



OPEN

Construction of N-7 methylguanine-related mRNA prognostic model in uterine corpus endometrial carcinoma based on multi-omics data and immune-related analysis

Junde Zhao^{1,3}, Jiani Zou^{1,3}, Wenjian Jiao¹, Lidong Lin¹, Jiuling Wang^{2,4}✉ & Zhiheng Lin^{1,4}✉

N-7 methylguanine (m7G) is one of the most common RNA base modifications in post-transcriptional regulation, which participates in multiple processes such as transcription, mRNA splicing and translation during the mRNA life cycle. However, its expression and prognostic value in uterine corpus endometrial carcinoma (UCEC) have not been systematically studied. In this paper, the data such as gene expression profiles, clinical data of UCEC patients, somatic mutations and copy number variants (CNVs) are obtained from the cancer genome atlas (TCGA) and UCSC Xena. By analyzing the expression differences of m7G-related mRNA in UCEC and plotting the correlation network maps, a risk score model composed of four m7G-related mRNAs (NSUN2, NUDT3, LARP1 and NCBP3) is constructed using least absolute shrinkage and selection operator (LASSO), univariate and multivariate Cox regression in order to identify prognosis and immune response. The correlation of clinical prognosis is analyzed between the m7G-related mRNA and UCEC via Kaplan–Meier method, receiver operating characteristic (ROC) curve, principal component analysis (PCA), t-SNE, decision curve analysis (DCA) curve and nomogram etc. It is concluded that the high risk is significantly correlated with ($P < 0.001$) the poorer overall survival (OS) in patients with UCEC. It is one of the independent risk factors affecting the OS. Differentially expressed genes are identified by R software in the high and low risk groups. The functional analysis and pathway enrichment analysis have been performed. Single sample gene set enrichment analysis (ssGSEA), immune checkpoints, m6A-related genes, tumor mutation burden (TMB), stem cell correlation, tumor immune dysfunction and rejection (TIDE) scores and drug sensitivity are also used to study the risk model. In addition, we have obtained 3 genotypes based on consensus clustering, which are significantly related to ($P < 0.001$) the OS and progression-free survival (PFS). The deconvolution algorithm (CIBERSORT) is applied to calculate the proportion of 22 tumor infiltrating immune cells (TIC) in UCEC patients and the estimation algorithm (ESTIMATE) is applied to work out the number of immune and matrix components. In summary, m7G-related mRNA may become a potential biomarker for UCEC prognosis, which may promote UCEC occurrence and development by regulating cell cycles and immune cell infiltration. It is expected to become a potential therapeutic target of UCEC.

UCEC is one of the three major malignant tumors in gynecology. Its incidence ranks first among malignant tumors of the female reproductive tract in developed countries such as Europe and the United States¹. The latest data show that the incidence and associated mortality of UCEC are on the rise worldwide with the delay in women's reproductive age as well as the use of hormone drugs and the rejuvenation of comorbidities such

¹Shandong University of Traditional Chinese Medicine, Jinan 250014, Shandong, China. ²Office of Medical Insurance Management, Qilu Hospital of Shandong University, Jinan 250012, China. ³These authors contributed equally: Junde Zhao and Jiani Zou. ⁴These authors jointly supervised this work: Jiuling Wang and Zhiheng Lin. ✉email: ayi-99@163.com; linzhiheng2020@163.com

as obesity, hypertension and hyperglycemia^{2–4}. Although endometrial cancer is usually diagnosed early in the course of illness and generally has a good prognosis, a small number of patients present or develop into metastatic or recurrent UCEC, with a 5-year survival rate of only 10–20%^{5,6}. The principal mode of initial treatment of UCEC is surgery. Corresponding adjunctive therapies are identified based on the recurrent risk and prognosis. Chemotherapy, targeted therapy, endocrine therapy and/or salvage radiotherapy are often used in the treatment of relapsed or advanced UCEC⁷. The overall prognosis is poor in view of recurrent or advanced UCEC. It has poor response to classical treatment. Therefore, it is urgent to explore new treatment approaches to improve prognosis.

m7G is one of the most common RNA base modifications in post-transcriptional regulation⁸. It is widely distributed in tRNA⁹, rRNA¹⁰ and the 5' cap region of mRNA in eukaryotes¹¹, playing a vital role in gene expression, processing, metabolism, protein synthesis and stable transcription¹². The m7G modification is catalyzed by the Trm8/Trm82 complex in yeast and the METTL1/WDR4 complex in human beings¹³. This modification is closely related to tRNA stability and human diseases¹⁴. Recent studies have found that the m7G is also present inside the mRNA of higher eukaryotes. It may regulate almost every process of the mRNA life cycle, including transcription, mRNA splicing and translation, etc.^{15–17}, suggesting that the presence of m7G modification sites in mRNAs may reveal new biological functions in future studies. At present, the research on identification of RNA modification sites is rapidly developing. However, the specific role and prognostic value of m7G-related mRNA remain unclear in the progression and metastasis of UCEC.

In this study, we aim to identify the m7G-related mRNAs with differential expression in UCEC, thereby constructing a new predictive prognostic risk score model and evaluating the model's predictive performance in low-risk and high-risk patients. Secondly, we have established three clustering subtypes and systematically analyzed their relationships with prognosis, TME and immune infiltration. Overall, this study will provide new insights into the role of m7G-related mRNA in predicting UEC prognosis and immunotherapy efficacy.

Materials and methods

Data download and processing. The RNA-sequencing (RNA-seq) gene expression profiles and clinical information of 587 samples were downloaded from the TCGA database (<https://portal.gdc.cancer.gov/>), including mRNA expression data from 552 UEC tissues and 35 normal endometrial tissues. Then, 539 patients with sufficient gene expression profiles and OS data were selected for follow-up analysis based on patient clinical data and their corresponding RNA-seq data. The mutation data such as 529 cases of somatic mutations and CNV data were downloaded from UCSC Xena (<https://xena.ucsc.edu/>). The TCGA samples, as validation datasets were randomly divided into two equal numbers of groups and they were named Test1 and Test2, respectively. In addition, GSE119041 was downloaded from the Gene Expression Omnibus (GEO) database as an external cohort for validation. Gene expression data were annotated (http://asia.ensembl.org/homo_sapiens/info/index). The mean expression level of duplicated genes was calculated. All gene expressions were converted to log₂ (gene expression value + 1) for follow-up analysis. A total of 29 RNA sequence datasets containing m7G modification sites were found in previous literature. The m7G gene expression level was obtained using the “LIMMA” package of R software and its expression profile was further normalized. The difference in m7G expression was compared using the Wilcoxon test between the normal sample group and tumor sample group.

Differentially expressed gene screening, related functional annotation and pathway enrichment analysis. The TCGA-UCEC samples were divided based on the quartiles of m7G expression. The lowest 25% of quartiles were defined as the low m7G expression group and the highest 25% of quartiles were defined as the high m7G expression group under difference analysis. Differentially expressed m7G genes (DEGs) were screened out according to criteria of $|\log_2 \text{FC}| > 0.1$ and false discovery rate (FDR) < 0.05 in the UCEC low and high expression groups. Heat maps of DEGs were drawn by “pheatmap” R package. DEGs-related network diagrams were drawn based on the DEGs expression and the application of “igraph” and “reshape2” package. Entrez ID was converted through “org.Hs.eg.db” R package. Gene ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathway analysis were performed on all DEGs using the “ClusterProfiler” R package.

Analysis of gene mutations and copy number variants. Gene mutation maps of UCEC patients were drawn using “maftools” R package. A CNV matrix was established for 28 m7G-related CNV genes in UCEC. The CNV matrix is a binary matrix. GAIN represents the occurrence of the CNV for a particular gene in a particular sample. Other matrix elements include the LOSS. Visualizing graphs of CNV frequency were plotted using “barplot” and “RCircos” packages.

Screening m7G-related prognostic mRNAs and constructing a prognostic risk model. The expression of 539 TCGA dataset samples was combined with clinical data to ensure that a prognostic model could predict effectively. Firstly, a univariate Cox proportional risk regression analysis of m7G-related DEGs was performed to obtain genes that were significantly related to prognosis ($P < 0.05$). Then, with survival status as the dependent variable and the selected gene expression value as the response variable, the number of genes was dimensionally reduced through Lasso COX regression analysis of 1000 times via glmnet software package, thus reducing the errors of the model to obtain a generalized linear model. After that, the multivariate Cox proportional risk regression analysis was performed to yield risk genes and construct a UCEC prognostic risk model. A disease risk score was used as a predictor of prognostic status in the model. The disease risk score was determined by the parameter X of the multivariate Cox proportional risk regression analysis and the expression Y of each gene in the sample. The risk score calculation formula is shown as follows:

$$\text{Risk score} = \sum_i^n X_i \times Y_i.$$

Construction and efficacy evaluation of prognostic risk score model. The sample data were divided into high-risk and low-risk groups according to the median of risk indexes. Combined with survival information and gene expression, the survival curve was plotted to yield the survival conditions of patients with high and low risks and high and low gene expressions, respectively, as so to evaluate whether the predictive effect of the model was significant ($P < 0.05$). The statistical method used in this process was a log-rank test. The receiver operating characteristic curve (ROC curve) was plotted using “survival ROC” package of R software in order to evaluate the predictive power of regression models over 1, 3 and 5-year survival. When the $\text{AUC} > 0.5$ and it is closer to 1, the better the prognosis will be. Based on the gene expression in the prognostic model, the principal component analysis (PCA) and t-SNE analysis were performed using “prcomp” and “Rtsne” R packages, respectively. The results were visualized using “ggplot” R package. The univariate and multivariate Cox regression analysis was performed to evaluate the independent prognostic value of risk scores in this model. DCA curve analysis was carried out using the “stdca” software package of R language to evaluate the practical application of the model in the clinical practice. A nomogram in predicting the 1, 3 and 5-year survival rates of UCEC patients was produced by the “survival” package and the “rms” package of R software. The fit of the model was assessed by calibration diagram. The sensitivity of the model was evaluated by risk scores and ROC curves. The differentiation of the model was evaluated using risk scores and ROC curves.

Functional enrichment analysis and immune-related analysis of differential genes in the two risk subgroups. Firstly, the “limma” R package was used to analyze the differential expression of genes between the high-risk and low-risk groups. Genes with $|\log_2 \text{FC}| > 0.1$ and $\text{FDR} < 0.05$ were differentially expressed between the high-risk and low-risk groups. The next step was to do the GO analysis. The infiltration scores of 16 immune cells and 13 immune functions were calculated for ssGSEA using the “GSVA” R package in UCEC. The activities of immune checkpoints and m6A RNA methylation regulatory factors were analyzed in the high-risk and low-risk groups. $P < 0.05$ was considered significant in all statistical analysis.

GSEA analysis of m7G genes between high and low expression groups. In order to elucidate the significant differences of m7G genes in function and pathway between the high and low expression groups, the gene set enrichment analysis (GSEA) was performed based on the UEC expression matrix in TCGA using the “clusterProfiler” package of R software^{18–20}. Only $\text{FDR} = q < 0.05$, the gene set was considered to be significantly enriched.

