



RNA modification writer expression profiles predict clinical outcomes and guide neoadjuvant immunotherapy in non-small cell lung cancer

Bolun Zhou,¹ Fenglong Bie,¹ Ruochuan Zang, Moyan Zhang, Peng Song, Lei Liu, Yue Peng, Guangyu Bai, Jun Zhao,* and Shugeng Gao*

Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Summary

Background RNA modifications, including adenosine-to-inosine RNA editing, alternative polyadenylation, m¹A and m⁶A, play a significant role in tumorigenesis and tumor immunity. However, the functions of RNA modification enzymes (writers) in immunotherapy and tumor microenvironment (TME) remain unknown.

Methods Nonnegative matrix factorization clustering was applied to identify RNA modification clusters in lung adenocarcinoma, one of the most prevalent subtypes of non-small cell lung cancer (NSCLC). CIBERSORT and ESTIMATE algorithms were performed to depict TME characteristics. Additionally, a scoring system called Writer-Score was established to quantify RNA modification patterns and subsequently predict clinical outcomes. We subsequently used RNA sequencing, targeted DNA sequencing and multiplex immunofluorescence to further evaluate the efficacy of Writer-Score in NSCLC patients receiving neoadjuvant immunotherapy.

Findings We identified three distinct RNA modification clusters and two DEGclusters, which were shown to be strongly associated with a variety of TME features and biological processes. Additionally, the Writer-Score served as an important factor in post-transcriptional events and immunotherapy. The Writer-Score was capable of properly predicting the prognosis of NSCLC patients receiving neoadjuvant PD-1 inhibitor therapy.

Interpretation Our work systematically analyzed four types of RNA modifications and constructed a scoring system to guide neoadjuvant immunotherapy in NSCLC, which highlighted the writers' roles in post-transcriptional events, TME and neoadjuvant immunotherapy.

Funding A full list of funding bodies that supported this study can be found in the Acknowledgements section.

Copyright © 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: RNA modification; Non-small cell lung cancer; Tumor microenvironment; Neoadjuvant immunotherapy

Introduction

Lung cancer is one of the most fatal and widespread cancers worldwide, with non-small cell lung cancer (NSCLC) accounting for around eighty percent of all occurrences.^{1,2} Previously published research revealed that the accumulation of genetic mutations, such as EGFR, KRAS, and ALK, contributed to the development of NSCLC.³ Apart from these, epigenetic alterations in certain genes may also contribute to the progression of NSCLC. And one of the essential parts of epigenetic

regulation, RNA modification, serves as a crucial factor in a variety of pathological processes, including cancer progression.⁴

Generally, more than one hundred types of modifications exist in RNA levels, but it is hard to enroll all of them in one individual study. The most heavily modified nucleotide on RNA is adenine and the most prevalent RNA modification is m⁶A.^{5,6} Some interactions also possibly exist among m⁶A and other RNA modifications. One previous study has revealed that the A-to-I modification can be negatively regulated by the m⁶A modification writer.⁷ Herein, we focused on adenine-related RNA modifications, including A-to-I RNA editing, alternative polyadenylation (APA), m¹A methylation and m⁶A methylation, all of which are mediated by

*Corresponding authors.

E-mail addresses: Drzhaojun@126.com (J. Zhao), gaoshugeng@cicams.ac.cn (S. Gao).

¹ Contributing equally.

eBioMedicine 2022;84:
104268

Published online xxx
<https://doi.org/10.1016/j.ebiom.2022.104268>

Research in context

Evidence before this study

We have searched the terms “[("model" OR "prediction" OR "immune model" OR "predictive model" OR "RNA modification" OR "RNA modification model" OR "prognostic model" OR "prognosis" OR "signature" OR "immune signature" OR "prognostic signature" OR "RNA modification signature") AND ("NSCLC" OR "non-small cell lung cancer" OR "lung cancer")]" via the PubMed on Jan.8, 2022. Some of the studies have investigated and constructed an immune-related model in NSCLC. However, none of the studies have focused on the relationship between RNA modification and tumor immunity, and no study has developed a prognostic scoring system for patients receiving neoadjuvant immunotherapy.

Added value of this study

Our study reveals distinct TME characteristics of different phenotypes classified by the RNA modification, suggesting a robust interaction between RNA modification and tumor immunity in NSCLC. We presented the RNA-seq data of NSCLC patients undergoing neoadjuvant PD-1 inhibitor treatment with a longtime follow-up, and assessed the function of RNA modification in neoadjuvant immunotherapy. Furthermore, we established a scoring system to quantify the RNA modification patterns and the immune infiltrating levels in individual cohorts, which could predict the clinical outcomes of patients receiving neoadjuvant immunotherapy.

Implications of all the available evidence

Our work provides a comprehensive assessment of the relationship between RNA modification writers and immune infiltration and constructs a reliable and accurate model to predict the prognosis of NSCLC patients receiving neoadjuvant PD-1 inhibitor treatment, providing important insights into the development of neoadjuvant immunotherapy.

particular enzymes (also called “writers”). The m⁶A modification usually functions in regulating translation, splicing and processing of the RNA, and has been linked to a variety of physiological and pathological processes.⁸ The m¹A can affect the stability and function of tRNA and rRNA, as well as hinder the translation of associated proteins.⁹ The A-to-I RNA editing (adenosine-to-inosine) can alter RNA transcripts’ nucleotides, increasing the complexity of the transcriptome.¹⁰ And the APA has the potential to control oncogene expression, remodel specific cellular pathways and interact with other biological processes.¹¹ However, the interplay of these four types of RNA modification writers in NSCLC is yet unknown, especially in the neoadjuvant immunotherapy cohort.

Immunotherapies involving immune checkpoint inhibitors (ICIs) have been conducted and shown to assist NSCLC patients.^{12,13} Recently, clinical studies of neoadjuvant ICI treatment were initiated to gain insight into its applicability in resectable lung cancer.^{14,15} Although the efficacy of neoadjuvant mono-immunotherapy has been reported throughout the years, no study has indicated the long-term prognosis for NSCLC patients. Numerous patients who underwent neoadjuvant immunotherapy experienced poor clinical outcomes, and we currently lack the prognostic biomarkers for their prognoses. Thus, extensively investigating the mechanisms underlying the tumor microenvironment (TME) in NSCLC patients is significant.¹⁶ Previous studies have established a striking correlation of TME with diverse types of RNA modifications. For example, Wang et al. have demonstrated that depleting METTL3/14 could increase the immune responses of anti-PD-1 therapy, as well as the secretion of CXCL10 and IFN- γ of the TME.¹⁷ Thus, it is worthwhile to explore the mechanism by which various RNA modification writers impact immune infiltrating cells of the TME, as this may aid in the improvement of efficient immunotherapeutic strategies.

In the present study, using genomic data of 1317 lung adenocarcinoma (LUAD) samples from the GEO and TCGA database and 30 NSCLC patients receiving neoadjuvant PD-1 inhibitor therapy from our hospital, we comprehensively evaluated the phenotypes of RNA modification and the correlation of multiple related writers with the TME. Three diverse RNA modification patterns related to different immune infiltrating levels were identified, suggesting RNA modification significantly affects the TME. The Writer-Score scoring system was established and performed well in the normal cohorts and NSCLC cohort receiving neoadjuvant PD-1 inhibitor therapy. To conclude, we showed the indispensable utility of the Writer-Score in differentiating the post-transcriptional events and its applicability in immunotherapy.

Methods

Ethics

The Ethics Committee and Institutional Review Boards of National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College approved this study (Approval No.17-151/1407), and the signed informed consents were obtained from all patients.

Collect and process available datasets

We got access to the TCGA database via UCSC Xena (<https://xena.ucsc.edu/>) and the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>) to download genomic data

and clinical annotations of LUAD patients. The copy number alteration data of the TCGA-LUAD cohort were obtained from the cBioportal database (<https://www.cbioportal.org/>). A total of 1317 patients from TCGA-LUAD and seven GEO cohorts were included^{18–24} (Table S1). The RNA-seq data were transformed to the transcripts per kilobase million (TPM) and $\log_2(\text{TPM} + 1)$ format. The same microarray platform was used in five GEO datasets (GSE29013, GSE30219, GSE31210, GSE37745 and GSE50081), and we used the “*gcrma*” R package (version 2.62.0) and the *sva* R package (version 3.38.0) to normalize and combine the five datasets into a meta-GEO cohort.

We also enrolled forty NSCLC patients receiving neoadjuvant PD-1 inhibitor and following surgery in the National Cancer Center in a clinical trial (ChiCTR-OIC-17013726).¹⁴ The detailed information about this cohort has been previously published.¹⁴ Thirty patients with RNA-seq data were enrolled in the NCC (National Cancer Center) cohort in this study.

