostic genes that were distinctly related to the OS of CRC patients.

Development and validation of the genetic risk signatures

A weighted prognostic risk score of m6A-related gene expression was constructed based on the following formula: Risk score = $\sum_{i=1}^{n} \mathrm{Coef}\ (\mathrm{Genei}) \times \mathrm{Expr}(\mathrm{Genei})$, where Coef ($Gene_i$) was the coefficient of genes correlated with CRC survival, and Expr ($Gene_i$) was the expression of genes. The prognostic value of the risk score was evaluated by Kaplan-Meier survival curves with log-rank tests in both TCGA and GEO study samples. Multivariate Cox regression analysis was performed to evaluate the prediction performance of the m6A-related gene signatures prognostic risk score. Patients with CRC were further stratified into low- and high-risk groups based on the median value of the prognostic risk score of m6A-related genes.

Analysis of the molecular characteristics in the low- and high-risk groups

To explore the biological function and alternative pathways of these m6A-related gene signatures, we performed a co-expression and pathway enrichment analysis based on the TCGA database, using the Kyoto Encyclopaedia of Genes and Genomes Pathway (KEGG pathway) as reference[21]. Linear regression was performed to detect co-expressed genes (FDR<0.05). In the gene mutation analysis, information on genetic alterations was obtained from the cBioPortal database. The quantity and quality of gene mutations were analysed in low- and high-risk groups by using the Maftools package in R.

Exploration of immunotherapeutic response between low- and high-risk groups

To depict immune characteristics of CRC patients, the entire expression data were imported into CIBERSORT (https://cibersort.stanford.edu/) and a deconvolution algorithm using support vector regression was used and iterated 1,000 times to determine the relative proportions of 22 immune cell types in tumours. The relative proportions of immune cell types and clinicopathologic factors were compared between the low- and high-risk groups. The tumour Immune Dysfunction and Exclusion (TIDE) score was calculated online (http://tide.dfci.harvard.edu/) to predict the likelihood of immunotherapeutic response between the low- and high-risk groups.

Prediction of potential compounds targeting therapeutic sensitivity in CRC patients

To obtain potential compounds with differential therapeutic sensitivity, we investigated the predictive capacity of the low- and high-risk groups in responding immunotherapy. The 50% inhibiting concentration half-maximal inhibitory concentration (IC_{50}) value of 138 anti-cancer drugs was inferred from the GDSC website based on the COAD/READ dataset of the TCGA project. The "pRRophetic" algorithm[22] was used to predict the IC_{50} of compounds in the low- and high-risk groups separately.

Statistical analysis

An independent *t*-test was performed to compare continuous variables between two groups. Categorical data were tested using the χ^2 test. Pearson correlation analysis was implemented to identify m6A-related genes (with the | Pearson r | >0.05 and p<0.001). Univariate survival analysis was performed by K-M survival analysis with the log-rank test to calculate the significance of differences in the OS. Multivariate survival analysis was performed using the Cox regression model to estimate the hazard ratio (HR). The time-dependent area under the receiver operating characteristic curve (AUC) was estimated to evaluate the predictive power of the risk score and TNM stage to the OS. Stratification analysis was performed to investigate the survival difference in subgroups, including age, sex, American Joint Committee on Cancer (AJCC) TNM stage, T stage, N stage, M stage, Radiation therapy history. A nomogram of the

risk score and other predictors was set up accordingly for the prediction of the 1-, 3-, 5- year OS. Unless otherwise stated, the *P* values were two-sided and $P<0.05$ was considered as statistically significant.