Consensus clustering analysis of m7G-related mRNA responses based on UCEC. In order further investigate the role of m7G-related mRNA in UCEC, we used the “ConensusClusterPlus” package to divide those UCEC patients of the TCGA cohort into groups and obtain the optimal number of clusters based on the cumulative distribution function (CDF). The differences in genotyping were compared based on the expression level of prognostic risk genes. The Kaplan–Meier survival curve was used to analyze the differences between the OS and PFS among different groups. Based on the sample expression level of UCEC patients in the TCG database, the matrix scores and immunity scores were calculated using the “ESTIMATE” package among genotypes. In order to quantify and analyze the immune cell infiltration levels of each genotyping sample, the “CIBERSORT” computational R package was used to plot heat maps and pairwise difference plots.

Calculation of tumor mutation burden and correlation analysis. The risk scores were calculated based on the risk score formula and the expression of genes in each genotyping sample. According to the median of risk indexes, the genotyping data were divided into high-risk and low-risk groups. The “maftools” R package was used to draw the gene mutation maps of patients in both risk subgroups. The difference in TMB was compared between the high-risk and low-risk groups. The correlation was compared between the TMB levels and risk scores. The new grouping is based on the Median value of TMB for the entire sample. Kaplan–Meier method was used to analyze the relationship of OS in the samples between the high and low TMB groups and the high-risk and low-risk groups within the TCGA cohort.

Correlation analysis between risk scores and stem cells, TIDE scores and drug sensitivity. The “ggExtra” and “ggpubr” R packages were applied to calculate the stem cell correlation between the two risk subgroups and the “ggplot” package was applied to visualize the data results. The TIDE scores were compared between the high-risk and low-risk groups and a violin diagram was plotted. Finally, in order to further guide the clinical precision medications, two R packages of “Limma” and “pRRophetic” were used to analyze the data of the two risk subgroups the “ggboxplot” was introduced to visualize the drug sensitivity.

Cell culture. Human endometrial cancer cell line HEC-1A (purchased from Shanghai Fuhengsheng Materials Co., LTD.) was cultured in Mccoy’s 5A + 10%FBS (both purchased from Gibco), and normal endometrial epithelial cells (purchased from Shanghai Fuhengsheng Materials Co., LTD.) were cultured in EMEM complete medium at 37 °C, 5%CO₂, pH 7.2–7.4, sterile constant temperature culture. When the degree of cell fusion reached 90%, cell passage was carried out.

Rt-PCR. Total RNA was collected using a RNeasy mini kit (RNA Fast 200, China), according to the manufacturer's instructions. Then, cDNA was synthesized from 1 µg of total RNA using a SureScript First-Strand cDNA Synthesis Kit (GeneCopoeia, USA). Finally, the expression levels of NSUN2, NUDT3, LARP1, NCBP3. The sequences of the primers were as follows:

NSUN2: forward 5'-ACCTGGCTCAAAGACCACACAG-3', Reverse 5'-TGGCTTGATGGACGAGCAGGTA-3'.
 NUDT3: forward 5'-GAAGCACAGGACGTATGTCTATG-3', reverse 5'-CTGCACGGGTTTGTGATACTG-3'.
 LARP1: forward 5'-GCTGTTTAGGAACAGCTGCC-3', reverse 5'-CCACAGG TGACAGGGAGAAG-3'.
 NCBP3: forward 5'-AGGAAATCGGCGTCCAAGTT-3', reverse 5'-TGCCTTGCCAGTCTTTGTCT-3'.
 GAPDH: forward 5'-GCACCGTCAAGGCTGAGAAC-3', reverse 5'-TGGTGAAGACGCCAGTGA-3'.

Melting curve analysis was used to confirm the amplification specificity. The quantification data were analyzed using LightCycler analysis software (version 4.0; Roche Applied Science, Mannheim, Germany). Relative expression was normalized to that of GAPDH.

Results

Transcriptomic difference analysis in m7G low and high expression groups in patients with UCEC. According to the m7G expression quartiles, TCGA-UCEC tumors were divided into two groups with high and low m7G expressions for differential expression analysis to identify 20 DEGs ($|\log_2 FC| > 0.1$, $FDR < 0.05$) with a difference of more than 2 times. The detailed distribution of 20 DEGs was shown by heat map (Fig. 1A) and a complete network of relationships among DEGs was plotted (Fig. 1B). The GO enrichment results showed that the DEGs of the m7G high and low expression groups were mainly involved in biological processes such as RNA cap binding, RNA 7-methylguanosine cap binding and catalytic activity, acting on RNA, etc. Its products were mainly involved in ribonucleoprotein granule, cytoplasmic ribonucleoprotein granule and RNA cap binding complex and other cellular components, playing a role in biological molecular functions, including organic cyclic compound catabolic process, aromatic compound catabolic process, cellular nitrogen compound catabolic process, heterocycle catabolic process and nucleobase-containing compound catabolic process, etc. (Fig. 1C). Meanwhile, the results of KEGG signaling pathway analysis showed that the m7G high expression was mainly enriched in RNA degradation, nucleocytoplasmic transport, mRNA surveillance pathway, spliceosome and so on signal pathways (Fig. 1D).

Gene mutation map and CNV analysis of UCEC patients. Of the 529 HCC patients in the TCGA database, 152 (28.73%) had genetic mutations (Fig. 2A). One of the most common types of gene mutations was the missense mutation. However, the proportion of missense mutations was small in some genes, such as EIF4G3 and LARP1. In addition, EIF4G3 mutations had the highest frequency, followed by AGO2, NCBP1, LARP1, GEMIN5, NSUN2, CYFIP1, DCP2, WDR4 and EIF4A1. However, NUDT4B and NCBP3 did not show any mutations. In order to analyze the copy number variants in the m7G-related genes, we selected 28 copy number variant genes in UCEC for further analysis. The status of m7G-related gene copy number variants in 28 UCEC patients was shown in Fig. 2B, where NCBP2, AGO2, NSUN2, EIF4E1B, LSM1, NCBP3 and NUDT16 had significant CNV amplification frequencies, and a small number of genes were missing, including EIF4G3, IFIT5, EIF4E2, and EIF4E3. Figure 2C showed the visual expression of all m7G-related copy number variant genes on chromosomes in the UCEC sample.

Construction of prognostic model of m7G-related mRNA. Firstly, in the TCGA cohort, the expression data of 20 m7G-related DEGs were extracted. The univariate COX regression was carried out for the overall survival (OS) to yield 5 m7G-related mRNAs significantly related to the OS ($P < 0.05$) (Fig. 3A). Then, 4 genes were obtained for subsequent analysis through Lasso COX regression analysis according to the optimal λ value (Fig. 3B,C). The multivariate COX proportional risk regression analysis was used to obtain a total of 4 risk genes (Fig. 3D), namely NSUN2, NUDT3, LARP1 and NCBP3 for the construction of a prognostic risk model.

Construction of a risk score model based on m7G-related prognostic mRNA. The risk scores of all patients were calculated. All UCEC patients in the TCGA cohort were divided into high and low risk groups based on the median risk score. The distribution of risk scores and survival status was shown in Fig. 4A,B. The Kaplan–Meier survival curve showed that the patients with hepatocellular carcinoma in the high-risk group and those with NSUN2, NUDT3 and LARP1 genes in the high expression group had a lower survival rate and shorter survival time (Fig. 4C–F). The patients in the high expression group with the NCBP3 gene had a higher survival rate and longer survival times (Fig. 4G). In addition, the ROC curve showed that the model also had a good predictive power for overall survival of UCEC patients in the TCGA cohort, with AUC values of 0.623 (1 year), 0.665 (3 years) and 0.637 (5 years), respectively (Fig. 4H). At the same time, an external cohort and two sets of validation datasets were constructed to explain the generalization capacity of the models and the effect in other cohorts (Fig. 4I,J, Supplementary Figs. S1A, S2A). The patients in the low-risk group have a higher survival rate and longer survival times (Fig. 4I, Supplementary Figs. S1B, Fig S2B). The predictive power of the model was also validated in the external cohort (Fig. 4J). PCA and t-SNE analyses showed that the difference in sample data distribution was significant between the two risk subgroups in the TCGA cohort (Fig. 4K,L).

Prognostic model as an independent prognostic factor in patients with UCEC. Univariate and multivariate COX regression analysis was used to evaluate whether the prognostic model was an independent

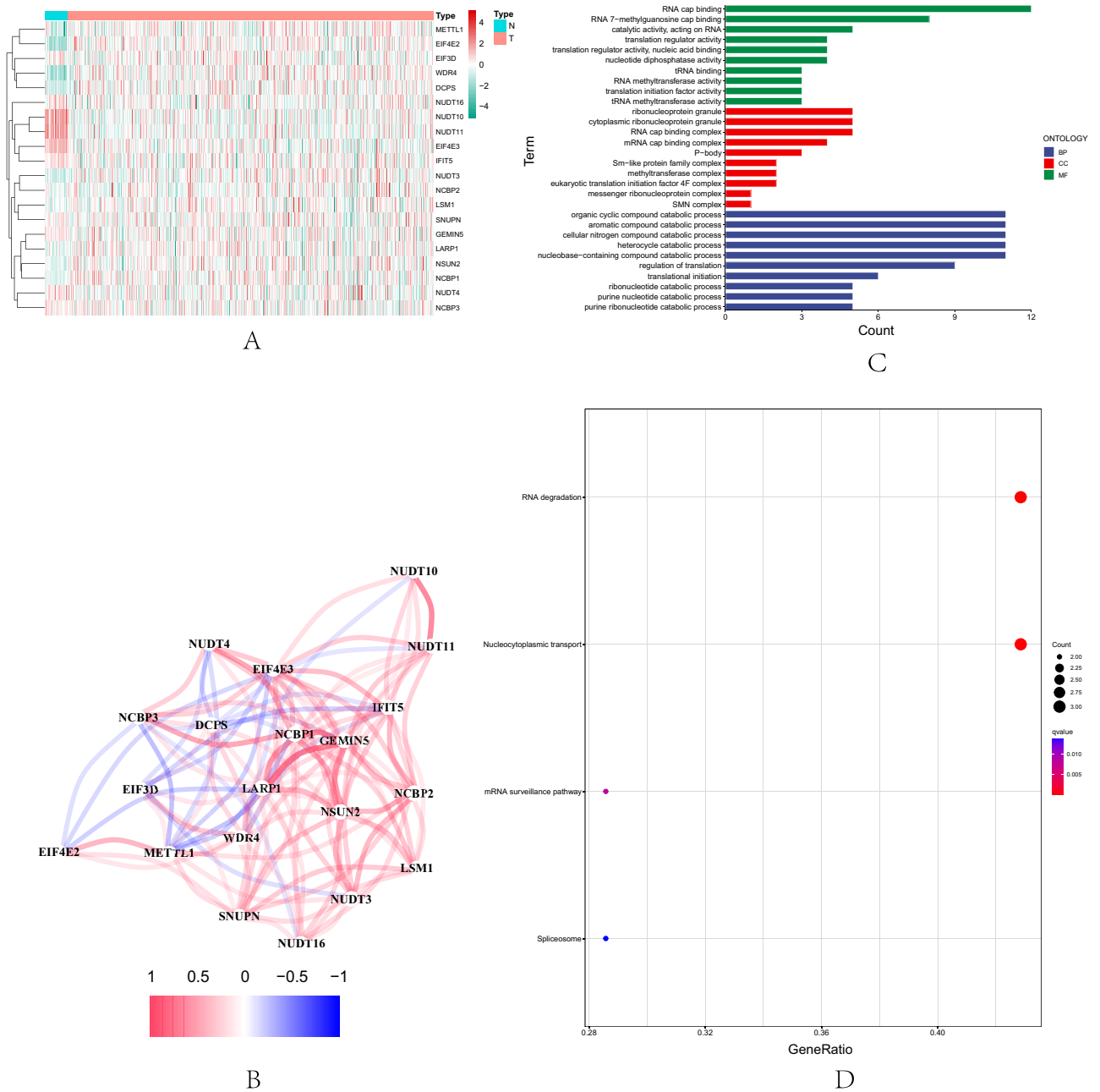


Figure 1. Transcriptomics analysis of m7G high and low expressions in UCEC tissues. **(A)** The heat map shows 20 DEGs. **(B)** Correlation network diagram of DEGs. **(C)** GO enrichment analysis results. **(D)** KEGG enrichment analysis results.