Consensus clustering with NMF

The first nonnegative matrix factorization (NMF) clustering was performed using 25 RNA modification writers' expression levels (Table S2). The second NMF clustering was conducted according to overlapping differentially expressed genes (DEGs) among three clusters of the first clustering, which subsequently resulted in the DEGclusters. We utilized the “*NMF*” R package (version 0.23.0) with the lee algorithm. Next, a multivariate Cox regression analysis was conducted using the “*survival*” R package (version: 3.2-10) to investigate the correlations of different clinical parameters with prognosis. And principal component analysis (PCA) was performed using the “*psych*” R package (version 2.0.12) based on the expression of RNA modification writers.

GSEA and functional enrichment analysis

We performed the Gene Set Enrichment Analysis (GSEA) via the “*clusterProfiler*” R package (version 3.18.1). The functional enrichment analysis was conducted using the “*clusterProfiler*” R package (version 3.18.1). $P_{\text{adjust}} < 0.05$ was statistically significant in the Gene Ontology (GO) analysis and GSEA, and $P < 0.05$ & $P_{\text{adjust}} < 0.2$ in the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis.

Evaluation of immune infiltration

We utilized ESTIMATE algorithm to analyze the immune infiltrating levels, which was also capable of inferring the stromal score and immune score.²⁵ We quantified the StromalScore, ImmuneScore and ESTIMATEScore to show the different immune characteristics of these clusters. We used the CIBERSORT to evaluate the proportion of 22 immune infiltrating cells

(<http://cibersort.stanford.edu/>). We calculated the Immunophenoscore (IPS) and showed intratumoral immune features and cancer antigenomes.²⁶

Establishment and validation of Writer-Score scoring system

According to the first NMF clustering in the meta-GEO cohort, we used the “*limma*” R package (version 3.46.0) to identify DEGs between every two clusters.²⁷ $P_{\text{adjust}} < 0.05$ and $|\text{fold change}| > 1$ were set as the statistically significant threshold. We then identify the overlapping DEGs by intersecting DEGs between every two clusters.

In order to select overlapping DEGs correlated with prognosis, we used the “*survival*” R package (version: 3.2-10) to conduct the univariate Cox regression analysis in the meta-GEO cohort. Next, we used the LASSO regression analysis and screened for the genes enrolled in the Writer-Score scoring system. And the following formula was the Writer-Score: $\text{Writer-Score} = \sum (\text{Coef}_i * \text{Expr}_i)$, where i indicates the number of hub genes, Coef_i refers to each gene's coefficient from the LASSO analysis and Expr_i refers to the expression level of each gene. The format of RNA-seq data used in the establishment and validation of the scoring system was $\log_2(\text{TPM} + 1)$. The cut-off value of high- and low-score groups was defined using the “*survminer*” R package (version 0.4.9). We then got Kaplan–Meier curves with the Log-rank test in the training dataset (meta-GEO cohort) and validation datasets.

The relationship between Writer-Score and miRNA

We downloaded the miRNA expression profiles from the TCGA via UCSC Xena. We then used the “*limma*” R package (version 3.46.0) to identify differentially expressed miRNAs between the high- and low-score groups.²⁷ $P_{\text{adjust}} < 0.05$ and $|\text{fold change}| > 1.5$ were set as the statistically significant threshold. And KEGG enrichment analysis was conducted to screen for targeted pathways among the targeted genes of the differentially expressed miRNAs.

RNA sequencing

A total of 30 NSCLC patients' tissue samples were processed into FFPE tissue using formalin fixation. The truXTRAC FFPE total na Kit (Covaris) was used to extract and purify total RNA from FFPE samples.

Based on Illumina HiSeq X Ten platform, total RNA was collected and subjected to high-throughput sequencing to generate RNA expression profile data. As the manufacturer's instruction previously reported, we used the Qiagen AllPrep DNA/RNA Mini Kit to isolate the total RNA from tumor tissue samples.²⁸ The Life Invitrogen Qubit3.0/4.0 was used to quantify the RNA and the Agilent 2100 Bioanalyzer system assay was used to assess the RNA. Then, based on the low-

throughput protocol, 50 ng of total RNA was utilized with the SMARTer Stranded Total RNA-Seq Kit v2 (Takara). Using PCR amplification and SMART[®] (Switching Mechanism At 5' End of RNA Template) cDNA synthesis technology, this kit builds Illumina-compatible libraries. The strand orientation of the original RNA was preserved by the template-switching reaction's directionality. The library containing adapters was determined using the Life Invitrogen Qubit3.0/4.0 and evaluated using the Agilent 2100 Bioanalyzer system assay. After that, the Illumina HiSeq X Ten Sequencing System was used to sequence the RNA.

Sequential data pre-processing and mapping

We removed rRNA removal, de-junction contamination, low quality sequences, etc to preprocess raw reads, in order to get high-quality sequences, which were then used in all following analyses. Using the ENSEMBL database, genome annotation files and reference genes were obtained. We used the HISAT2 to align clean data to the reference genome.²⁹ We used HTSeq to calculate gene expression levels (<http://www.huber.embl.de/users/anders/HTSeq/doc/overview.html>). The FPKM (Fragments Per Kilobase Million Mapped Reads) technique was used to calculate gene expression levels.

Targeted DNA sequencing

According to previous publication, we conducted DNA extraction, library preparation, sequencing, and somatic mutation calling.³⁰ With the formula: Absolute mutation count $\times 1000000$ / Panel exonic base num, we calculated TMB using the absolute mutation counts of the tumor samples. TMB was measured in mutations per Mb.

Multiplex immunofluorescence (mIF) staining assay

On paraffin-embedded NSCLC tissue blocks serially sectioned into 3 μ m sections, multiplex immunohistochemistry staining was conducted. We deparaffinized, rehydrated, and then boiled the slides for 20 minutes at 97°C in Tris-EDTA buffer (pH = 9; Clinipath #643901, the Netherlands). We then used Antibody Diluent/Block (PerkinElmer #72424205, USA) to inhibit endogenous peroxidase for 10 minutes. We detected only one antigen in every round, and labeling of the subsequent antibody following epitope retrieval and protein blocking as previously. At room temperature, we incubated primary antibodies for CD4 (1:50; ZM0418; ZSGB-Bio, China), CD8 (1:50; ZA-0508; ZSGB-Bio, China), LAG3 (1:100; ab40466; Abcam), TIM3 (1:100; CST45208S; CST), FOXP3 (1:400; ab20034; Abcam) and PD-1 (1:100; ZM0381; ZSGB-Bio, China) for 1 h. Following that, samples were incubated with the antibody at 37°C for 10 minutes using the Opal polymer anti-rabbit/mouse horseradish peroxidase (HRP) Kit (PerkinElmer

#2414515, USA). The opal multiplex immunohistochemistry kit (NEL797B001KT, PerkinElmer, USA) was subsequently used to visualize the TSA, which contains TSA Coumarin system (NEL703001KT, PerkinElmer, Massachusetts, USA), Opal 650 (FOXP3, PD-1), Opal 620 (CD4, LAG3), Opal 540 (CD8), Opal 690 (TIM3) and fluorophores (DAPI). After labeling each panel's antigens, we conducted microwave treatment (MWT) for 20 minutes at 97°C to eliminate the TSA-antibody complex with Tris-EDTA buffer (pH = 9; Klinipath #643901, the Netherlands). Using 4',6-Diamidino-2-Phenylindole (DAPI), we counterstained all slides for five minutes and used NobleRyder Antifade Mounting Medium (NobleRyder #10052, China) to prepare for imaging. Each staining batch contained a quality control sample of tonsil tissue from the autopsy as a positive control and to measure interexperimental repeatability.

Tissue imaging

The PerkinElmer Vectra imaging system (Vectra 3.0.5; PerkinElmer, USA) was used to scan the slides. After a low resolution (4x or 10x) scan, we used the Phenochart viewer (Akoya Bioscience) to select regions of interest (ROI). Next, at a higher resolution (20x), we scanned these ROIs and subjected them to inForm analysis. Using inForm Advanced Image Analysis software (inForm 2.3.0; PerkinElmer, Massachusetts, USA), we unmixed multispectral images based on spectral images constructed from images of tissues stained individually with each reagent. Using a selection of 5–10 representative original multispectral images, TheinForm software was trained. We estimated the percentage of positively stained cells by dividing the number of positively stained cells by the total number of nucleated cells.

Reagent validation

All the antibodies have been validated and the relevant documentation was provided in the Supplemental Data (Reagent Validation file).

Statistical analysis

All statistical analyses were performed using R (version 4.0.4) software. To compare the difference between the two groups, the Wilcoxon rank sum test was used. To assess the difference between three or more groups, the Kruskal–Wallis test was used. The correlation was analyzed using Spearman's technique of correlation. $P < 0.05$ was regarded as the threshold for statistical significance.

Role of funding source

The funders played no role in the design of the research, collection, analysis and interpretation of data, the

writing of the paper and the decision to submit the manuscript for publication.