Results

Landscape of genetic variation of m6A regulators in CRC patients

A total of 21 m6A regulators, namely 8 "writers", 2 "erasers", and 11 "readers", were included in this study. We firstly assessed the prevalence of somatic mutations and copy number variations (CNV) of these 21 m6A regulators. Among the 551 samples, 169 (30.67%) had mutations in any of the m6A modification regulators (**Fig. 1A**). It was found that *ZC3H13* exhibited the highest mutation frequency (23%) followed by *KIAA1429* (18%) and *YTHDC2* (15%), while demethylases *ALKBH5* (2%) and *WTAP* (3%) showed low number of mutations in CRC samples. We then examined the somatic copy number alterations of these m6A regulators and found that *METTL14* (34%), *METTL16* (56%), *ALKBH5* (58%) and *YTHDF2* (38%) had a widespread frequency of CNV deletions (**Fig. 1B** and **Fig. 1C**). To ascertain whether the above genetic variations influenced the expression of m6A regulators in CRC patients, we investigated the mRNA alterations of the m6A regulators between paired normal and tumour samples of CRC patients. This showed that alterations of CNV were prominent factors resulting in perturbations on the m6A regulators expression. Compared to the normal colon tissues, regulators with CNV gain demonstrated markedly higher expression in CRC tissues (e.g., *YTHDF1* and *KIAA1429*) (**Fig. 1B and Supplementary Fig. S1**). And vice versa, some regulators showed downregulated mRNA expression but with high frequency of CNV loss (e.g., *ALKBH5*). This analysis presented the high heterogeneity of genetic and expressional alteration landscape of m6A regulators between normal and tumour samples, demonstrating that the expression imbalance of m6A regulators may be important in the onset and progression of CRC.

Identification of m6A-related genes in patients with CRC

A total of 551 COAD/READ patients from the TCGA database were included in our study to

calculate the prognostic risk score of m6A-related genes. The detailed workflow for risk model construction and subsequent analyses is shown in **Fig. 2**. We abstracted the matrix expression of 21 m6A regulators and 19,982 mRNAs from the TCGA database. Correlations between these 21 m6A regulators and 19,982 mRNAs were examined and we identified 4274 mRNAs that were significantly correlated with m6A regulators base on the criteria of |Pearson R|>0.5 and p<0.001. To identify m6A-related genes that correlated with the OS of CRC patients, we screened from 4274 m6A-associated mRNAs in the TCGA training set using univariate Cox regression analysis. At FDR<0.05, fifty-seven m6A-related mRNAs correlated significantly with OS (**Supplementary Table S1**).

Construction of the prognostic risk score based on m6A-related gene expression signatures

To avoid overfitting, the LASSO-Cox regression was applied to optimise the selection of gene signatures in relation to the OS. Consequently, 18 m6A-related mRNAs (*PMM2*, *ERI1*, *NEK9*, *USP53*, *CNOT3*, *CDK5RAP2*, *ING5*, *HMGXB4*, *SH3D19*, *UBE2H*, *CLK1*, *SFPQ*, *UBP1*, *PDCD6IP*, *ZNF248*, *SCL25A53*, *CLCC1* and *GPR125*) were finally selected to construct a m6A-related gene signatures prognostic risk score for CRC survival (**Supplementary Fig. S2A and S2B**). The correlation between m6A regulators and m6A-related gene signatures in the TCGA dataset is showed in **Supplementary Fig. S3**. CRC patients were separated into high- and low-risk groups based on the median value of the prognostic risk score constructed by the m6A-related gene expression signatures. The distribution of risk scores between the low- and high-risk groups is depicted in **Fig. 3A**, and the survival status and survival time of CRC patients in the low- and high-risk groups are shown in **Fig. 3B**. The expression levels of the 18 m6A-related genes in the low- and high-risk groups are shown in **Fig. 3C**. Kaplan–Meier survival curves showed that CRC patients with higher risk scores had worse clinical outcomes (lower OS rates and a shorter OS time, HR=1.30, 95%CI: 1.21-1.41; P<0.0001, log-rank test) (**Fig. 3D**). PCA analysis was further conducted to test the difference between the low- and high-risk groups based on the entire gene expression profiles, 21 m6A regulators and the expression profile of the 18 m6A-related genes (**Supplementary Fig. S4A-4C**). As showed **Supplementary Fig. S4A-4B**, the gene expression profiles of the low- and high-risk groups

were differently distributed (**Supplementary Fig. S4C**).

Validation of the prognostic risk score based on m6A-related gene expression signatures

Detailed clinicopathologic characteristics of CRC patients in TCGA and GEO datasets are shown in **Supplementary Table S2** and **Supplementary Table S3**. The expression of 18 m6A-related genes was closely correlated with the OS of CRC patients as determined by K-M analysis (**Supplementary Fig. S5**). **Supplementary Fig. S6** showed the m6A-related gene signatures affecting the OS of patients with CRC. According to the subgroups classified by sex, age, AJCC TNM stage or tumour stage, the OS of the low-risk group continued to be superior to that of the high-risk group (**Supplementary Fig. S7A-7H**). To validate the prognostic capability, we calculated the risk scores for CRC patients in two GEO (GSE39582, GSE17536) datasets using the same formula. As showed in **Fig. 4A-4H**, patients stratified into the high-risk group had a significantly worse prognosis than those in the low-risk group which was consistent with the results of TCGA dataset (P=1.38e-11, log-rank test).