prognostic factor for the patients with UCEC. In the TCGA cohort, the univariate COX regression analysis showed a significant correlation [risk ratio (HR) 1.809; 95% confidence interval (CI) 1.404–2.332; $P < 0.001$] between the risk scores and overall survival of patients with UCEC (Fig. 5A). The multivariate COX regression further showed that the risk score was an independent prognostic factor for overall survival [HR 1.809; 95% CI 1.404–2.332; $P = 0.020$] (Fig. 5B). In order to understand the clinical effectiveness of the model, it was known from the DCA curves that the score-clinical variable integration model had a more significant net benefit than the traditional age, treatment protocols of all patients or non-treatment protocol in prognostic prediction (Fig. 5C). In the prognostic nomogram, each differential factor was based on the scores corresponding to the first ruler and the total score was obtained by added them up. Then according to the scores of the second ruler, the 1, 3 and 5-year survival rates were corresponded downward. As can be seen from the chart, the higher the age, the higher the score and the worse the survival prognosis (Fig. 5D). In clinical practice, a scoring system can be constructed for patients with different conditions, so as to obtain individualized survival predictions. In the calibration plot, the abscissa represented a predicted value and the ordinate represents an actual value. This curve and the 45° dashed line (i.e., an ideal state) represented the differences between the actual and predicted values. From the calibration chart, the 1-year survival prediction was close to the same with a high predictive

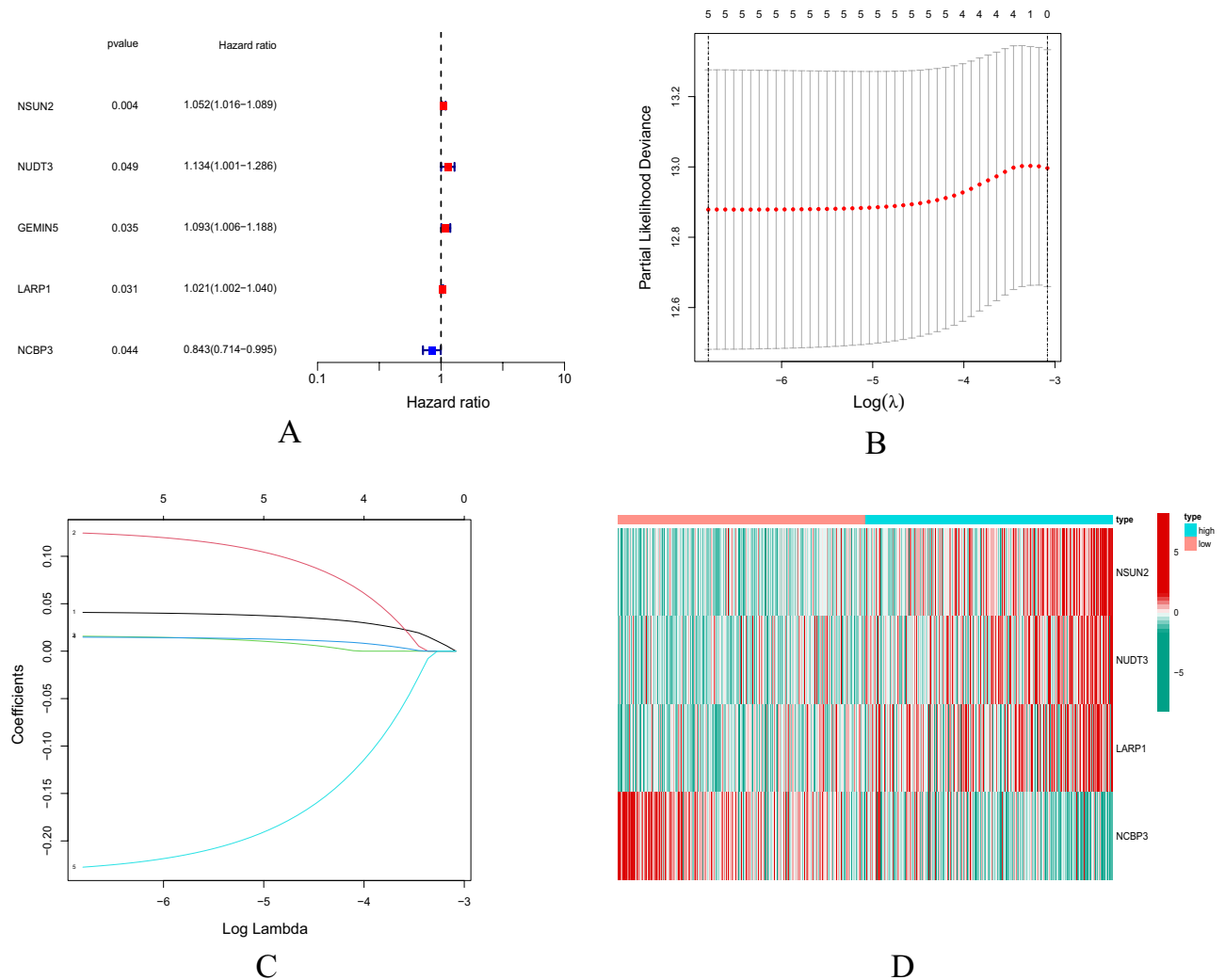


Figure 3. Construction of m7G-related mRNA prognostic model. (A) Screening prognostic mRNAs using univariate COX regression analysis. (B) LASSO regression analysis. (C) Multivariate COX proportional risk regression analysis of four prognostic mRNAs for establishment of a prognostic model. (D) Heatmap of prognostic-related m7G-related mRNAs. Green represents low expression of m7G-related mRNA, red represents high expression of m7G-related mRNA.

IFN Response were higher in the low-risk group than those in the high-risk group. Moreover, the difference was statistically significant ($P < 0.05$) between the two risk subgroups (Fig. 6C). In terms of immune checkpoints, the difference in immune-related genes was significantly between the two risk subgroups ($P < 0.05$). Moreover, the expression levels of HHLA2, TNFRSF14, TNFRSF25, CD244, LGALS9, CD40LG, CD200, TNFSF14, TNFRSF4, TNFSF15 and CD44 in the low-risk subgroup were higher than those in the high-risk subgroup (Fig. 6D). As shown in Fig. 6E, the expression levels of M6A-related genes RBM15, YTHDF2, YTHDF1, WTAP and HNRNPC in the high-risk subgroup were significantly higher than those in the low-risk subgroup ($P < 0.01$).

GSEA-based m7G-related signal pathways. The low and high expressions of m7G were detected with the “cluserPorfiler” package. The signal pathways of UECG participating in TCGAs were identified by GSEA analysis. Signaling pathways such as B CELL RECEPTOR, CYTOSOLIC DNA, PENTOSE PHOSPHATE and INSULIN were enriched in the m7G highly expressed UECG phenotype (Fig. 7A–D), as well as significantly enriched in endometrial cancer, bladder cancer, non-small cell lung cancer and pancreatic cancer and other diseases (Fig. 7E–H). Biological processes such as alpha linolenic acid metabolism, transporters, hedgehog signaling pathway and arachidonic acid metabolism were enriched in the m7G lowly expressed phenotype (Fig. 7I–L).

Analysis of survival, tumor microenvironment and immune cell infiltration among 3 genotypes. Based on the expression similarity of m7G-related mRNAs in the UECG samples, the gene set was finally divided into three stable clusters by choosing $k=3$, with the stability of clusters ascending between $k=2$ and $k=9$ (Fig. 8A–D). The differential distribution of the four prognostic risk genes in the three types was shown in Fig. 8E. As shown in the Kaplan–Meier survival curves in Fig. 8F,G, the difference in survival state was