Results

The genetic landscape of 25 RNA modification writers in LUAD

According to recent studies, we enrolled 25 RNA modification writers in our study.^{31–34} We first analyzed the somatic mutation frequency of these writers in LUAD. Of the 355 LUADs samples, 131 (36.9%) had somatic mutations of these writers (Figure 1a). Next, we evaluated the copy number alterations (CNAs) of the 25 RNA modification writers (Figure 1b). Specifically, ZC3H13 had the highest mutation frequency (4%), followed by PCF11 (3%) and RBM15 (2%). Moreover, the principal component analysis (PCA) was used and the results revealed that these 25 RNA modification writers stratified LUAD samples from normal samples (Figure 1c). Besides, we calculated correlations among 25 writers expression and the results showed that the majority of the correlations were positive (Figure 1d). Our results revealed that most of writers expression were elevated in the tumor samples (Figure 1e–h). In detail, we found that expression levels of writers with CNA gain were relatively high in tumor samples, suggesting that CNA may serve as an essential factor in regulating writers expression.

Identification of distinct patterns and correlated immune features based on RNA modification writers

In order to classify patients into distinct RNA modification phenotypes, we conducted the NMF clustering based on 25 RNA modification writers' expression in the meta-GEO cohort (Figure S1a–b). Three distinct clusters were identified, including Cluster1 (n = 192), Cluster2 (n = 191) and Cluster3 (n = 162) (Table S3). The expression levels of these writers were different (Figure 2a). The Kaplan-Meier curves revealed that patients of Cluster1 were correlated with the prominent survival advantage, whereas patients of Cluster3 had an unfavorable prognosis (Figure 2b, $P = 0.012$, Log-rank test). Furthermore, we utilized the multivariate Cox regression analysis and found that these clusters were associated with patients prognosis (Figure S1c).

The RNA modification clusters correlated with distinct immune features

We explored the immune landscape among the three clusters (Figure 2c, Table S4). According to the results, the proportion of T cells follicular helper, mast cells resting, T cells CD4 memory resting were the highest in Cluster1; the proportion of T cells CD4 naive, macrophages Mo, mast cells activated and T cells CD4 memory activated were the highest in Cluster2; the

proportion of dendritic cells resting and neutrophils were the highest in Cluster3 (Figure 2d). We then accessed the immune-related score through the ESTIMATE algorithm (Table S5). Compared with Cluster2, the results indicated that all of the scores of the Cluster1 and Cluster3 were relatively higher (Figure 2e).

Construction and validation of the Writer-Score scoring system

Among the three clusters, we subsequently found that 466 overlapping DEGs (Figure 3a). GO enrichment analysis was performed based on the DEGs (Figure 3b). We then performed NMF clustering in the meta-GEO cohort based on overlapping DEGs, and stratified patients into two phenotypes: DEGcluster1 and DEGcluster2 (Figure S2a–b). The results showed that patients with pathological stage IA were mostly enrolled in DEGcluster1, whereas patients with pathological stage II were predominantly classified by DEGcluster2 (Figure S2c). As for the prognostic value of these two DEGclusters, the Kaplan-Meier curves showed that patients in DEGcluster1 had a more favorable prognosis than patients in DEGcluster2 (Figure 3c). And we conducted the multivariate Cox regression analysis, indicating that pathological stage, age and DEGcluster were related to prognosis (Figure S2c). Furthermore, we compared the expression levels of PD-L1 and CTLA-4 between DEGcluster1 and DEGcluster2, indicating that expression of these immune checkpoints were remarkably higher in DEGcluster2 (Figure 3d).

Due to the complexity of the clusters based on the RNA modification writers, we established a scoring system (Writer-Score) based on the overlapping DEGs, aiming to quantify the RNA modification phenotypes of independent cohorts. 123 out of 243 patients in the Cluster2 overlap with the high-score group and 148 out of 302 patients in the Cluster1 overlap with the low-score group (Figure 3e). 180 out of 243 patients in the DEGcluster2 overlap with the high-score group and 270 out of 302 patients in the DEGcluster1 overlap with the low-score group (Figure 3f). The relationships among Cluster, DEGcluster and Writer-Score were analyzed (Figure 4a). We compared the distribution of Writer-Score in different clusters, which revealed that Writer-Score of Cluster2 was the highest and Writer-Score of Cluster1 was the lowest (Figure 4b). And the DEGcluster2 had a predominantly higher Writer-Score than DEGcluster1 (Figure 4c). The Writer-Score was established according to the results of the univariate Cox regression analysis (Table S6) and LASSO regression analysis (Figure 4d). The coefficients of the Writer-Score scoring system were obtained from the LASSO regression analysis (Table S7).

Next, we stratified patients into high- and low-score subgroups. We used the meta-GEO cohort as the training cohort and the results indicated that patients of the

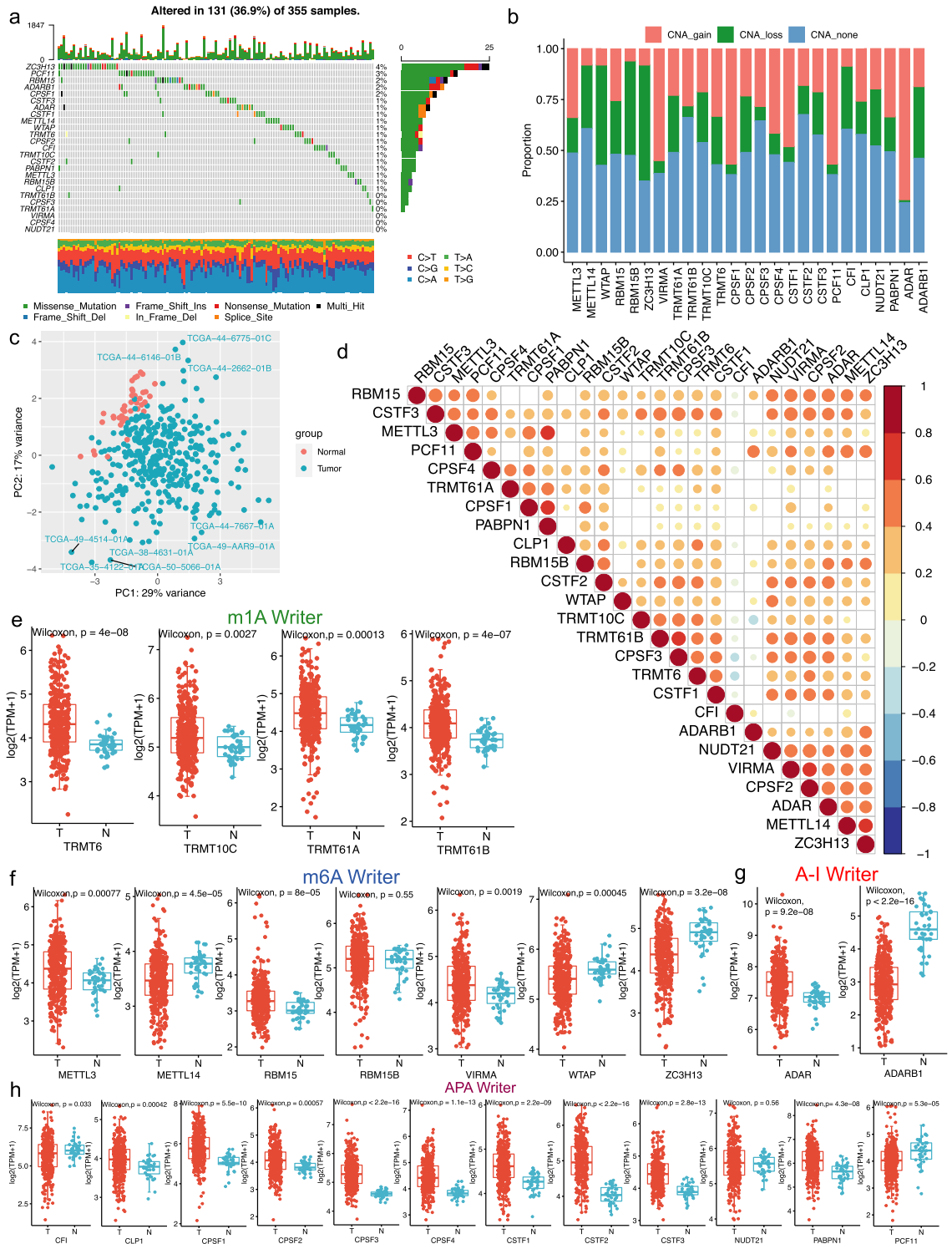


Figure 1. The transcriptional and genetic features of 25 RNA modification writers in LUAD (TCGA cohort). (a) 131 of 355 LUAD patients had genetic alterations of the above writers. The fraction of every variant type was presented in the right bar graph and the proportion of conversions was presented in the bottom bar graph. (b) The CNV mutation features of 25 RNA modification writers were shown in the bar graph. CNV loss referred to hemizygous and homozygous deletion, CNV gain referred to gain and high-level amplification. (c) Classification of cancer and normal samples via the principal component analysis of 25 RNA modification writers. *P < 0.05, **P < 0.01, and ***P < 0.001. (d) The Spearman correlation analysis among the 25 RNA modification writers. *P < 0.05, **P < 0.01, and ***P < 0.001. (e-h) The expression distribution of four kinds of RNA modification writers (m¹A, m⁶A, A-I and APA) between tumor and normal samples.