Molecular characteristics of the low- and high-risk groups stratified by the prognostic risk score

To demonstrate the potential mechanisms and pathways involved in the molecular heterogeneity leading to the different outcomes between the low- and high-risk groups, we performed functional enrichment analysis with annotation of KEGG gene set. We found that m6A-related gene expression signatures were differentially enriched (FDR<0.05) in the pathways related to cancer, immune response, and neural signaling between the two groups (**Supplementary Table S4**), and pathways that more than half of the gene signatures enriched in were summarized in **Supplementary Fig. S8**. When examining the somatic mutations, we found that the top 20 cancer driver genes mutated more frequently in the high-risk group than in the low-risk group (**Fig.5A-5B**), and significant co-occurrences were also observed among mutations of these genes (as shown in **Fig. 5C**).

Estimation of the tumour immune microenvironment and cancer immunotherapy response

To analyse the composition of immune cells in different risk groups, we used the Wilcoxon test to compare the distribution of immune cells. As shown in **Fig. 6A,** we found that CD8 T cells, Tregs regulatory T cells, resting natural killer (NK) cells, and M0 macrophages were more abundant in the high-risk group, while plasma cells, resting memory CD4 T cells, activated memory CD4 T cells and M2 macrophages were more abundant in the low-risk group. We next investigated the correlations between the m6A-related signature model and immunotherapeutic biomarkers. Higher TIDE prediction score represented a higher potential for immune evasion, which suggested that the patients were less likely to benefit from Immune checkpoint inhibitor (ICI) therapy. In our results, the low-risk group had a lower TIDE score than the high-risk group, implying that low-risk patients may have a better immunotherapy response than high-risk patients (**Fig. 6B-6E**). Also, we found that the high-risk group had a higher microsatellite instability (MSI) score, T-cell exclusion score and dysfunction score. To find the potency of m6A-related prognostic score as a biomarker for predicting the response of CRC patients to drugs, we used "pRRophetic" algorithm to infer the therapeutic response based on the IC50 value of the 138 anti-cancer drugs in TCGA-COAD/READ patients. We found 50 chemotherapeutic drugs displaying differential IC50 between these two groups (**Supplementary Fig. S9)**.

Construction of nomogram based on prognostic risk score and clinical characteristics

We next investigated the distribution of the risk score of patients with CRC using different conventional clinical information (including sex, T stage, N stage, M stage and AJCC TNM stage), and found that CRC patients with higher T, N or TNM stage had a higher risk score (**Fig. 7A**). Univariate Cox analysis showed that age, radiation history, T stage, N stage and the prognostic risk score were significantly associated with the prognosis of CRC (**Fig. 7B**). Multivariate Cox analysis confirmed that the prognostic risk score based on m6A-related gene expression signatures was an independent predictor of CRC survival (**Fig. 7C**). Multivariate Cox prediction models combing prognostic risk score and clinical characteristics yielded C-statistics of 0.854 (95%CI: 0.795-0.913), 0.844 (95%CI: 0.790-0.898) and 0.796 (95%CI: 0.708-0.883) for the prediction of the 1-, 3-, 5-year OS (**Fig. 8A-8C**), which displayed superior predictive performance over the model that only included clinical characteristics with C-

statistics of 0.808 (95%CI: 0.740-0.875), 0.793 (95%CI: 0.730-0.856) and 0.755 (95%CI: 0.665-0.845). Calibration plots showed that the observed vs. predicted rates of 1-, 3-, 5-year OS had good concordance (**Fig. 8D-8F**). Accordingly, a prognostic nomogram was established based on the risk score and clinical characteristics for the prediction of OS in CRC patients as shown in **Fig. 8G**.

Discussion

Here, we developed a prognostic risk score based on m6A-related gene expression signatures and performed external validation to assess its prediction accuracy. Our study indicated that the m6A-based prognostic risk score was an independent predictor for CRC survival and had improved the prediction accuracy of CRC survival when combined with clinical characteristics. When stratified by this risk score, the high-risk group was associated with a worse survival rate, lower immunogenicity, and greater number of somatic mutations than patients in the low-risk group. Moreover, the low-risk group had a lower TIDE score than the high-risk group for predicting immunotherapy response, implying that low-risk patients could benefit more from immunotherapy than high-risk patients.