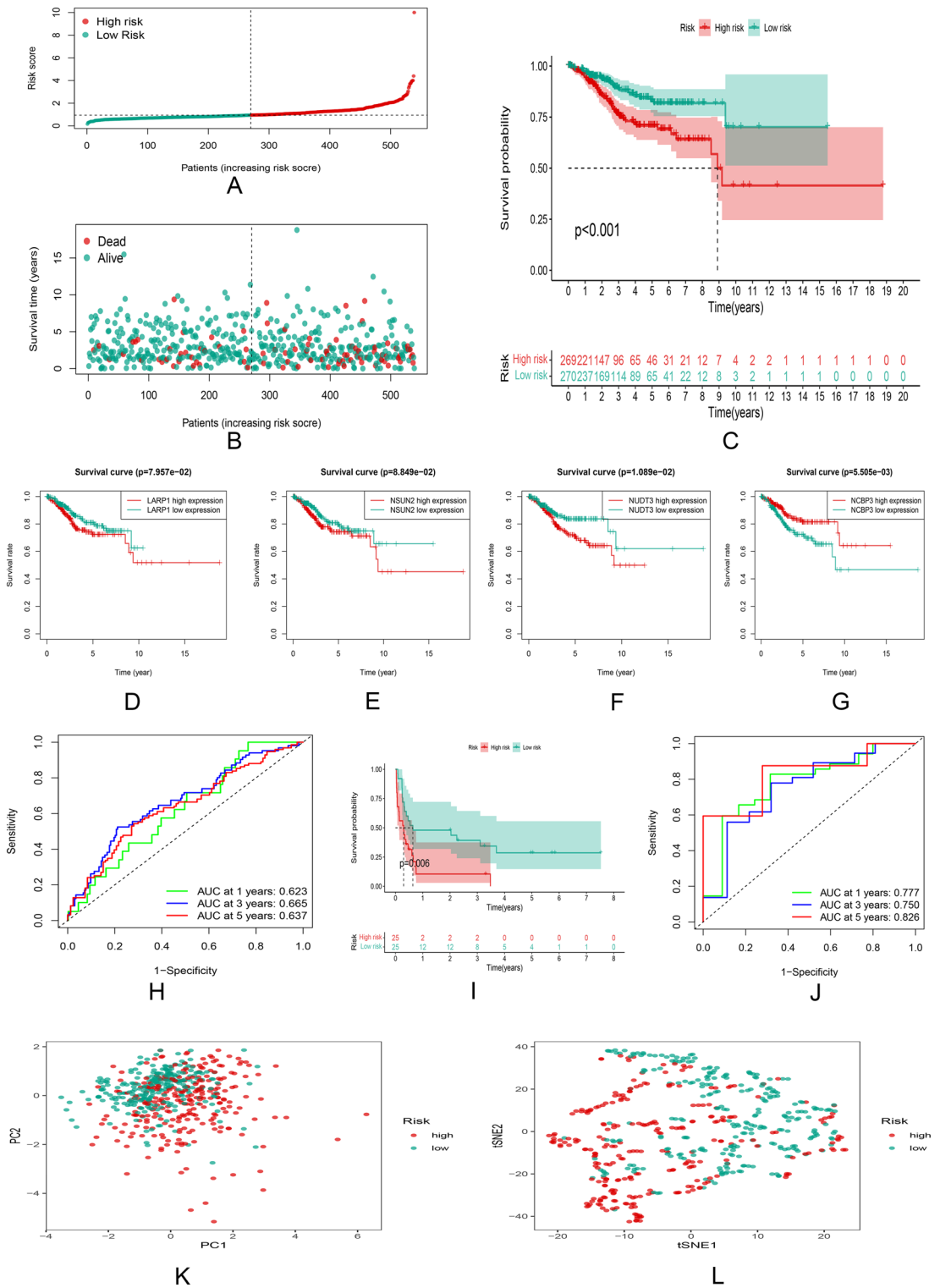


Figure 4. Survival analysis and comparison between the high and low risk groups. (A,B) Median distribution of risk scores and survival status of UCEC patients. (C) Kaplan–Meier survival curves of patients in the high and low risk groups in the TCGA cohort. (D–G) Survival status of high and low expression groups with NSUN2, NUDT3, LARP1 and NCBP3 genes. (H) ROC curves and AUC values of the model in the TCGA cohort. (I) PCA analysis results. (J) t-SNE analysis results.

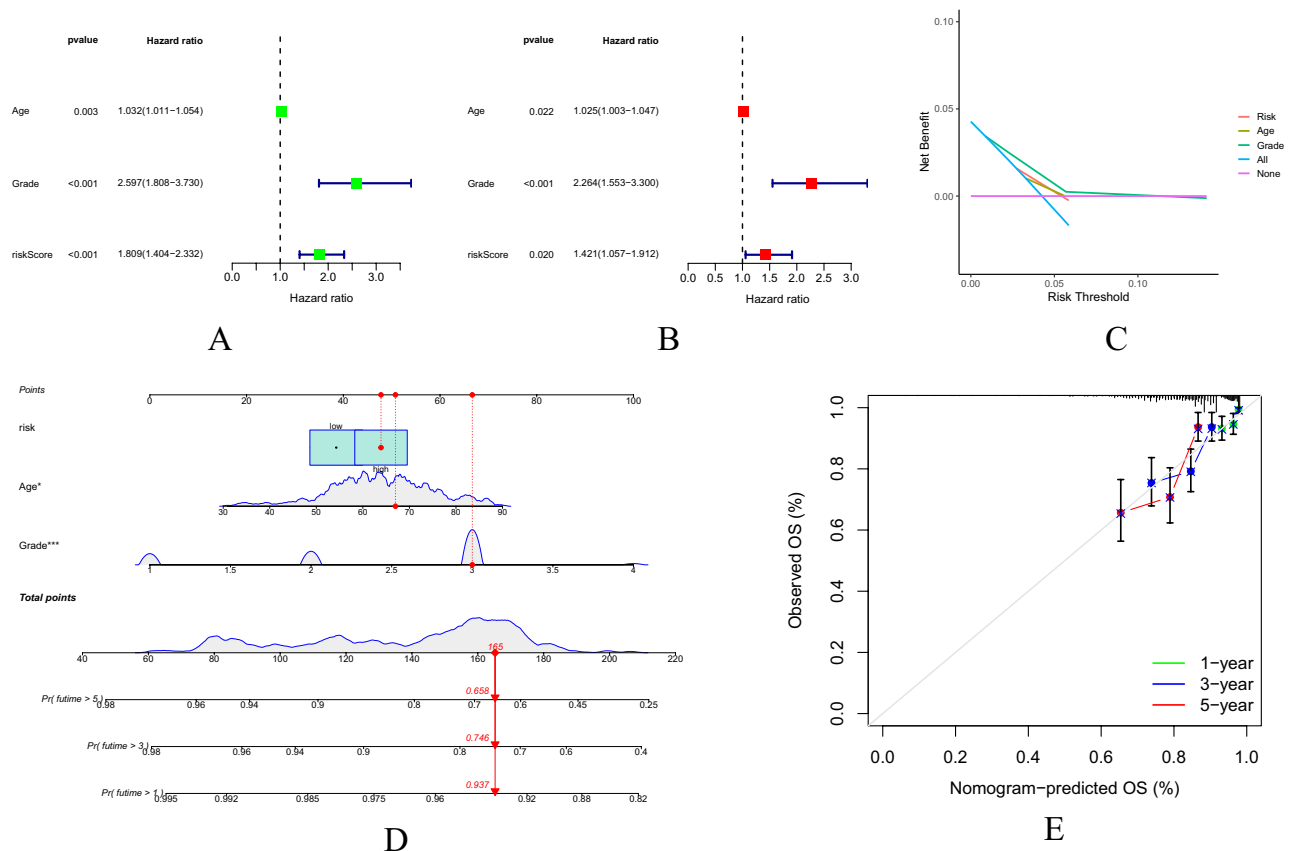


Figure 5. Independent prognostic analysis of risk score model. (A,B) Univariate and multivariate COX regression analysis of OS in the TCGA cohort. (C) DCA curve of scores and score-clinical variable integration mode. (D) Survival prognosis nomogram of UCEC patients. (E) Calibration charts for predicting 1, 3 and 5-year survival of UCEC patients.

significantly among three groups. Besides, the OS and PFS of the third group were the best ($P < 0.001$). TME, ESTIMATE Score, Immune Score, Stromal Score, Tumor Purity were significantly correlated with ($P < 0.01$) types C1–3 (Fig. 8H–K).

Based on the gene expression profiles of these samples, CIBERSORT was used to analyze the main components of the immune cell infiltration matrix of three types. Figure 8L showed that the difference in immune cell infiltration was significant among the three types. After performing the differential analysis of the same immune cells for the 3 types, the results showed that the difference in T cells CD8, T cells CD4 memory resting, T cells CD4 memory activated, T cells follicular helper, T cells regulatory (Tregs), T cells gamma delta, NK cells activated, macrophages M1, macrophages M2, dendritic cells resting and dendritic cells activated was significant (all $P < 0.05$) among three types. See Fig. 8M.

Significant correlation between the TMB levels and OS in patients with UCEC. Of the 255 UCEC patients in the high-risk group, 250 (98.04%) had genetic mutations in the gene mutation map (Fig. 9A) while 266 UCEC patients in the low-risk group, 262 (98.5%) had genetic mutations in the gene mutation map (Fig. 9B). There was no significant difference in TMB levels in the high-risk and low-risk groups (Fig. 9C). There was no significant correlation between the TMB levels and risk scores in UCEC patient samples (Fig. 9D). The TMB level in UCEC patients was calculated. Having combining risk scores with TMB levels, it was concluded that the patients with UCEC had a higher TMB level and longer survival time in the low-risk group ($P < 0.001$). See Fig. 9E,F.

Risk score model closely related to stem cells, TIDE scores and drug sensitivity. As shown in Fig. 10A,B, Spearman correlation test suggested that the risk score model was positively correlated with RNAss ($P < 0.001$) and not significantly correlated with DNAss ($P = 0.37$). UCEC sample data in the low-risk group had scored higher in the TIDE scores and the difference was statistically significant ($P < 0.01$) between the two risk subgroups (Fig. 10C). Drug sensitivity analysis revealed that there was a difference in drug sensitivity (all $P < 0.001$) between the high-risk and low-risk groups. Moreover, the bicalutamide sensitivity (IC₅₀) of bicalutamide, EHT.1864 and temsirolimus in the low-risk group was lower than that in the high-risk group (Fig. 10D–F). The IC₅₀ of methotrexate, midostaurin and parthenolide in the high-risk group was lower than that in the low-risk group (Fig. 10G,I).

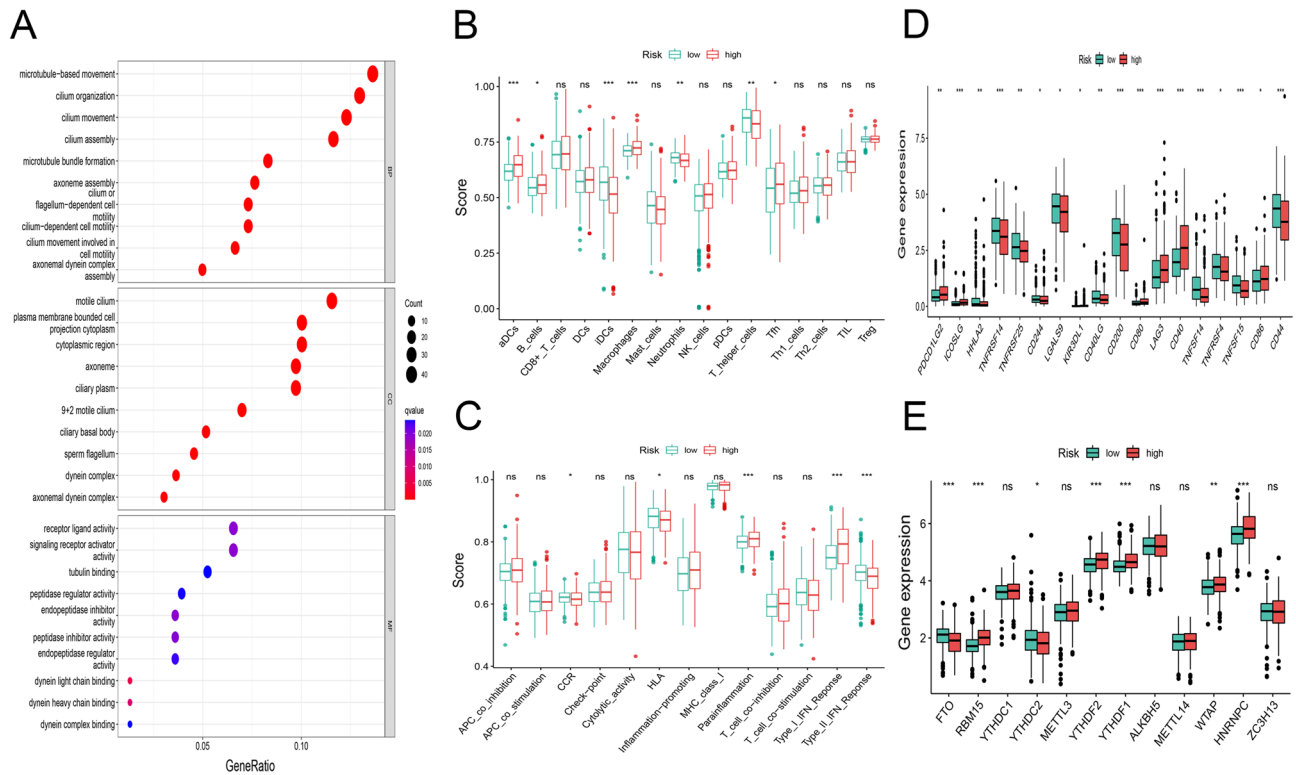


Figure 6. Functional enrichment analysis and immune infiltration levels of differential genes in the two risk subgroups. **(A)** GO analysis of differentially expressed genes. **(B)** Infiltration levels of immune cells. **(C)** Infiltration scores of immune pathways. Differential expression of 18 common immune checkpoints **(D)** and 12 M6A-related genes **(E)** between the two risk subgroups.