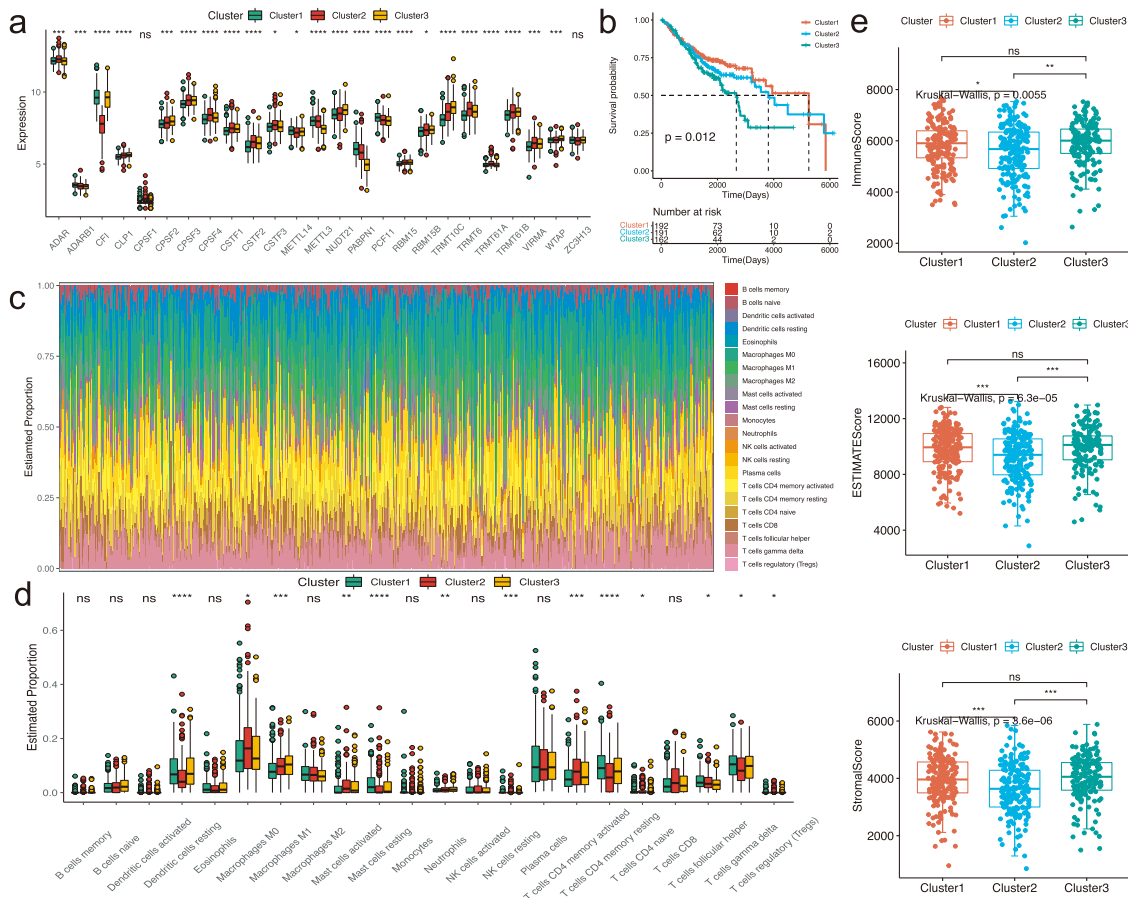


Figure 2. Different clusters according to 25 RNA modification writers and associated immune characteristics in the meta-GEO cohort. (a) The box plot revealed the expression levels of 25 RNA modification writers in three different clusters. Kruskal-Wallis H test was applied to compare the statistical difference of these clusters. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$. (b) Log-rank test was used to compare the overall survival among three different clusters and the results was visualized via the Kaplan-Meier curves. 192, 191 and 162 patients belonged to Cluster1, Cluster2 and Cluster3, respectively. (c) The estimated fraction of various immune infiltrating cells was calculated by the CIBERSORT algorithm. (d) The box plot revealed the estimated proportion of different immune infiltrating cells via the CIBERSORT algorithm among three clusters. The difference of the estimated fraction was compared using the Kruskal-Wallis H test. The scattered dots represented the fraction of different immune cells. The median proportion was represented by the thick line. The 25th and 75th percentiles were symbolized by the top and bottom of each box. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. (e) The comparison of ESTIMATEScore, ImmuneScore and StromalScore in three distinct clusters.

high-score subgroup revealed an unfavorable prognosis (Figure 4e). We then applied it in TCGA, GSE72094 and GSE13213 cohorts (Table S8-10). The results demonstrated that patients of the low-score subgroups were correlated with prominent survival benefits in all three cohorts (Figure 4f-h). According to recent publications, TMB may serve as a significant factor in predicting immunotherapy's response.³⁵ The results showed that the TMB of the high-score subgroup was remarkably higher than the low-score subgroup (Figure 4i). In addition, we found that the mutation frequency was relatively higher in the high-score group than the other (Figure 4j-k).

The role of Writer-Score in post-transcriptional regulation

We analyzed the association between the Writer-Score and processes related to RNA modification. We evaluated the pathways enriched by the miRNA-targeted genes, which showed that the miRNA-targeted genes were differentially expressed between the low- and high-score groups (Figure S3a). And we have noticed that MAPK signaling pathway, PD-L1 expression and PD-1 checkpoint pathway, microRNAs in cancer, PI3K-Akt signaling pathway and others were significantly correlated with the differentially expressed miRNA-targeted genes (Table S11). GO enrichment showed that these

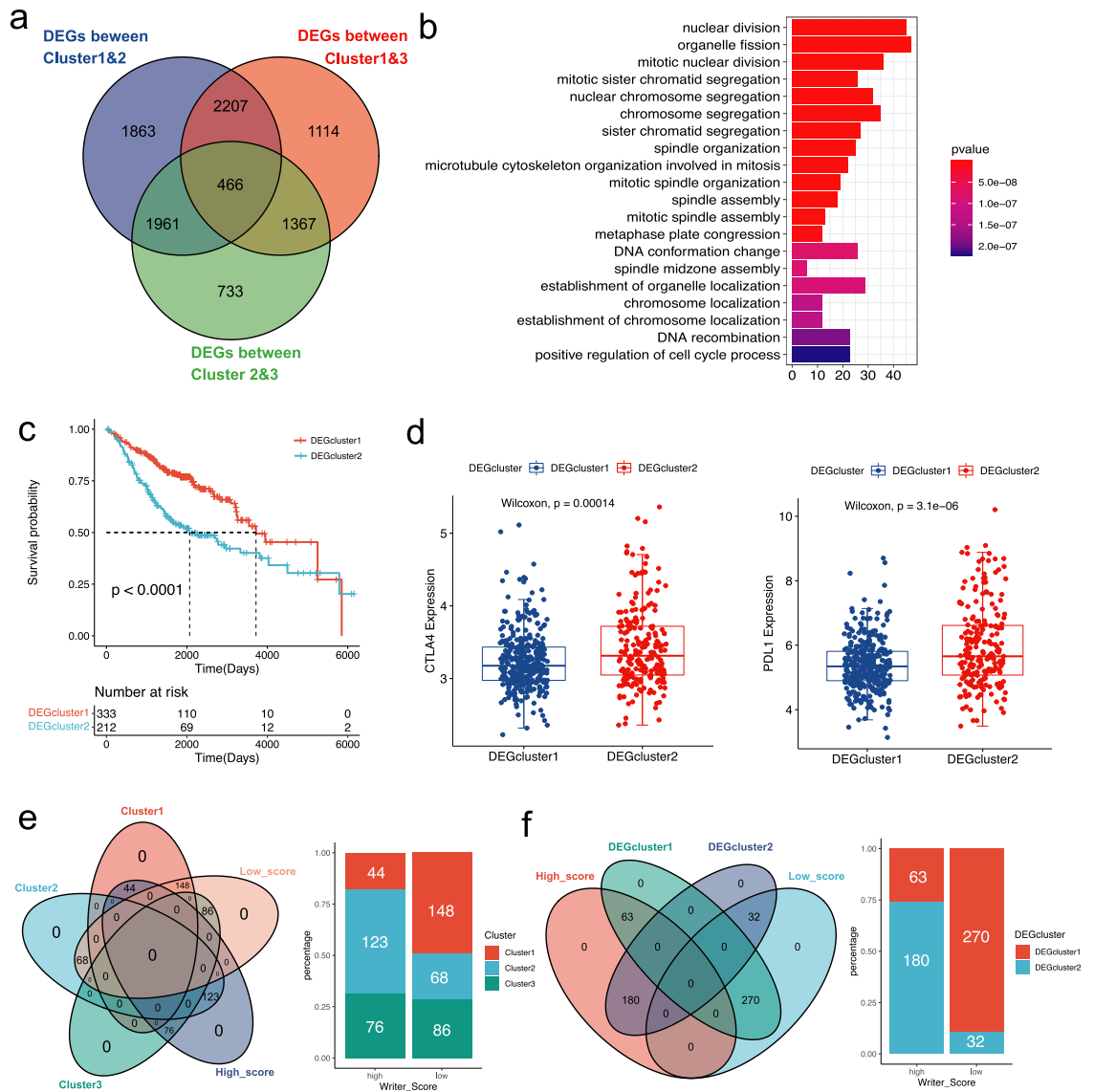


Figure 3. The analysis of NMF clustering based on differentially expressed genes (DEGs) among three clusters in the meta-GEO cohort. (a) The Venn diagram showed 466 overlap DEGs among the three clusters. (b) The GO enrichment analysis of the DEGs. The gene counts of each GO term were presented in the x-axis and the statistical significance was indicated by the brightness of the column color. (c) The prognosis of patients in two DEGclusters was compared in the Kaplan-Meier curves via the Log-rank test. (d) The distribution of CTLA-4 and PD-L1 between two DEGclusters. (e) The Venn diagram and bar plot revealed the relationship between the three clusters and two Writer-Score groups. (f) The Venn diagram and bar plot revealed the relationship between the two DEGclusters and two Writer-Score groups.

genes were significantly correlated with biological processes related to the development and regulation of dendrite (Figure S3b). In addition, our findings demonstrated that 63 miRNAs expressed differentially between the high- and low-score groups in the TCGA cohort (Figure S3c).