Studies have shown that m6A modification of mRNAs can affect the occurrence and development of tumours.[23] Functional enrichment analyses in this study indicated that *CDK5R4P2*, *CLK1*, *CNOT3*, *GPR125*, *ING5*, *SFPQ* and *UBE2H* are mainly involved in the neural, destabilization and metabolic processes of mRNA signatures, and influence the growth, differentiation and communications of multiple colon cell types. Our analysis suggested that metabolism signaling pathways (including key mRNAs of *CDK5RAP2*, *CNOT3*, *CLK1* and *SFPQ* genes) may contribute to tumorigenesis and cancer development. Interestingly, *GPR125* and *SFPQ* were enriched in a neural signaling pathway in relation to Spinocerebellar ataxia. Additionally, *CLK1* was reported as a novel inhibitor of CLK kinases that impairs the growth of CRC cell lines and organoids, inhibited anchorage-independent colony formation, cell migration, and promotes cytotoxicity.[24] *UBE2H* was identified as a m6A-related hub gene closely related to the clinicopathology and prognosis of CRC using a prognostic signature model.[25] In concordance with our findings, Cejas *et al* also found that *CNOT3* overexpression

in colon tissues was associated with worse prognosis outcomes of CRC.[26]

Our study firstly developed a prognostic risk score based on 18 m6A-related gene expression signatures that could be used as an index to predict the OS of CRC patients, and further validated its predictive performance in two independent external datasets. Time-dependent AUC showed that the m6A-based prognostic risk score had a good accuracy in predicting the OS of CRC patients in both the TCGA and the validation datasets. The combination of the prognostic risk score with TNM stage and age displayed superior predictive performance over the model that only included clinical characteristics. The stratified analysis also confirmed that the risk score could predict CRC survival with good performance in different clinical subgroups (age, T stage, AJCC TNM stage). Taken together, this m6A-based prognostic risk score could be used as an independent predictor for CRC survival and the application of risk score in combination with clinical characteristics could improve the prediction accuracy of CRC survival.

Using this m6A-related prognostic risk score as a classifier, CRC patients were stratified into low- and high-risk groups to gain further biological insight into the gene mutations and immunologic nature of CRC patients in different risk groups. We found that m6A-related gene expression signatures were differentially enriched in the pathways related to cancer, immune response, and neural signaling between the two groups. When examining the somatic mutations, we found that the top 20 cancer driver genes mutated more frequently in the high-risk group than in the low-risk group, and significant co-occurrences were also observed among mutations of these genes. By examining the immunologic nature of CRC patients in different risk groups, we found high-risk group generally had higher monocytes and macrophages M1 infiltration and fractions of T cells CD8, and lower memory resting CD4 T cells than low-risk patients. It has been reported that CRC patients enriched with M1 phenotype and the high islet density of M1 macrophages would have poor prognosis, which are consistent with the findings from our study. This indicates that the m6A-related gene expression signatures may modulate the TME phenotypes to influence the survival of CRC patients.

Emerging pieces of evidence showed that different TME phenotypes might have different

degrees of benefit from immunotherapeutic treatment.[27] A Tumour Immune Dysfunction and Exclusion (TIDE) score has been increasingly used as an index for predicting immunotherapeutic response.[28] Using the TIDE algorithm, we estimated the immune response and found that patients in the low-risk group have a superior response to immunotherapy. Chemotherapy results indicated that the high-risk patients with CRC were more sensitive to 24 chemotherapies than low-risk patients. These results suggested that the poorer prognosis for high-risk patients could be due to higher immunosuppression in the TME, and that TME may influence the response of chemotherapy and immunotherapy. Based on these findings, this m6A-based risk score might also be used as an indicator for predicting immunotherapy response among CRC patients.

Our study also provides insight for future studies on the process and mechanism of m6A modification of gene expression signatures. However, we are also aware of several limitations in this study. Although the m6A-related gene signatures prognostic risk score showed superior performance on the prediction of CRC survival and the response to immunotherapy, it should be prospectively validated in real clinical settings and the clinicopathological factors should also be considered. Moreover, both the TIDE and MSI scores focused on the function and status of T cells, which could not fully reflect the complexity of the TME involved in the immunotherapeutic response. Thus, future observational studies should be performed to further validate the application of this prognostic risk score in the prediction of CRC survival and to understand how these m6A-related gene expression signatures modulate the TEM and influence the response to immunotherapy.