RT-PCR. Compared with normal endometrial epithelial cell line hEEC, the expression levels of NSUN2, NUDT3 and LARP1 in human endometrial carcinoma cell line HEC-1A were significantly increased ($P < 0.001$), while the expression levels of NCBP3 were significantly decreased ($P < 0.001$), further supporting the conclusion of the above analysis (Fig. 11).

Discussion

UCEC is a common gynecologic malignant tumor. The principal mode of initial treatment for UCEC is surgery and corresponding adjuvant therapies decided based on the recurrent risk and prognosis. For the treatment of recurrent or advanced endometrial cancer, the targeted therapy, endocrine therapy, and/or palliative radiation therapy (radiotherapy) are mostly performed based on factors such as histological type, differentiation, lymph node metastasis and depth of muscle invasion (chemotherapy) etc. Non-focal recurrent or metastatic UCEC has a poor prognosis, with a 5-year survival rate of only 10–20%. Considering that there is currently no standard post-line treatment protocol for relapsed or advanced UCEC, poor overall prognosis and poor response to classical therapy, new prognostic targets and treatment strategies need to be explored to improve prognosis.

As one of the basic processes of epigenetic regulation, RNA modification is involved in many biological physiological and pathological functions^{21–24}. Studies have revealed that RNA modification disorders promote the processing and translation of carcinogenic transcription subsets, and are closely associated with the metastasis and progression of prostate cancer, esophageal squamous cell carcinoma, gastrointestinal tumors and so on human cancers^{25–29}. However, current research on the role of m7G-related mRNAs in the occurrence and development of UCEC is still limited. In this paper, we have constructed a risk score model for the first time based on m7G-related mRNA and systematically analyzed the prognostic and immunotherapeutic value of this model for patients with UCEC. Firstly, we have screened m7G-related mRNAs that are differentially expressed in UECC tissues and paracancer tissues and analyzed the correlation among the mRNAs. By combining the expression of UCEC samples in the TCGA dataset with clinical data, and performing univariate, Lasso COX regression analysis and multivariate Cox proportional risk regression analysis, risk genes are obtained and a UCEC risk prognosis model is established.

We have then comprehensively investigated the relationship between the four m7G-related mRNAs (NSUN2, NUDT3, LARP1 and NCBP3) and UCEC clinical data variables. Existing studies have shown that these genes play vital roles in the pathogenesis and progression of human cancer. AS a post-transcriptional regulatory factor, NOP2/Sun domain family member 2 protein (NSUN2) plays a key role in catalyzing tRNA methylation, promoting gene and protein translation and maintaining transcriptional stability^{30–33}. Besides, NSUN2 has a dual role in tumor cells, playing different functions at different stages of the cell cycle in animal tissues. Although a

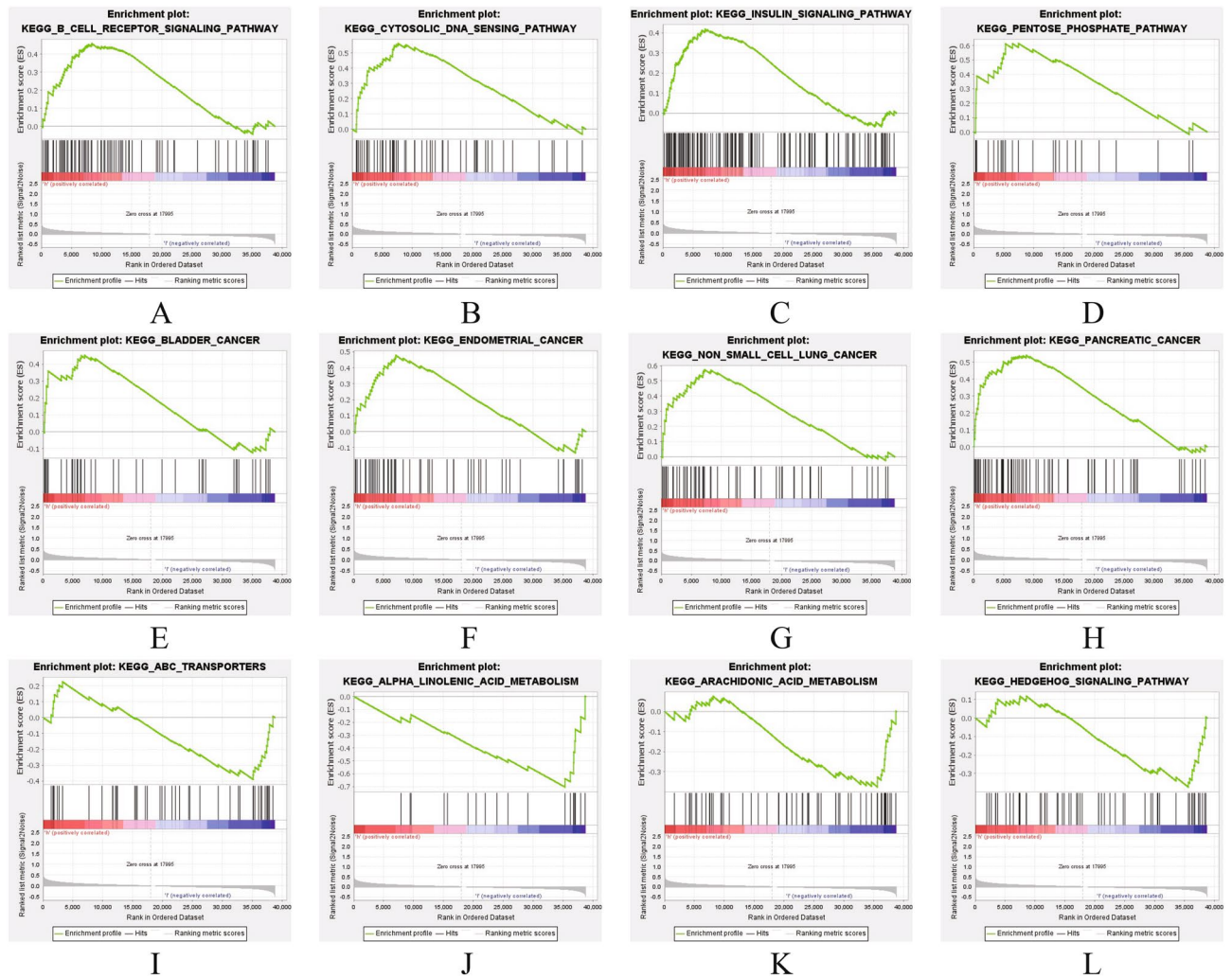


Figure 7. GSEA of M7G high expression (A–H) and low expression (I–L) samples.

study report reveals that the NSUN2 is overexpressed in epithelial tumor cells³⁴ and ovarian cancer³⁵ compared to normal tissue. Moreover, the patients with a low level of NSUN2 show a better overall survival rate than those with a high level of the NSUN2. However, in human skin cancer, the metastatic capacity of cancer cells is negatively correlated with the expression of the NSUN2³⁴. In addition, some studies have shown that the NSUN2 has the properties of delaying protein synthesis and controlling the stem cell cycle in stem cells, which may lead to tumor recurrence and chemotherapy resistance³⁴. NUDT3 is a cytoplasmic protein with mRNA dissociation activity in the family of diphosphoinositol polyphosphate phosphohydrolases (DIPP)³⁶. It is found that through genome-wide and NUDT3 gene knockout cell analysis of NUDT3 damaged cells, NUDT3 is a mRNA dissociation enzyme that directly or indirectly affects the stability of mRNAs and regulates expression levels to adjust the cell migration, such as MCF-7 breast cancer cell migration³⁷. In addition, NUDT3 has been found to be a key link in proliferative biological pathways. Small molecule drugs act as targeted agents for Triple-negative breast cancers (TNBCs) by inhibiting NUDT3³⁸. La-related protein 1 (LARP1) belongs to RNA-binding proteins (RBPs). As one of the post-transcriptional regulatory factors, the carcinogenic properties of LARP1 has been demonstrated in many studies. LARP1 promotes the occurrence of NSCLC and CC by targeting its enhancers to positively regulate the expression of mTOR in non-small cell lung cancer (NSCLC) and cervical cancer (CC)^{39,40}. At the same time, LARP1 can be used as a prognostic marker of colorectal cancer (CRC). Its overexpression is significantly correlated with the poor prognosis of the CRC and the OS^{41,42}. In addition, LARP1 can also indirectly regulate the expression of bcl2 and BIK and the development of ovarian cancer³⁹. Nuclear cap-binding protein 3 (NCBP3) (or C17orf85) is a new cap-binding protein that may bind directly to RNA caps^{43,44}. NCBP3 may act as a bridge between RBPs in the biological function of mRNA, upregulating the expression of downstream target genes in non-small cell lung cancer by interacting with NCBP1, thereby promoting the progression of lung cancer⁴⁵. Furthermore, NCBP3 is found in glioma tissues and cells to bind directly to small nucleolar RNA host gene 6 (SNHG6) and inhibit the transcription of gastrulation brain homeobox 2 (GBX2) in a manner that relies on polycomb repressive complex 2 (PRC2), thereby promoting malignant progression of gliomas⁴⁶. These results show that the four m7G-related mRNAs of NSUN2, NUDT3, LARP1 and NCBP3 are closely related to tumorigenesis and prognosis. This confirms our conclusions in UCEC.