The role of Writer-Score in predicting the immunotherapeutic effect

IPS has been considered as a crucial factor to predict immunotherapeutic responses in some publications. Due to the close relationship between tumor microenvironment and Writer-Score, we further investigated

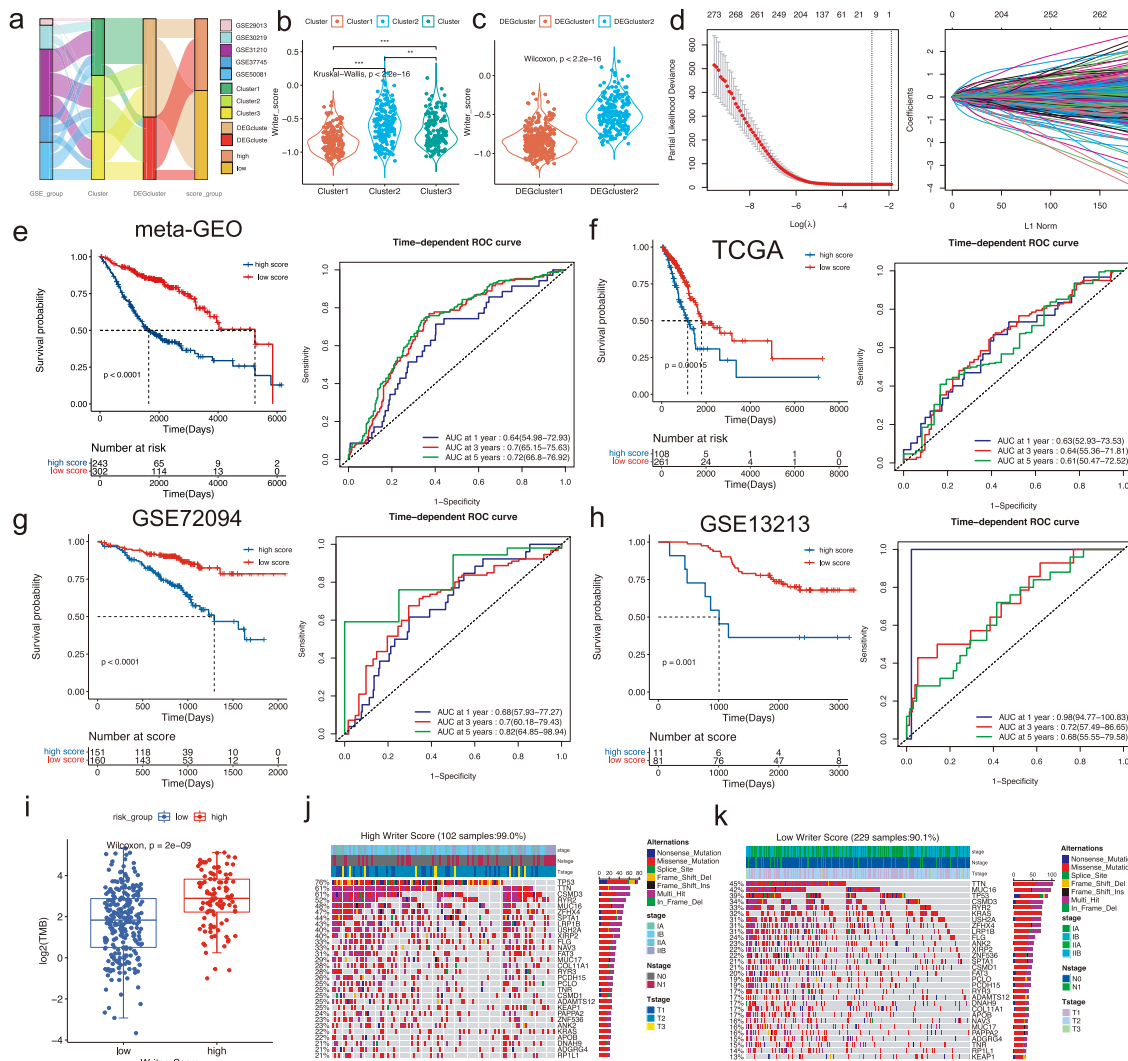


Figure 4. The construction and evaluation of the Writer-Score model. (a) The alluvial diagram of different clusters with various GSE groups, DEGclusters and Writer-Score groups. (b) The distribution diagram of Writer-Score in three distinct clusters. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. (c) The distribution diagram of Writer-Score in two distinct DEGclusters. (d) LASSO regression was applied to establish the Writer-Score model. (e-h) The Kaplan-Meier curves of high and low Writer-Score patients and their time-dependent ROC curves in meta-GEO, TCGA, GSE72094 and GSE13213 cohorts (Log-rank test). (i) The comparison of the tumor mutation burden (TMB) between high and low Writer-Score patients in the TCGA cohort. (j-k) Mutational landscape of the top mutational genes between high and low Writer-Score groups in the TCGA cohort. The upper bar plot referred to the annotations, including stage, Nstage and Tstage.

whether the Writer-Score can act as an important factor in predicting immunotherapeutic responses. We found that the Writer-Score was strongly correlated with the expression levels of PD-L1 and CTLA-4 (Figure 5a). And the expression levels of CTLA-4 and PD-L1 in the high-score subgroup were relatively higher in the meta-GEO cohort (Figure 5b). The differences of the IPS were also compared and the results replied that the IPS was higher in the low-score subgroup than the other (Figure 5c). In addition, GSEA in the high-score

subgroup revealed that chromosome segregation, mitotic cell cycle and cell cycle were the top three biological processes (Figure 5d).

Validation of Writer-Score in NCC cohort with neoadjuvant immunotherapy

Based on the RNA-seq data of the NSCLC patients receiving neoadjuvant PD-1 inhibitor treatment, we used the Writer-Score to predict the prognosis of these

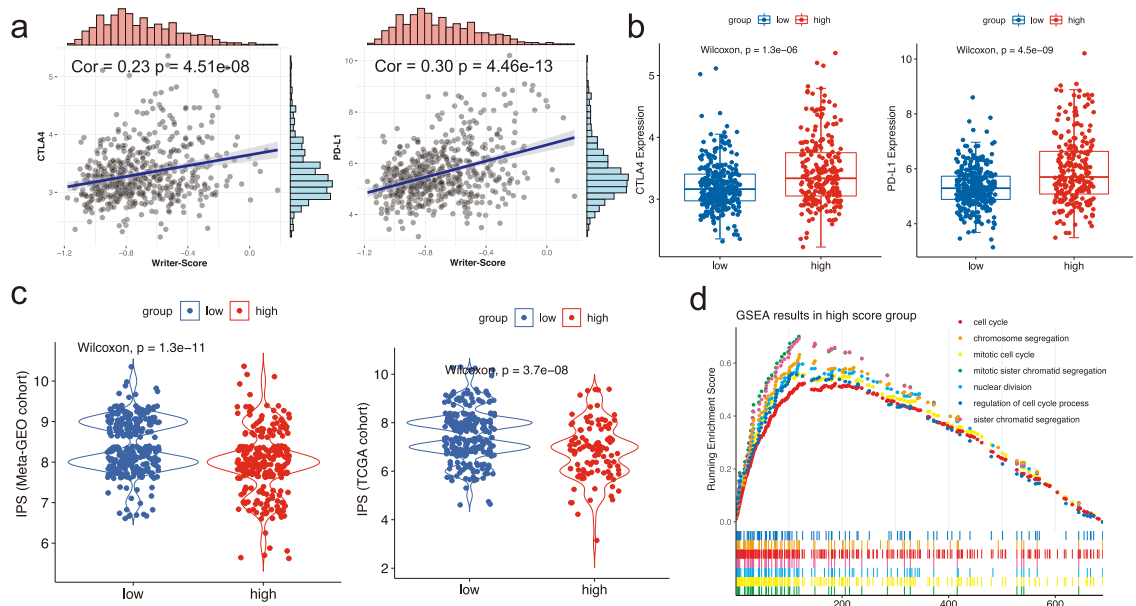


Figure 5. The relationship between the Writer-Score and tumor microenvironment and some related pathways in the meta-GEO cohort. (a) The Spearman correlation analysis of the Writer-Score with the expression levels of CTLA-4 and PD-L1. (b) The distribution of CTLA-4 and PD-L1 expression between the high and low Writer-Score subgroups. (c) The distribution of IPS scores between the high and low Writer-Score groups in meta-GEO and TCGA cohorts. (d) The results of the GSEA analysis in the high Writer-Score group.

