

RESEARCH

Open Access



# Comprehensive analysis reveals the prognostic and immunogenic characteristics of DNA methylation regulators in lung adenocarcinoma

Jing Huang<sup>1†</sup>, Chujian Huang<sup>2†</sup>, Can Huang<sup>3†</sup>, Zichang Xiang<sup>4</sup>, Yao Ni<sup>4</sup>, Jian Zeng<sup>5\*</sup> and Songhua Cai<sup>2\*</sup>

## Abstract

DNA methylation regulators (DMRs) play a key role in DNA methylation, thus mediating tumor occurrence, metastasis, and immunomodulation. However, the effects of DMRs on clinical outcomes and immunotherapy response remain unexplored in lung adenocarcinoma (LUAD). In this study, eight LUAD cohorts and one immunotherapeutic cohort of lung cancer were utilized. We constructed a DNA methylation regulators-related signature (DMRRS) using univariate and multivariate COX regression analysis. The DMRRS-defined low-risk group was preferentially associated with favorable prognosis, tumor-inhibiting microenvironment, more sensitivity to several targeted therapy drugs, and better immune response. Afterward, the prognostic value and predictive potential in immunotherapy response were validated. Collectively, our findings uncovered that the DMRRS was closely associated with the tumor immune microenvironment and could effectively predict the clinical outcome and immune response of LUAD patients.

**Keywords** Lung adenocarcinoma, DNA methylation regulator, Tumor microenvironment, Prognosis, Immune response

<sup>†</sup>Jing Huang, Chujian Huang and Can Huang contributed equally to this work.

\*Correspondence:

Jian Zeng

291552521@qq.com

Songhua Cai

caisonghua@hcamssz.ac.cn

<sup>1</sup>Department of Thoracic Surgery, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, 107 Yanjiang West Road, Guangzhou 510120, China

<sup>2</sup>Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen 518116, China

<sup>3</sup>Eight-year MD program, Peking Union Medical College, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100087, China

<sup>4</sup>Shenzhen University Medical School, Shenzhen 518055, Guangdong, China

<sup>5</sup>Department of Anesthesiology, Longgang District Central Hospital of Shenzhen, Shenzhen 518116, Guangdong, China



© The Author(s) 2024. **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 Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Introduction

Lung cancer remains one of the leading causes of cancer-related deaths worldwide [31], among which more than 40% of patients have lung adenocarcinoma (LUAD) [30]. Numerous therapies, such as surgery, radiotherapy, chemotherapy, molecular targeted therapy, and immunotherapy, have been developed for the clinical treatment of LUAD in recent decades. However, the 5-year survival rate of advanced LUAD is still less than 25% due to the complexity of the tumor formation mechanism [8, 32]. Therefore, it is urgent to find new prognostic biomarkers or therapeutic targets.

Immune checkpoint blocker therapy has achieved impressive success in treating various tumor types [1, 7, 16]. However, only a few patients can benefit from immunotherapy [35]. To identify patients who respond well to immunotherapy, the researchers have made great efforts in exploring potential biomarkers. The tumor mutation burden (TMB), microsatellite instability (MSI) status, PD-L1 expression, and some mutated genes have shown a great future [14, 25, 27]. Unfortunately, the predictive power of these markers varies due to tumor heterogeneity. Better biomarkers may be uncovered by focusing on specific cancer types.

As one of the most important epigenetic modifications, DNA methylation plays a crucial role in various biological processes [29]. The potential prognosis of DNA methylation regulators has been noted in several types of cancer [41, 43]. Nevertheless, the value of DNA methylation regulators in LUAD remains largely unknown. In addition, there is growing evidence of a link between DNA methylation and tumor immunity [24, 33]. For example, TET1 mutation has been strongly associated with increased tumor immunogenicity and could be used as a potential biomarker for immune checkpoint blocker therapy in multiple cancers [38]. A valuable methylation score has been established based on DNA methylation regulators, which has helped develop personalized immunotherapy strategies for gastric cancer [23].

This study aimed to conduct a comprehensive analysis of DNA methylation regulators in LUAD based on the TCGA and GEO databases. Toward this goal, a DNA methylation regulators-related signature (DMRRS) was constructed to predict the prognosis of patients with LUAD effectively. Then the relationships between immune scores, immune cell infiltration, chemotherapy and targeted therapy sensitivity, immune response, and risk signature were thoroughly analyzed.

## Materials and methods

### Dataset source and preprocessing

Seven microarray datasets (GSE19188, GSE30219, GSE31210, GSE3141, GSE37745, GSE50081, and GSE72094) were obtained from Gene Expression

Omnibus (GEO) database. The expression matrices of the above datasets and corresponding clinical data were downloaded using the GEOquery R package [5]. Since GSE19188, GSE30219, GSE31210, GSE37745, and GSE50081 were all conducted in the same platform: Affymetrix Human Genome U133 Plus 2.0 Array, we then processed the raw CEL data of these five datasets by the robust multichip average (RMA) algorithm for background correction and normalization [10]. Finally, using the ComBat function of the sva R package for batch removal, we integrated them into a large GEO cohort and referred to it as the Large-GEO cohort hereinafter [19].