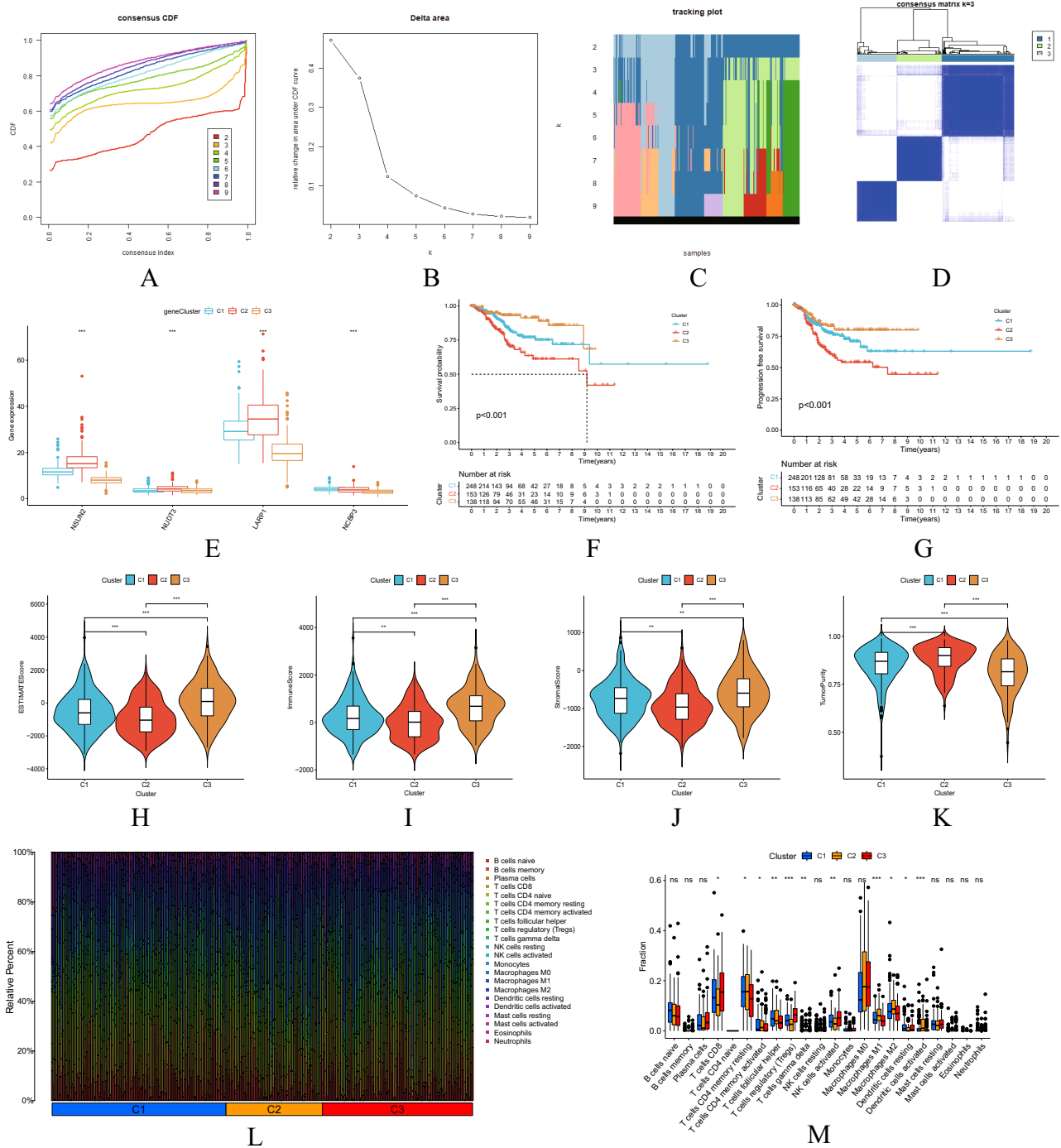


Figure 8. Survival, TME and immune cell infiltration analysis of consensus clustering recognition and typing. (A) Consensus clustering CDFs of $k=2-9$. (B) Relative change in area under CDF curve of $k=2-9$. (C) Tracking plot of $k=2-9$. (D) Consensus clustering matrix of $k=3$. (E) Differential distribution of 4 prognostic risk genes in 3 types among the three groups with respect to OS (F) and PFS (G) Kaplan–Meier survival curve. (H) ESTIMATEScore. (I) ImmuneScore; (J) StromalScore. (K) TumorPurity; Principal component analysis plot of immune cell infiltration matrix of 3 types (L) and variance analysis plot (M).

It has been found that the UCEC could evade the immune system to achieve tumor progression and metastasis by mimicking the immune tolerance mechanisms that occur in the maternal fetus^{47,48}. Currently, an immunotherapy regimen with immune checkpoint inhibitors—lenvatinib plus pembrolizumab has become an emerging field of progressive advanced UCEC research and treatment after systemic treatment^{49–51}. We have divided the UCEC patients into high-risk and low-risk groups based on a risk score model and compared differentially expressed genes between the two risk subgroups. The results show that these four m7G-related prognostic mRNAs of NSUN2, NUDT3, LARP1 and NCBP3 in the risk score model are not only associated with the patient’s

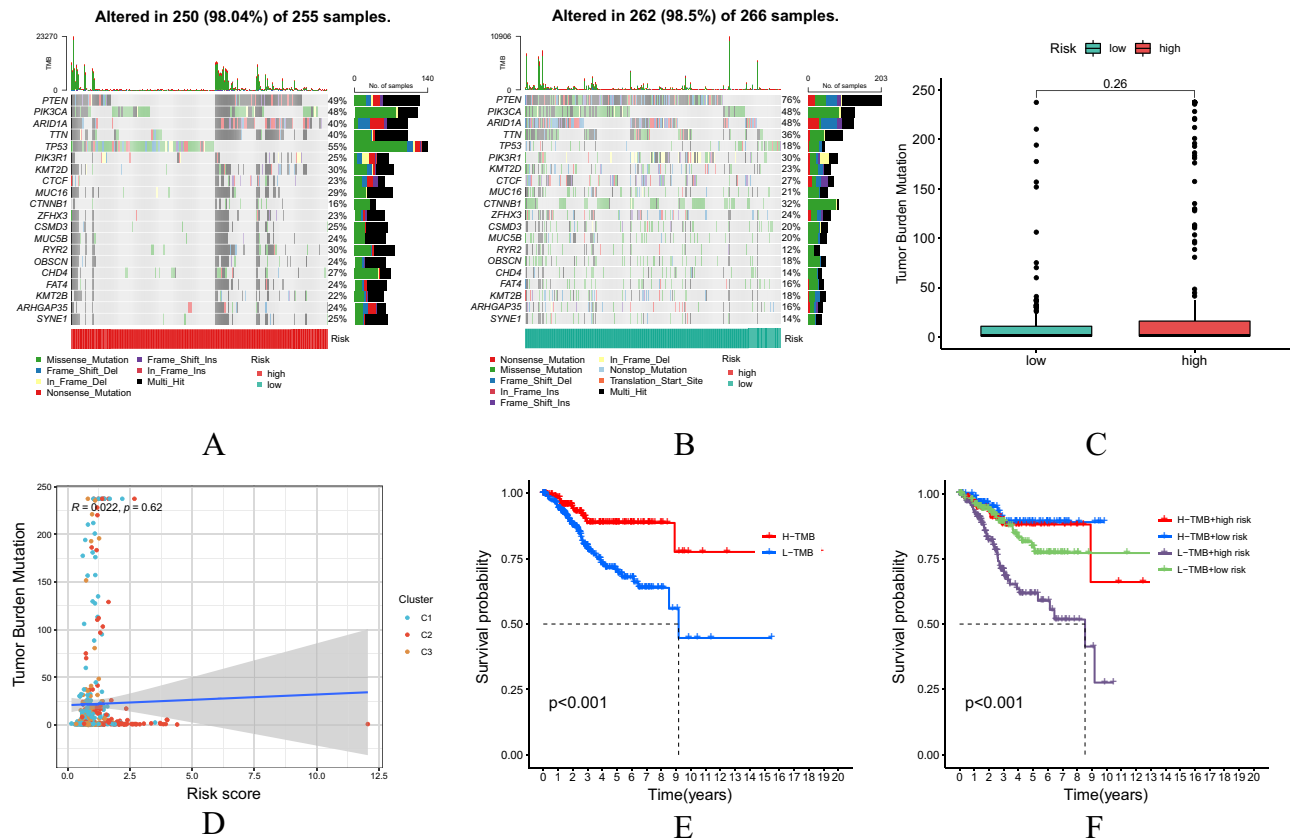


Figure 9. Correlation analysis of TMB levels between the two risk subgroups. (A,B) Distribution of the first 20 mutant genes in the high and low risk groups. (C) Difference analysis of TMB in the two risk subgroups. (D) Correlation between TMB and risk score. (E,F) Analysis of OS in UCEC patients in combination with TMB levels and risk scores.

OS, but also significantly related to immune cell infiltration, immune checkpoints and immune-related signaling pathways, etc. These findings may contribute to further insight into the effects of m7G-related prognostic mRNA on UCEC prognosis and immunotherapy. A variety of prognostic markers have been proved to have considerable prospects in the immune and tumor microenvironment, providing a new direction for immunotherapy^{52,53}. ssGSEA analysis shows a significant increase in immune cells that promote the tumor proliferation and metastasis such as infiltrating macrophages in the UCEC samples of high-risk group^{54,55}. The poor prognosis of patients in the high-risk group may be related to pro-cancer infiltrating immune cells. In addition, compared with the m7G-related prognostic mRNA low-risk group, UCEC samples in the high-risk group mostly have a lower level of immune checkpoint molecular expression, suggesting that the patients in the low-risk group are more likely to benefit from the treatment with immune checkpoint inhibitors. It also shows that our m7G-related prognostic mRNA risk score model may be able to predict the clinical efficacy of immune checkpoint blocking therapy in patients with UCEC.

In this paper, based on the expression similarity of m7G-related mRNAs in the UCEC samples, the gene set is finally divided into three stable clusters by choosing $k = 3$. There have been many previous reports on UCEC subtypes, such as the quantification of immunophenoscore (IPS) by the pattern of gene differential expression and the construction of five immune molecular subtypes⁵⁶. There are also methods such as clustering EC samples in TCGA into immune signature cluster modeling using GVSA enrichment analysis, evaluated immune cell profiling in UCEC cohorts and defined four immune subtypes of EC⁵⁷.

Expression of TMB has gradually emerged as the best biomarker chosen for a variety of tumor immune checkpoints, including lung cancer, colorectal cancer, prostatic cancer and breast cancer⁵⁸. In this study, we have obtained 3 genotypes based on consensus clustering analysis of UCEC responses to m7G-related mRNAs. Through somatic mutation analysis of UCEC samples in the TCGA database, we have found that the patients in the low-risk group have a higher level of TMB, and those with high TMB in the low-risk group have the best prognosis. As a novel biomarker, TMB has been associated with the efficacy of immunosuppressants on non-small cell lung cancer and malignant melanoma⁵⁹. Some studies have suggested that the patients with a high TMB have a better prognosis in non-small cell lung cancer and malignant melanoma, compared with renal clear cell carcinoma, colon cancer and prostatic cancer⁶⁰. This means that the patients with a low risk and a high TMB are likely to be recognized by immune cells and may benefit more from immunotherapy.