patients (Figure 6a). Patients in the low-score group had more favorable disease-free survival (DFS), suggesting the good performance of the Writer-Score in NSCLC patients receiving neoadjuvant immunotherapy ($P = 0.021$, Figure 6b, Log-rank test). Responders or the patients with major pathological response (MPR) had relatively lower scores (Figure 6c-d). Furthermore, we evaluated the relationship between Writer-Score and various ICIs, implying Writer-Score was negatively correlated with most ICIs, such as PDCD1 and CD27 (Figure 6e). We also compared the mutation profiles of high- and low-score groups. TP53, POLD1 and KMT2D were the top genes in the high-score group, while TP53, NFE2L2 and CUL3 were those in another group (Figure 6f). In addition, the tumor mutation burden (TMB) was negatively correlated with Writer-Score in NCC cohort (Figure 6g).

To further reveal the TME of the high- and low-score groups, we used paraffin-embedded NSCLC tissues and conducted the multiplex immunofluorescence (mIF) immunohistochemical staining (Figure 7a). Our results showed that the proportion of CD8⁺ T cells of tumor was much higher in the low-score group. Particularly, the proportion of the non-exhausted CD8⁺ T cells (CD8⁺LAG3⁻, CD8⁺TIM3⁻ and CD8⁺PD1⁻) was significantly higher in the low-score group with better prognosis, whereas exhausted CD8⁺ T cells (CD8⁺LAG3⁺, CD8⁺TIM3⁺ and CD8⁺PD1⁺) were insignificant. The

proportion of other immune cells, including CD4⁺ T cells and Tregs (CD4⁺FOXP3⁺), did not show any difference between low- and high-score groups (Figure 7b, Wilcoxon rank sum test).

Discussion

Emerging studies indicate that RNA modification plays a pivotal role in tumor immunity through multiple RNA modification writers.³⁶ For example, we have known that m⁶A, as the prevalent epigenetic modification of RNA, could modify tumor immunogenicity and immune infiltrating cells within the TME, hence affecting the responses and prognosis of patients receiving immunotherapy.³⁷ However, previous publications have concentrated on a small number of RNA modification writers, leaving the complex interaction among various RNA modification writers poorly understood. Furthermore, a major barrier to immunotherapy, especially neoadjuvant immunotherapy, is the absence of predictive biological markers that may be used to identify certain subpopulations with superior therapeutic outcomes. In neoadjuvant immunotherapy, the prognostic value of RNA modification remains largely unknown.

In the present study, we categorized three clusters related to RNA modification and developed a scoring system called the Writer-Score, which could stratify

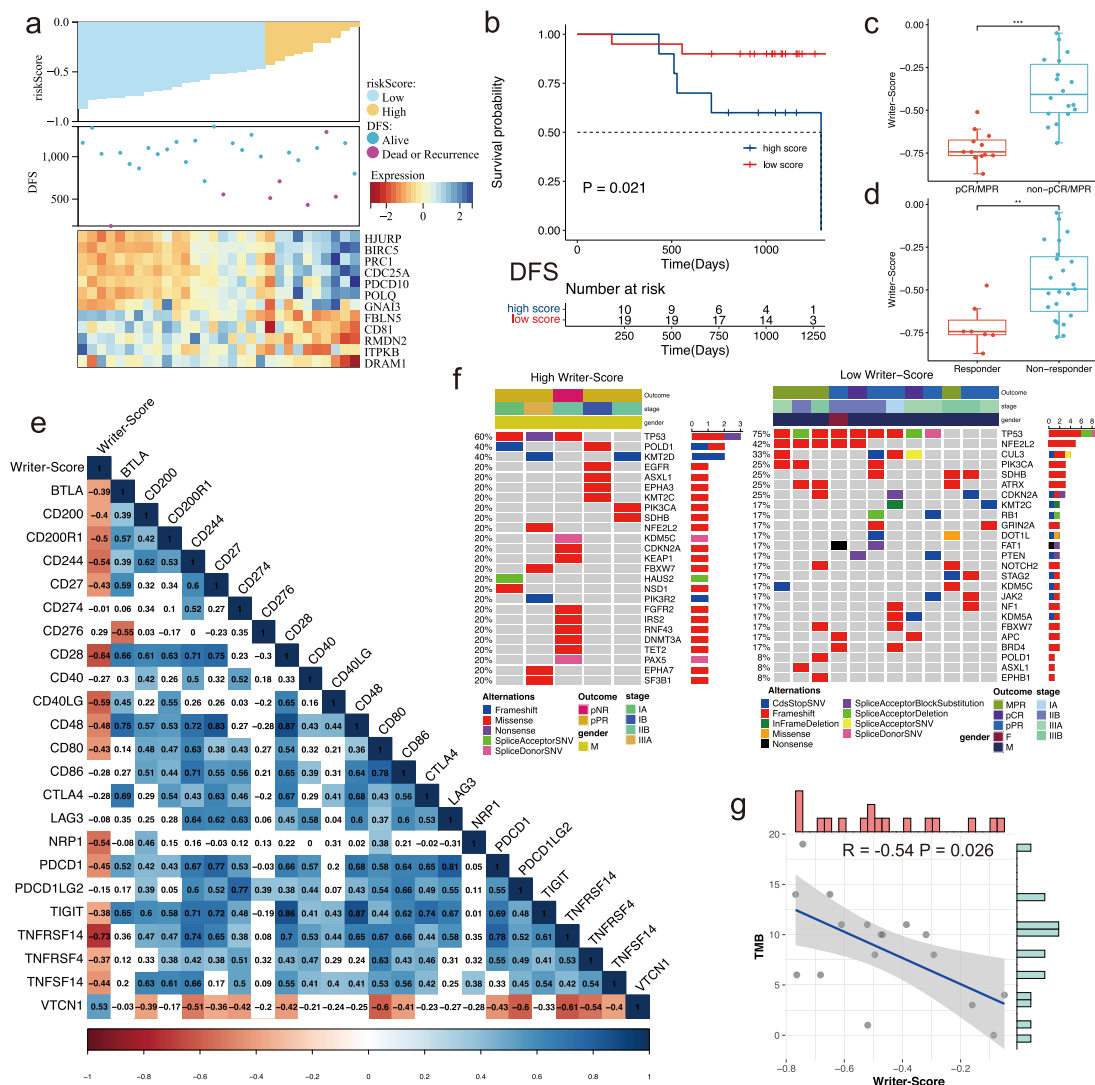


Figure 6. Validating the efficacy of the Writer-Score in the NCC cohort with neoadjuvant immunotherapy. (a) The distribution of Writer-Score in all patients and the heatmap of key genes in the NCC cohort. (b) The Kaplan-Meier curves indicating the difference of DFS between high- and low-score groups. DFS, disease-free survival. (c) The comparison of the Writer-Score between pCR/MPR group and the non-pCR/MPR group. pCR, pathological complete response; MPR, major pathological response. (d) The comparison of the Writer-Score between responder group and the non-responder group. (e) The correlation of the Writer-Score with different immune checkpoint inhibitors. (f) Mutational landscape of the top mutational genes between high and low Writer-Score groups in the NCC cohort. (g) The correlation of Writer-Score with TMB in the NCC cohort. TMB, tumor mutation burden.

patients into high- and low-score subgroups correlated with contrasting clinical outcomes and TME characteristics. Furthermore, Writer-Score performed well in predicting the prognosis of NSCLC patients receiving neoadjuvant immunotherapy and subsequent surgery. Additionally, the proportion of non-exhausted CD8⁺ T cells was considerably greater in low-score subgroup with better clinical outcomes, indicating the high accuracy and good performance of Writer-Score. Our research uncovers a previously unknown link between

RNA modification and tumor immunity, which aids to inform the clinical management of NSCLC.