In conclusion, we developed a prognostic risk score based on the expression signature of 18 genes associated with m6A modification to predict the OS of CRC patients and their response to immunotherapy. This work highlights the clinical implications of this risk score in distinguishing immune and molecular characteristics and identifying response of target treatments. The derived m6A-related risk score showed the potential to be used as a prognostic and therapeutic indicator for the prediction of CRC prognosis and the development of individualized CRC treatment strategy.

Acknowledgement

Consent to publication

Not applicable.

Data availability statement

All data relevant to the study are included in the article or uploaded as supplementary information.

Competing interests

No potential conflicts of interest were disclosed.

Funding information: XL is supported by the Natural Science Fund for Distinguished Young Scholars of Zhejiang Province (LR22H260001). ET is supported by a CRUK Career Development Fellowship (C31250/A22804). This work is also funded by a grant to MGD as Project Leader with the MRC Human Genetics Unit Centre Grant (U127527198).

Figure legends

Figure 1. Landscape of genetic of m6A regulators in colorectal cancer. (**A**) The mutation frequency of 21 m6A regulators in 169 patients with CRC from TCGA cohort. (**B**) Bar graphs showing the frequency of CNV gain (green), loss (blue) and non CNV (yellow) of m6A regulators in TCGA-COAD/READ cohort. (**C**) Principal component analysis for the expression profiles of 21 m6A regulators to distinguish tumors from normal samples in TCGA cohort. (**D**) Expression of 21 m6A regulators between normal tissues and CRC tissues.

Figure 2. Flow chart of this study.

Figure 3. Prognostic value of the risk patterns of the 18 m6A-related gene signatures in the TCGA training dataset. (**A**) Distribution of m6A-related gene expression model-based risk score. (**B**) Different patterns of survival status and survival time between the high- and low-risk subgroups. (**C**) Clustering analysis heatmap shows the expression standards of the 18 prognostic genes for each patient. (**D**) Kaplan-Meier survival curves of the OS of patients in the high- and low-risk subgroups.

Figure 4. Prognostic value of the risk model of the 18 m6A-related gene signatures in GSE39582 and GSE17536 dataset. (**A**) Distribution of m6A-related gene expression model-based risk score for the GSE39582. (**B**) Different patterns of survival status and survival time between the high- and low-risk subgroups for the GSE39582. (**C**) Clustering analysis heatmap shows the expression standards of the 18 prognostic genes for each patient for the GSE39582. (**D**) Kaplan-Meier survival curves of the OS of patients in the high- and low-risk subgroups for the GSE39582. (**E**) Distribution of m6A-related gene expression model-based risk score for the GSE17536. (**F**) Different patterns of survival status and survival time between the high- and low-risk subgroups for the GSE17536. (**G**) Clustering analysis heatmap shows the expression standards of the 18 prognostic genes for each patient for the GSE17536. (**H**) Kaplan-Meier survival curves of the OS of patients in the high- and low-risk subgroups for the GSE17536.

Figure 5. Molecular characteristics of different risk subgroups. (**A and B**) Waterfall plot displays tumour somatic mutation information of the genes with high mutation frequencies in

the high-risk subgroup (**A**) and low-risk subgroup (**B**). Mutated genes (rows, top 20) are ordered by mutation rate; samples (columns) are arranged to emphasize mutual exclusivity among mutations. The right shows the mutation percentage, and the top shows the overall number of mutations. The color coding indicates the mutation type. (**C**) The co-expression patterns of top 20 mutated genes in CRC patients.

Figure 6. The landscape and estimation of the tumor immune microenvironment using the m6A-related gene signatures model. (**A**) The proportions of TME cells in different risk subgroups. Significant statistical differences between the two subgroups were assessed using the Wilcoxon test, the asterisks represented the statistical p value (blank, not significant; *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001). (**B-E**) TIDE (**B**), MSI (**C**), and T-cell exclusion (**D**) and dysfunction (**E**) score in the high- and low-risk patients. The scores between the two risk subgroups were compared through the Wilcoxon test (ns, not significant; ***P < 0.001).

Figure 7. Correlation between the 18-gene expression signatures and clinical characteristics. (**A**) Difference analysis of the distribution of risk scores in different T, N, M, AJCC TNM stages, gender, and radiation history. Statistical difference of two groups was compared by the Wilcoxon test and three or more groups were compared by the Kruskal–Wallis test (*P < 0.05; **P < 0.01; ***P < 0.001; ns not significant). (**B and C**) Univariate (**B**) and multivariate (**C**) Cox regression analyses of correlations between the 18-gene expression signatures and clinical characteristics with OS, and revealed that the risk score based on the m6A-related gene expression signatures was an independent prognostic predictor in the TCGA dataset.