Bicalutamide, a first-generation non-steroidal androgen receptor (AR) antagonist, has become a key part of prostate cancer (PCA) treatment by acting directly on the AR⁶¹. In Hepatocellular carcinoma (HCC), low

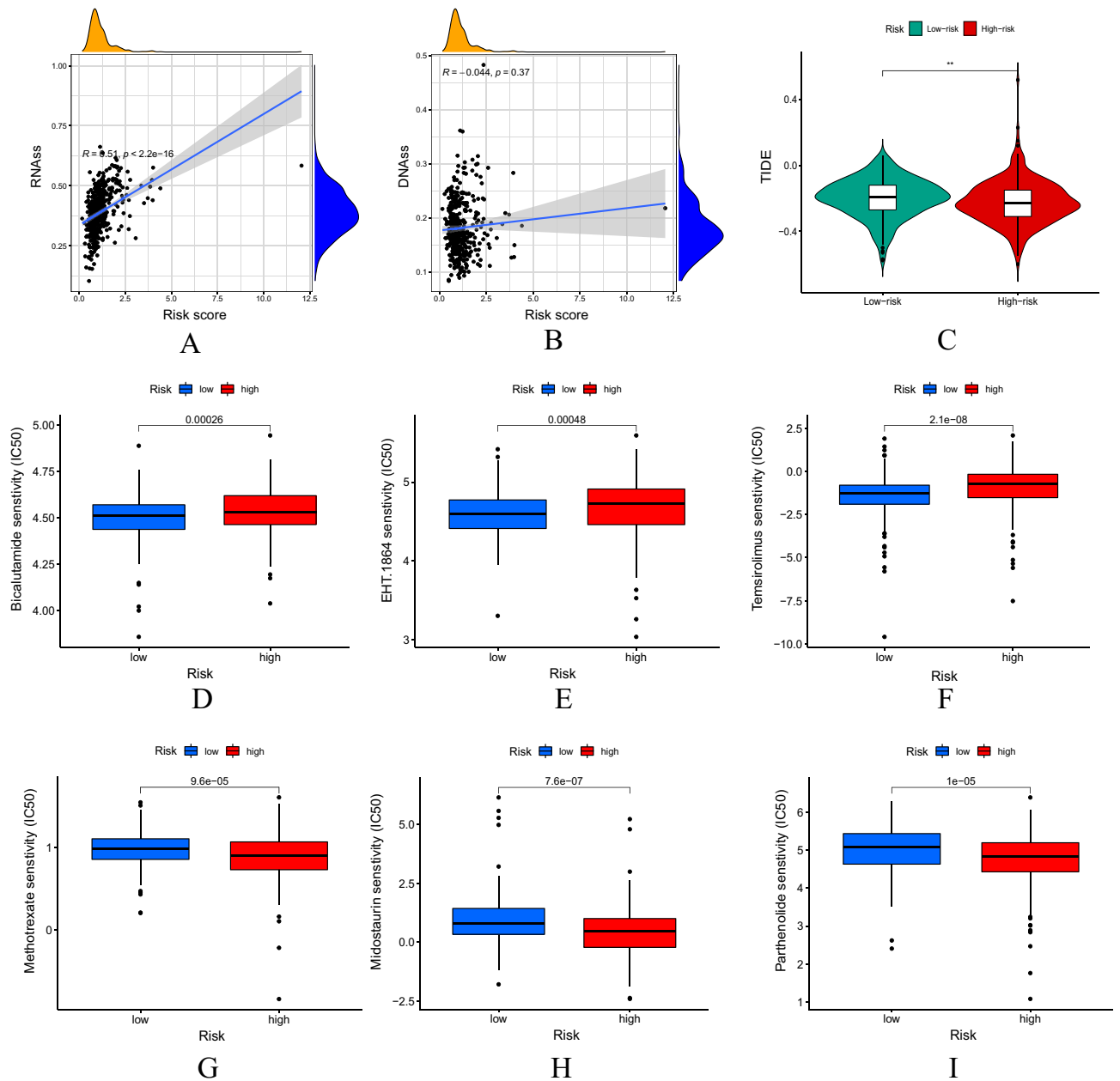


Figure 10. The risk score model is closely related to stem cells, TIDE scores and drug sensitivity. Correlation between the risk scores with RNAss (A) and DNAss (B) stem cells. (C) Differential expression of TIDE scores between the low and the high risk groups. (D–I) Drug sensitivity analysis of two risk subgroups.

expression of SNHG1/SNHG3 was more sensitive to bicalutamide⁶². In a lncRNA prognostic risk model of gastric cancer patients, high-risk patients were more sensitive to chemotherapeutic agents such as bicalutamide compared to the low-risk group⁶³.

Limitations. However, our current research still has some shortcomings. Firstly, the AUC value of the verification group is not ideal, so more external databases and clinical samples are needed to validate its predictive ability. In addition, the specific mechanisms by which the 4 risk genes affect UCEC prognosis still need to be further studied, which may become a new target for the UCEC treatment.

Conclusion

To sum up, in this study, bioinformatics methods have been used to analyze multi-omics data of UCEC samples in the TCGA database. The first UCEC prognostic risk score model is established based on 4 m7G-related mRNAs (NSUN2, NUDT3, LARP1 and NCBP3), which can make individualized predictions of survival rate in UCEC patients and is an independent prognostic factor for UCEC patients. In addition, this risk model has a certain reference value for predicting immune cell infiltration, immune function, immune checkpoints, TMB levels, m6A-related genes, tumor immune microenvironment, stem cell correlation, TIDE scores and drug sensitivity

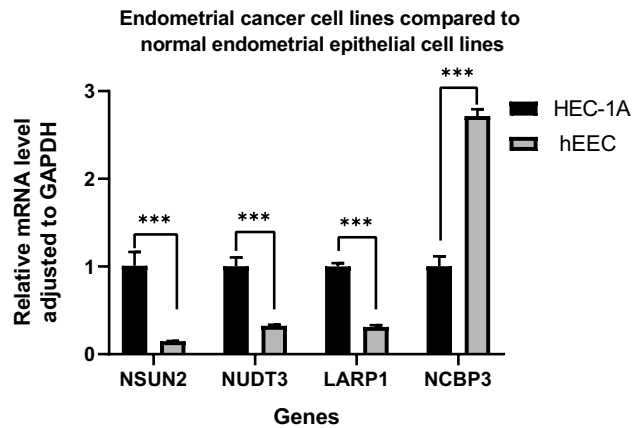


Figure 11. The result of RT-PCR.

in UCEC, indicating that targeting m7G-related mRNA may become a promising target for the treatment of UCEC. It will provide new directions and ideas for follow-up clinical diagnosis and treatment of UCEC, and has great clinical significance for UCEC-specific therapeutic drugs.

Data availability

The datasets generated and/or analyzed during the current study are available in the [The Cancer Genome Atlas (TCGA)] repository, [<https://portal.gdc.cancer.gov/>] and UCSC Xena repository [<https://xenabrowser.net/datapages/>].

Received: 10 May 2022; Accepted: 20 October 2022

Published online: 05 November 2022

References

- Crosbie, E. J. *et al.* Endometrial cancer. *Lancet* **399**(10333), 1412–1428. [https://doi.org/10.1016/S0140-6736\(22\)00323-3](https://doi.org/10.1016/S0140-6736(22)00323-3) (2022).
- Makker, V. *et al.* Endometrial cancer. *Nat. Rev. Dis. Primers* **7**(1), 88. <https://doi.org/10.1038/s41572-021-00324-8> (2021).
- Lu, K. H. & Broaddus, R. R. Endometrial cancer. *N. Engl. J. Med.* **383**(21), 2053–2064. <https://doi.org/10.1056/NEJMra1514010> (2020).
- Van Arsdale, A. *et al.* Association of obesity with survival in patients with endometrial cancer. *Gynecol. Oncol.* **154**(1), 156–162. <https://doi.org/10.1016/j.ygyno.2019.03.258> (2019).
- Francis, S. R. *et al.* Recurrent early stage endometrial cancer: Patterns of recurrence and results of salvage therapy. *Gynecol. Oncol.* **154**(1), 38–44. <https://doi.org/10.1016/j.ygyno.2019.04.676> (2019).
- de Boer, S. M. *et al.* Adjuvant chemoradiotherapy versus radiotherapy alone for women with high-risk endometrial cancer (PORTEC-3): Final results of an international, open-label, multicentre, randomised, phase 3 trial. *Lancet Oncol.* **19**(3), 295–309. [https://doi.org/10.1016/S1470-2045\(18\)30079-2](https://doi.org/10.1016/S1470-2045(18)30079-2) (2018).
- Morice, P., Leary, A., Creutzberg, C., Abu-Rustum, N. & Darai, E. Endometrial cancer. *Lancet* **387**(10023), 1094–1108. [https://doi.org/10.1016/S0140-6736\(15\)00130-0](https://doi.org/10.1016/S0140-6736(15)00130-0) (2016).
- Furuichi, Y. Discovery of m(7)G-cap in eukaryotic mRNAs. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **91**(8), 394–409. <https://doi.org/10.2183/pjab.91.394> (2015).
- Cartlidge, R. A. *et al.* The tRNA methylase METTL1 is phosphorylated and inactivated by PKB and RSK in vitro and in cells. *EMBO J.* **24**(9), 1696–1705. <https://doi.org/10.1038/sj.emboj.7600648> (2005).
- Létoquart, J. *et al.* Structural and functional studies of Bud23-Trm112 reveal 18S rRNA N7–G1575 methylation occurs on late 40S precursor ribosomes. *Proc. Natl. Acad. Sci. U.S.A.* **111**(51), E5518–E5526. <https://doi.org/10.1073/pnas.1413089111> (2014).
- Cowling, V. H. Regulation of mRNA cap methylation. *Biochem. J.* **425**(2), 295–302. <https://doi.org/10.1042/BJ20091352> (2009).
- Cole, M. D. & Cowling, V. H. Specific regulation of mRNA cap methylation by the c-Myc and E2F1 transcription factors. *Oncogene* **28**(9), 1169–1175. <https://doi.org/10.1038/onc.2008.463> (2009).
- Guy, M. P. & Phizicky, E. M. Two-subunit enzymes involved in eukaryotic post-transcriptional tRNA modification. *RNA Biol.* **11**(12), 1608–1618. <https://doi.org/10.1080/15476286.2015.1008360> (2014).
- Shaheen, R. *et al.* Mutation in WDR4 impairs tRNA m(7)G46 methylation and causes a distinct form of microcephalic primordial dwarfism. *Genome Biol.* **16**, 210. <https://doi.org/10.1186/s13059-015-0779-x> (2015).
- Lindstrom, D. L. *et al.* Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol. Cell Biol.* **23**(4), 1368–1378. <https://doi.org/10.1128/MCB.23.4.1368-1378.2003> (2003).
- Drummond, D. R., Armstrong, J. & Colman, A. The effect of capping and polyadenylation on the stability, movement and translation of synthetic messenger RNAs in *Xenopus* oocytes. *Nucleic Acids Res.* **13**(20), 7375–7394. <https://doi.org/10.1093/nar/13.20.7375> (1985).
- Murthy, K. G., Park, P. & Manley, J. L. A nuclear micrococcal-sensitive, ATP-dependent exoribonuclease degrades uncapped but not capped RNA substrates. *Nucleic Acids Res.* **19**(10), 2685–2692. <https://doi.org/10.1093/nar/19.10.2685> (1991).
- Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
- Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
- Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).
- Warda, A. S. *et al.* Human METTL16 is a N6-methyladenosine (m6A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. *EMBO Rep.* **18**(11), 2004–2014. <https://doi.org/10.15252/embr.201744940> (2017).