Recent studies have implied that immune infiltrating cells of the TME may play a substantial role in carcinogenesis and influence immunotherapy responses.³⁸ And three different clusters in our study exhibit different immunological characteristics. Previous research has established a link between several kinds of immune infiltrating cells and various clinical outcomes in cancer. In patients with lung cancer receiving anti-PD-1

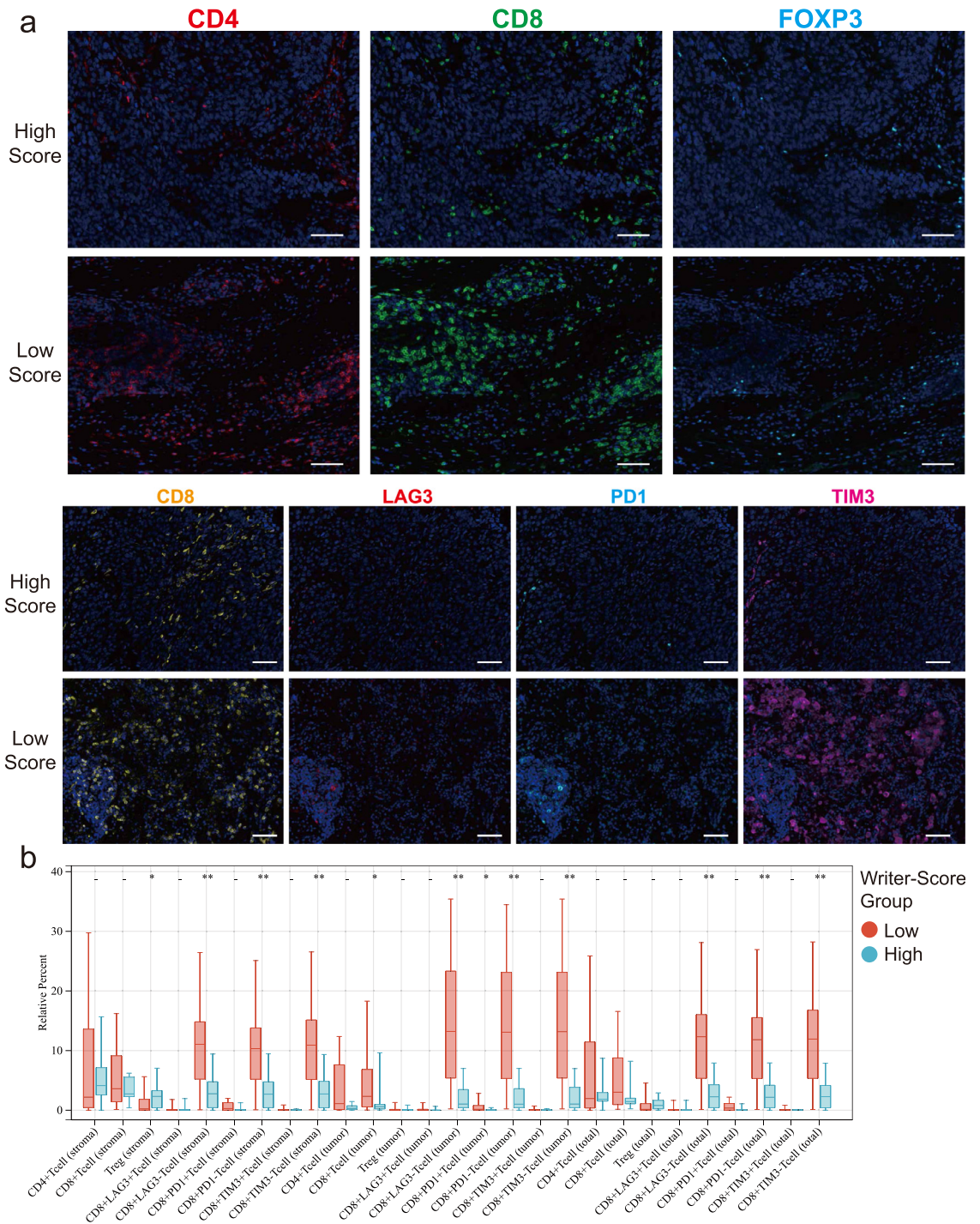


Figure 7. The comparison of different immune infiltrating cells within the TME. (a) The representative mIF staining figures of two patients from the high- and low-score groups (scale bar: 2mm). (b) The box plot depicting the difference of multiple immune infiltrating cells between the high- and low-score groups. *P < 0.05; **P < 0.01 (Wilcoxon rank sum test).

therapy, it was revealed that circulating T follicular helper cells increased the antibody production via promoting the activation of B cells, hence elevating

patients' response rate and survival time.³⁹ Cluster1 was proven to have a prominent survival advantage, which matched the findings of this investigation. Additionally,

neutrophils⁴⁰ and regulatory T cells⁴¹ could promote the progression and metastasis of lung cancer, which are related to unfavorable prognosis. Cluster3 had the same TME characteristics and poor clinical outcomes, indicating the potential power of the RNA modification clusters to predict the prognosis.⁴² In summary, our RNA modification patterns with specific immunological characteristics were proven to be reliable in LUAD, which might aid in the development of immunotherapy.

In contrast to Cluster1-3, we developed two DEG clusters using DEGs. Between the two DEG clusters, the difference in prognosis and immune characteristics was more significant. Moreover, we established the Writer-Score scoring system to stratify patients into two distinct subgroups with unique TME characteristics and converse clinical outcomes. The Writer-Score of Cluster3 with an unfavorable prognosis was higher than the score of Cluster1 with the best clinical outcome. Furthermore, we examine the predictive value of the Writer-Score in four independent cohorts, including our NCC cohort receiving neoadjuvant immunotherapy. The Writer-Score was found to be significantly effective in predicting prognosis of patients in validation cohorts. In addition, Chae et al. have discovered that TMB of the circulating tumor DNA was predominantly correlated with unfavorable clinical outcomes in NSCLC patients.⁴³ And we found that TMB was relatively higher in the high-risk subgroup of the TCGA cohort. Interestingly, we discovered that TMB was inversely correlated with Writer-Score in NCC cohort, implying TMB may be related to favorable prognosis among patients receiving immunotherapy. A research demonstrated that blood-based TMB was linked with good clinical outcomes in NSCLC patients treated with atezolizumab (a PD-L1 inhibitor), indicating that the predictive value of TMB may vary across NSCLC patients with and without immunotherapy.⁴⁴ Besides, differences in clinical characteristics of the two cohorts, such as race and sample size, may also lead to contradictory results.

Neoadjuvant immunotherapy using ICIs has recently emerged as a unique treatment for a variety of malignancies, including NSCLC. In comparison to chemotherapy, ICIs cause less adverse events and improve 1-year survival by 10% in patients with NSCLC.⁴⁵ Several clinical trials evaluating the efficacy of neoadjuvant immunotherapy for patients with resectable NSCLC have been completed, such as Checkmate-159,¹⁵ LCMC3⁴⁶ and ChiCTR-OIC-17013726.¹⁴ However, few study has uncovered the long-term follow-up, let alone effective predictive biomarkers to identify patients with better prognosis. And the precise link between RNA modification and efficacy of immunotherapy, particularly neoadjuvant immunotherapy, is still poorly understood.³⁷ Herein, we developed Writer-Score based on the cross-talk of four different types of RNA modification writers, which could accurately predict the clinical

outcomes of NSCLC patients receiving neoadjuvant immunotherapy. CD8⁺ T cells are the cornerstone of current efficient immunotherapies for cancer.⁴⁷ Additionally, we found non-exhausted CD8⁺ T cells were considerably concentrated in both stroma and tumor regions of low-score subgroups associated with favorable clinical outcomes, implying the reliability and high accuracy of Writer-Score. In addition, we found a negative correlation of Writer-Score with a variety of ICIs expression, indicating higher expression of ICIs was related to better prognosis for NSCLC patients receiving neoadjuvant immunotherapy. However, recent publications have only concentrated on the predictive value of limited number of ICI, such as PD-L1.^{14,15} And many other ICIs have not been examined, which might be potential predictive biomarkers for neoadjuvant immunotherapy and warrant further investigation.

In addition, we found RNA modification writers played a significant role in regulating post-transcriptional events in LUAD. The results demonstrated that miRNAs expressed differentially in high- and low-score subgroups, which targeted multiple genes correlated with numerous pathways. In the low-score subgroup, the PI3K-Akt, Ras and MAPK signaling pathways were predominantly activated, while cGMP-PKG and mTOR signaling pathways were highly activated in the high-score subgroup. The low-score subgroup was mostly associated with cell cycle and normal cell growth, whereas the high-score subgroup was notably correlated with the regulation of vascular smooth muscle cells, cell apoptosis and tumor metabolism.

Certainly, this research also possesses several drawbacks that warrant consideration. First, although we used RNA-seq data from one of the largest cohort of NSCLC patients receiving neoadjuvant immunotherapy, more patients are needed for future study. Second, we used the retrospective FFPE specimens in the NCC cohort and fresh specimens are supposed to be collected in future study. Third, inevitable bias may exist in our retrospective study and prospective study is needed for further validation.

To summarize, we comprehensively analyzed the exact role of four different kinds of RNA modification writers, revealing the intricate interactions among them and their correlations with patients' prognosis and TME features in NSCLC. The Writer-Score also performed well in predicting prognosis of NSCLC patients receiving neoadjuvant anti-PD-1 treatment.

Contributors

BZ, JZ, and SG designed and supervised the study. BZ, FB, RZ, MZ, PS analyzed the data. BZ wrote the original draft. LL, YP, GB and SG edited the draft. BZ, FB, RZ, MZ, PS, LL, YP, GB, JZ and SG have verified the underlying data. All authors have read and approved the final manuscript.