Figure 8. Assessment of the prognostic risk model of the m6A-related gene expression signatures and clinical features in CRC. (**A-C**) Time-dependent receiver operating characteristic (ROC) curves for the nomogram, risk score, and clinical characteristics in the TCGA dataset on predicting 1- (**A**), 3- (**B**), and 5-year (**C**) OS. (**D-F**) The calibration plot of the nomogram predicts the probability of the 1- (**D**), 3- (**E**), and 5- (**F**) year OS. (**G**) Nomogram for predicting the 1-, 3-, and 5-year OS of patients with CRC.

Reference

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**: 394-424.
2. Bramsen JB, Rasmussen MH, Ongen H, et al. Molecular-Subtype-Specific Biomarkers Improve Prediction of Prognosis in Colorectal Cancer. *Cell Rep* 2017; **19**: 1268-80.
3. Sun T, Wu R, Ming L. The role of m6A RNA methylation in cancer. *Biomed Pharmacother* 2019; **112**: 108613.
4. Niu Y, Lin Z, Wan A, et al. RNA N6-methyladenosine demethylase FTO promotes breast tumor progression through inhibiting BNIP3. *Mol Cancer* 2019; **18**: 46.
5. 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**: 616-24.
6. Fu Y, Dominissini D, Rechavi G, He C. Gene expression regulation mediated through reversible m^{6}A RNA methylation. *Nat Rev Genet* 2014; **15**: 293-306.
7. Hong K. Emerging function of N6-methyladenosine in cancer. *Oncol Lett* 2018; **16**: 5519-24.
8. Chen H, Gao S, Liu W, et al. RNA N(6)-Methyladenosine Methyltransferase METTL3 Facilitates Colorectal Cancer by Activating the m(6)A-GLUT1-mTORC1 Axis and Is a Therapeutic Target. *Gastroenterology* 2021; **160**: 1284-300.e16.
9. Zhong L, Liao D, Zhang M, et al. YTHDF2 suppresses cell proliferation and growth via destabilizing the EGFR mRNA in hepatocellular carcinoma. *Cancer Lett* 2019; **442**: 252-61.
10. Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* 2012; **21**: 309-22.
11. Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med* 2012; **366**: 2443-54.
12. Li N, Kang Y, Wang L, 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**: 20159-70.
13. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med* 2013; **19**: 1423-37.
14. Binnewies M, Roberts EW, Kersten K, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med* 2018; **24**: 541-50.
15. Fang H, Declerck YA. Targeting the tumor microenvironment: from understanding pathways to effective clinical trials. *Cancer Res* 2013; **73**: 4965-77.
16. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012; **487**: 330-7.
17. Marisa L, de Reyniès A, Duval A, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* 2013; **10**:

e1001453.
18. Smith JJ, Deane NG, Wu F, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology* 2010; **138**: 958-68.
19. Yang W, Soares J, Greninger P, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 2013; **41**: D955-61.
20. Colaprico A, Silva TC, Olsen C, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 2016; **44**: e71.
21. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000; **28**: 27-30.
22. Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* 2014; **9**: e107468.
23. Chen Z, Qi M, Shen B, et al. Transfer RNA demethylase ALKBH3 promotes cancer progression via induction of tRNA-derived small RNAs. *Nucleic Acids Res* 2019; **47**: 2533-45.
24. Sohail M, Shkreta L, Toutant J, et al. A novel class of inhibitors that target SRSF10 and promote p53-mediated cytotoxicity on human colorectal cancer cells. *NAR Cancer* 2021; **3**: zcab019.
25. Zhang Z, Zhang X. Identification of m6A-Related Biomarkers Associated with Prognosis of Colorectal Cancer. *Med Sci Monit* 2021; **27**: e932370.
26. Cejas P, Cavazza A, Yandava CN, et al. Transcriptional Regulator CNOT3 Defines an Aggressive Colorectal Cancer Subtype. *Cancer Res* 2017; **77**: 766-79.
27. Wang S, He Z, Wang X, Li H, Liu XS. Antigen presentation and tumor immunogenicity in cancer immunotherapy response prediction. *Elife* 2019; **8**.
28. Jiang P, Gu S, Pan D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* 2018; **24**: 1550-8.