22. Haussmann, I. U. *et al.* m6A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. *Nature* **540**(7632), 301–304. <https://doi.org/10.1038/nature20577> (2016).
23. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**(7481), 117–120. <https://doi.org/10.1038/nature12730> (2014).
24. Yang, X. *et al.* 5-methylcytosine promotes mRNA export—NSUN2 as the methyltransferase and ALYREF as an m5C reader. *Cell Res.* **27**(5), 606–625. <https://doi.org/10.1038/cr.2017.55> (2017).
25. Nusinow, D. P. *et al.* Quantitative proteomics of the cancer cell line encyclopedia. *Cell* **180**(2), 387–402. <https://doi.org/10.1016/j.cell.2019.12.023> (2020).
26. Bhat, M. *et al.* Targeting the translation machinery in cancer. *Nat. Rev. Drug Discov.* **14**(4), 261–278. <https://doi.org/10.1038/nrd4505> (2015).
27. Xie, S. *et al.* Emerging roles of RNA methylation in gastrointestinal cancers. *Cancer Cell Int.* **20**(1), 585. <https://doi.org/10.1186/s12935-020-01679-w> (2020).
28. Han, H. *et al.* N7-methylguanosine tRNA modification promotes esophageal squamous cell carcinoma tumorigenesis via the RPTOR/ULK1/autophagy axis. *Nat. Commun.* **13**(1), 1478. <https://doi.org/10.1038/s41467-022-29125-7> (2022).
29. Zhang, Q. *et al.* Malignant evaluation and clinical prognostic values of M6A RNA methylation regulators in prostate cancer. *J. Cancer* **12**(12), 3575–3586. <https://doi.org/10.7150/jca.55140> (2021).
30. Ivanov, P., Emará, M. M., Villen, J., Gygi, S. P. & Anderson, P. Angiogenin-induced tRNA fragments inhibit translation initiation. *Mol. Cell* **43**(4), 613–623. <https://doi.org/10.1016/j.molcel.2011.06.022> (2011).
31. Sobala, A. & Hutvagner, G. Small RNAs derived from the 5' end of tRNA can inhibit protein translation in human cells. *RNA Biol.* **10**(4), 553–563. <https://doi.org/10.4161/rna.24285> (2013).
32. Spriggs, K. A., Bushell, M. & Willis, A. E. Translational regulation of gene expression during conditions of cell stress. *Mol. Cell* **40**(2), 228–237. <https://doi.org/10.1016/j.molcel.2010.09.028> (2010).
33. Tuorto, F. *et al.* RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat. Struct. Mol. Biol.* **19**(9), 900–905. <https://doi.org/10.1038/nsmb.2357> (2012).
34. Blanco, S. *et al.* Stem cell function and stress response are controlled by protein synthesis. *Nature* **534**(7607), 335–340. <https://doi.org/10.1038/nature18282> (2016).
35. Yang, J. C. *et al.* Association of tRNA methyltransferase NSUN2/IGF-II molecular signature with ovarian cancer survival. *Future Oncol.* **13**(22), 1981–1990. <https://doi.org/10.2217/fon-2017-0084> (2017).
36. McLennan, A. G. The Nudix hydrolase superfamily. *Cell Mol. Life Sci.* **63**(2), 123–143. <https://doi.org/10.1007/s00018-005-5386-7> (2006).
37. Grudzien-Nogalska, E., Jiao, X., Song, M. G., Hart, R. P. & Kiledjian, M. Nudt3 is an mRNA decapping enzyme that modulates cell migration. *RNA* **22**(5), 773–781. <https://doi.org/10.1261/rna.055699.115> (2016).
38. Hao, Q. *et al.* Comp34 displays potent preclinical antitumor efficacy in triple-negative breast cancer via inhibition of NUDT3-AS4, a novel oncogenic long noncoding RNA. *Cell Death Dis.* **11**(12), 1052. <https://doi.org/10.1038/s41419-020-03235-w> (2020).
39. Hopkins, T. G. *et al.* The RNA-binding protein LARP1 is a post-transcriptional regulator of survival and tumorigenesis in ovarian cancer. *Nucleic Acids Res.* **44**(3), 1227–1246. <https://doi.org/10.1093/nar/gkv1515> (2016).
40. Mura, M. *et al.* LARP1 post-transcriptionally regulates mTOR and contributes to cancer progression. *Oncogene* **34**(39), 5025–5036. <https://doi.org/10.1038/ncr.2014.428> (2015).
41. Xie, C. *et al.* LARP1 predict the prognosis for early-stage and AFP-normal hepatocellular carcinoma. *J. Transl. Med.* **11**, 272. <https://doi.org/10.1186/1479-5876-11-272> (2013).
42. Ye, L. *et al.* Overexpression of LARP1 predicts poor prognosis of colorectal cancer and is expected to be a potential therapeutic target. *Tumour Biol.* **37**(11), 14585–14594. <https://doi.org/10.1007/s13277-016-5332-3> (2016).
43. Mazza, C., Ohno, M., Segref, A., Mattaj, I. W. & Cusack, S. Crystal structure of the human nuclear cap binding complex. *Mol. Cell.* **8**(2), 383–396. [https://doi.org/10.1016/s1097-2765\(01\)00299-4](https://doi.org/10.1016/s1097-2765(01)00299-4) (2001).
44. Schulze, W. M., Stein, F., Rettel, M., Nanao, M. & Cusack, S. Structural analysis of human ARS2 as a platform for co-transcriptional RNA sorting. *Nat. Commun.* **9**(1), 1701. <https://doi.org/10.1038/s41467-018-04142-7> (2018).
45. Zhang, H. *et al.* NCBP1 promotes the development of lung adenocarcinoma through up-regulation of CUL4B. *J. Cell Mol. Med.* **23**(10), 6965–6977. <https://doi.org/10.1111/jcmm.14581> (2019).
46. Li, X. *et al.* NCBP3/SNHG6 inhibits GBX2 transcription in a histone modification manner to facilitate the malignant biological behaviour of glioma cells. *RNA Biol.* **18**(1), 47–63. <https://doi.org/10.1080/15476286.2020.1790140> (2021).
47. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA Cancer J. Clin.* **69**(1), 7–34. <https://doi.org/10.3322/caac.21551> (2019).
48. Amant, F., Mirza, M. R., Koskas, M. & Creutzberg, C. L. Cancer of the corpus uteri. *Int. J. Gynaecol. Obstet.* **143**(Suppl 2), 37–50. <https://doi.org/10.1002/ijgo.12612> (2018).
49. De Felice, F., Marchetti, C., Tombolini, V. & Panici, P. B. Immune check-point in endometrial cancer. *Int. J. Clin. Oncol.* **24**(8), 910–916. <https://doi.org/10.1007/s10147-019-01437-7> (2019).
50. Gómez-Raposo, C., Merino Salvador, M., Aguayo Zamora, C., García de Santiago, B. & Casado Sáenz, E. Immune checkpoint inhibitors in endometrial cancer. *Crit. Rev. Oncol. Hematol.* **161**, 103306. <https://doi.org/10.1016/j.critrevonc.2021.103306> (2021).
51. Mehnert, J. M. *et al.* Immune activation and response to pembrolizumab in POLE-mutant endometrial cancer. *J. Clin. Investig.* **126**(6), 2334–2340. <https://doi.org/10.1172/JCI84940> (2016).
52. Jiang, F., Hu, Y., Liu, X., Wang, M. & Wu, C. Methylation pattern mediated by m6A regulator and tumor microenvironment invasion in lung adenocarcinoma. *Oxid. Med. Cell Longev.* <https://doi.org/10.1155/2022/2930310> (2022).
53. Jiang, F. *et al.* A hypoxia gene-based signature to predict the survival and affect the tumor immune microenvironment of osteosarcoma in children. *J. Immunol. Res.* **2021**, 5523832. <https://doi.org/10.1155/2021/5523832> (2021).
54. Henning, A. N., Roychoudhuri, R. & Restifo, N. P. Epigenetic control of CD8+ T cell differentiation. *Nat. Rev. Immunol.* **18**(5), 340–356. <https://doi.org/10.1038/nri.2017.146> (2018).
55. Hodgins, J. J., Khan, S. T., Park, M. M., Auer, R. C. & Ardolino, M. Killers 2.0: NK cell therapies at the forefront of cancer control. *J. Clin. Investig.* **129**(9), 3499–3510. <https://doi.org/10.1172/JCI129338> (2019).
56. Liu, C., Zhang, Y. & Hang, C. Identification of molecular subtypes premised on the characteristics of immune infiltration of endometrial cancer. *Ann. Transl. Med.* **10**(6), 337. <https://doi.org/10.21037/atm-22-301> (2022).
57. Li, B. L. & Wan, X. P. Prognostic significance of immune landscape in tumour microenvironment of endometrial cancer. *J. Cell Mol. Med.* **24**(14), 7767–7777. <https://doi.org/10.1111/jcmm.15408> (2020).
58. Lee, D. W. *et al.* Tumor mutation burden and prognosis in patients with colorectal cancer treated with adjuvant fluoropyrimidine and oxaliplatin. *Clin. Cancer Res.* **25**(20), 6141–6147. <https://doi.org/10.1158/1078-0432.CCR-19-1105> (2019).
59. Goodman, A. M. *et al.* Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol. Cancer Ther.* **16**(11), 2598–2608. <https://doi.org/10.1158/1535-7163.MCT-17-0386> (2017).
60. Heeke, S. *et al.* In-house Implementation of tumor mutational burden testing to predict durable clinical benefit in non-small cell lung cancer and melanoma patients. *Cancers (Basel)* **11**(9), 1271. <https://doi.org/10.3390/cancers11091271> (2019).
61. Crawford, E. D. *et al.* Androgen receptor targeted treatments of prostate cancer: 35 years of progress with antiandrogens. *J. Urol.* **200**(5), 956–966 (2018).

62. Yang, H. *et al.* Identification of m6A-related lncRNA to predict the prognosis of patients with hepatocellular carcinoma. *Biomed. Res. Int.* **2022**, 4169150. <https://doi.org/10.1155/2022/4169150> (2022).
63. Gao, L. *et al.* A risk model based on autophagy-related lncRNAs for predicting prognosis and efficacy of immunotherapy and chemotherapy in gastric cancer patients. *Aging* **13**(23), 25453–25465. <https://doi.org/10.18632/aging.203765> (2021).

Acknowledgements

Thanks to everyone who contributed to this article.

Author contributions

J.D.Z. and J.N.Z. conceived and designed the study; J.N.Z. and W.J.J. searched the related articles; J.D.Z., W.J.J. and L.D.L. analyzed the data; J.D.Z., J.N.Z. and L.D.L. wrote the manuscript. Z.H.L. and J.L.W. supervised the whole process. All authors read and approved the final manuscript.

Funding

The APC was funded by Zhiheng Lin.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-22879-6>.

Correspondence and requests for materials should be addressed to J.W. or Z.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022