Data sharing statement

GEO data was downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/gds/>) under the accession number(s) GSE13213, GSE29013, GSE30219, GSE31210, GSE37745, GSE50081 and GSE72094. TCGA RNA-seq data were downloaded from the TCGA database via the UCSC Xena (<https://xena.ucsc.edu/>) under the accession number(s) LUAD-FPKM. The copy number alteration data of the TCGA cohort were downloaded from the cBioportal database (<https://www.cbioportal.org/>). The data used in this study are available from the corresponding author on reasonable request.

Declaration of interests

The authors report no conflict of interest.

Acknowledgements

The work was supported by the Medical and Health Science-Technology Innovation Project of Chinese Academy of Medical Sciences (2021-12M-1-015), National Key R&D Program of China (2021YFC2500900) and Central Health Research Key Projects (2022ZD17).

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2022.104268.

References

- Hirsch FR, Scagliotti GV, Mulshine JL, et al. Lung cancer: current therapies and new targeted treatments. *Lancet*. 2017;389(10066):299–311.
- Zhou B, Zang R, Zhang M, et al. Worldwide burden and epidemiological trends of tracheal, bronchus, and lung cancer: a population-based study. *EBioMedicine*. 2022;78:103951.
- Kleczo EK, Kwak JW, Schenk EL, Nemenoff RA. Targeting the complement pathway as a therapeutic strategy in lung cancer. *Front Immunol*. 2019;10:954.
- Barbieri I, Kouzarides T. Role of RNA modifications in cancer. *Nat Rev Cancer*. 2020;20(6):303–322.
- Liang Z, Kidwell RL, Deng H, Xie Q. Epigenetic N6-methyladenosine modification of RNA and DNA regulates cancer. *Cancer Biol Med*. 2020;17(1):9–19.
- Roundtree IA, Evans ME, Pan T, He C. Dynamic RNA modifications in gene expression regulation. *Cell*. 2017;169(7):1187–1200.
- Xiang JF, Yang Q, Liu CX, Wu M, Chen LL, Yang L. N(6)-Methyladenosines modulate A-to-I RNA editing. *Mol Cell*. 2018;69(1):126–35.e6.
- Zhang H, Shi X, Huang T, et al. Dynamic landscape and evolution of m6A methylation in human. *Nucleic Acids Res*. 2020;48(11):6251–6264.
- Zhang C, Jia G. Reversible RNA modification N(1)-methyladenosine (m1A) in mRNA and tRNA. *Genomics Proteomics Bioinform*. 2018;16(3):155–161.
- Nakano M, Nakajima M. Significance of A-to-I RNA editing of transcripts modulating pharmacokinetics and pharmacodynamics. *Pharmacol Ther*. 2018;181:13–21.
- Zhang Y, Ye Y, Lou Y, et al. Comprehensive characterization of alternative polyadenylation in human cancer. *J Natl Cancer Inst*. 2018;110(4):379–389.
- Chae YK, Arya A, Iams W, et al. Current landscape and future of dual anti-CTLA4 and PD-1/PD-L1 blockade immunotherapy in cancer: lessons learned from clinical trials with melanoma and non-small cell lung cancer (NSCLC). *J Immunother Cancer*. 2018;6(1):39.
- Zhou B, Gao S. Construction and validation of a novel immune and tumor mutation burden-based prognostic model in lung adenocarcinoma. *Cancer Immunol Immunother*. 2022;71(5):1183–1197.
- Gao S, Li N, Gao S, et al. Neoadjuvant PD-1 inhibitor (Sintilimab) in NSCLC. *J Thorac Oncol*. 2020;15(5):816–826.
- Forde PM, Chaft JE, Smith KN, et al. Neoadjuvant PD-1 blockade in resectable lung cancer. *N Engl J Med*. 2018;378(21):1976–1986.
- Ramos-Casals M, Brahmer JR, Callahan MK, et al. Immune-related adverse events of checkpoint inhibitors. *Nat Rev Dis Primers*. 2020;6(1):38.
- Wang L, Hui H, Agrawal K, et al. m(6) A RNA methyltransferases METTL3/14 regulate immune responses to anti-PD-1 therapy. *Embo J*. 2020;39(20):e104514.
- Xie Y, Xiao G, Coombs KR, et al. Robust gene expression signature from formalin-fixed paraffin-embedded samples predicts prognosis of non-small-cell lung cancer patients. *Clin Cancer Res*. 2011;17(17):5705–5714.
- Rousseaux S, Debernardi A, Jacquiau B, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci Transl Med*. 2013;5(186):186ra66.
- Okayama H, Kohno T, Ishii Y, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res*. 2012;72(1):100–111.
- Botling J, Edlund K, Lohr M, et al. Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. *Clin Cancer Res*. 2013;19(1):194–204.
- Der SD, Sykes J, Pintilie M, et al. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. *J Thorac Oncol*. 2014;9(1):59–64.
- Schabath MB, Welsh EA, Fulp WJ, et al. Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. *Oncogene*. 2016;35(24):3209–3216.
- Tomida S, Takeuchi T, Shimada Y, et al. Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. *J Clin Oncol*. 2009;27(17):2793–2799.
- Yoshihara K, Shahmoradgoli M, Martinez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612.
- Charoentong P, Finotello F, Angelova M, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep*. 2017;18(1):248–262.
- Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
- Liu Y, Zhang C, Fan J, et al. Comprehensive analysis of clinical significance of stem-cell related factors in renal cell cancer. *World J Surg Oncol*. 2011;9:121.
- Chang LC, Das B, Lih CJ, et al. RefCNV: identification of gene-based copy number variants using whole exome sequencing. *Cancer Inform*. 2016;15:65–71.
- Chen Y, Mao B, Peng X, et al. A comparative study of genetic profiles of key oncogenesis-related genes between primary lesions and matched lymph nodes metastasis in lung cancer. *J Cancer*. 2019;10(7):1642–1650.
- Cao G, Li HB, Yin Z, Flavell RA. Recent advances in dynamic m6A RNA modification. *Open Biol*. 2016;6(4):160003.
- Di Giammartino DC, Nishida K, Manley JL. Mechanisms and consequences of alternative polyadenylation. *Mol Cell*. 2011;43(6):853–866.
- Huang H, Tan BZ, Shen Y, et al. RNA editing of the IQ domain in Ca(v)1.3 channels modulates their Ca²⁺-dependent inactivation. *Neuron*. 2012;73(2):304–316.
- Li X, Xiong X, Wang K, et al. Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. *Nat Chem Biol*. 2016;12(5):311–316.
- Samstein RM, Lee CH, Shoushtari AN, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet*. 2019;51(2):202–206.
- Shulman Z, Stern-Ginossar N. The RNA modification N(6)-methyladenosine as a novel regulator of the immune system. *Nat Immunol*. 2020;21(5):501–512.

- 37 Li X, Ma S, Deng Y, Yi P, Yu J. Targeting the RNA m(6)A modification for cancer immunotherapy. *Mol Cancer*. 2022;21(1):76.
- 38 Fukumura D, Kloepper J, Amoozgar Z, Duda DG, Jain RK. Enhancing cancer immunotherapy using antiangiogenics: opportunities and challenges. *Nat Rev Clin Oncol*. 2018;15(5):325–340.
- 39 Sánchez-Alonso S, Setti-Jerez G, Arroyo M, et al. A new role for circulating T follicular helper cells in humoral response to anti-PD-1 therapy. *J Immunother Cancer*. 2020;8(2):1–13.
- 40 Faget J, Groeneveld S, Boivin G, et al. Neutrophils and snail orchestrate the establishment of a pro-tumor microenvironment in lung cancer. *Cell Rep*. 2017;21(11):3190–3204.
- 41 Marshall EA, Ng KW, Kung SH, et al. Emerging roles of t helper 17 and regulatory T cells in lung cancer progression and metastasis. *Mol Cancer*. 2016;15(1):67.
- 42 Jiang Y, Zhan H. Communication between EMT and PD-L1 signaling: new insights into tumor immune evasion. *Cancer Lett*. 2020;468:72–81.
- 43 Chae YK, Davis AA, Agte S, et al. Clinical implications of circulating tumor DNA tumor mutational burden (ctDNA TMB) in non-small cell lung cancer. *Oncologist*. 2019;24(6):820–828.
- 44 Gandara DR, Paul SM, Kowanetz M, et al. Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. *Nat Med*. 2018;24(9):1441–1448.
- 45 Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med*. 2015;373(17):1627–1639.
- 46 Rusch VW, Chaft JE, Johnson B, et al. Neoadjuvant atezolizumab in resectable non-small cell lung cancer (NSCLC): initial results from a multicenter study (LCMC3). *J Clin Oncol*. 2018;36(15_suppl):8541.
- 47 Raskov H, Orhan A, Christensen JP, Gögenur I. Cytotoxic CD8(+) T cells in cancer and cancer immunotherapy. *Br J Cancer*. 2021;124(2):359–367.