The RNA sequencing data (FPKM value) of gene expression in TCGA databases were downloaded from the Genomic Data Commons (GDC, <https://portal.gdc.cancer.gov/>) using the R package TCGAbiolinks [3]. The corresponding clinical information of LUAD was downloaded from Xena public data hubs. The data of 533 LUAD samples and 59 adjacent normal tissues were downloaded on April 10, 2021. The following inclusion criteria were used: (1) histologically confirmed LUAD; (2) simultaneously available information on mRNA expression profile data and OS and (3) Survival time  $\geq 30$  days. Lastly, 490 patients with LUAD and the corresponding clinicopathological information, including age, gender, TNM staging, and grade, were enrolled for further analysis.

Publicly available immunotherapeutic datasets of lung cancer with both gene expression and corresponding clinical data were thoroughly searched. We just found the GSE126044 cohort, containing 16 lung cancer patients treated with anti-PD-1 antibody.

The detailed clinical information of the above datasets is shown in **Supplementary Table 1**.

### Multi-omic landscape of DNA methylation regulators in TCGA-LUAD

About 20 DNA methylation regulators were collected from previously published studies, including three writers (DNMT1, DNMT3A, DNMT3B), three erasers (TET1, TET2, TET3), and 14 readers (MBD1, MBD2, MBD3, MBD4, ZBTB33, ZBTB38, ZBTB4, UHRF1, UHRF2, MECP2, UNG, TDG, NTHL1, SMUG1). The somatic mutation and Copy Number Variation (CNV) data were acquired from the TCGA database and further analyzed. In short, we downloaded the mutation and CNV data using the GDCquery package. The mutation data was visualized by the maftools package. As for CNV data, we thresholded them by a noise cutoff of 0.2. Genes with CNV values smaller than  $-0.2$  were categorized as “loss”, while larger than 0.2 were as “gain”. Next, the differential expression of those regulators was compared between the tumor tissues and adjacent normal pairs.

### Construction of the DNA methylation regulators-related signature (DMRRS)

We first performed univariate Cox regression analysis in both TCGA and Large-GEO cohorts to evaluate the DNA methylation regulators' prognostic value. Then, we utilized multivariate Cox regression analysis to construct the powerful prognostic signature in the Large GEO training cohort with the backward stepwise regression method. A risk score for each patient was calculated as the sum of each gene's score, which was generated using the following formula:

$$Risk\ score = \sum_{i=1}^n Coef(i) * Expression(i)$$

Where  $n$  represents the number of genes,  $Coef(i)$  is the coefficient of relative prognostic genes in the model, and  $Expression(i)$  is the expression of each selected gene. The sensitivity and specificity of the prognostic signature were further accessed by receiver operating characteristic (ROC) curves and the area under the ROC curves (AUC values). According to this equation, the risk score of each patient was calculated in the Large-GEO training, TCGA, GSE3141, and GSE72094 cohorts. The patients were then divided into high- and low-risk groups using the surv-cutpoint function in the survminer package.

### Nomogram development and evaluation

We utilized the univariate and multivariate Cox regression analysis to choose the independent prognostic factors in TCGA-LUAD. Based on the TNM grade and risk score, we developed the nomogram using the rms R package. To evaluate the prediction accuracy of the nomogram, we calculated Harrell's consistency index and plotted the calibration curves.

### Gene set variation analysis (GSVA) and functional annotation

To investigate the biological process of two risk groups, Gene set variation analysis (GSVA) was performed using the GSVA R package [13]. Differential analysis was then conducted to determine the significantly enriched pathways in each group. The gene sets of 'c2.cp.kegg.v7.4.symbols' were downloaded from the MSigDB database for running GSVA analysis. Adjusted P with a value less than 0.05 was considered statistically significant.

### Estimation of immune score and immune cell infiltrates

We used the ESTIMATE algorithm to calculate each patient's stromal and immune scores [40]. The fraction of 22 immune cell types for each sample was yielded by estimating relative subsets of RNA transcripts (CIBERSORT; <https://cibersort.stanford.edu/>). The algorithm of 1,000 permutations was adopted. Only samples with

a CIBERSORT  $p$  of  $<0.05$  were included to perform the subsequent analysis. Correlation between different risk groups and abundances of each cell type were analyzed and visualized by radar chart.

### Tumor immune estimation resource (TIMER) database

The TIMER database was used to evaluate the correlation between DMRRS-related genes and immune cell infiltrates in the tumor microenvironment. In addition, using the Mutation module of the TIMER database, we compared the contents of immune cells in wild and mutant types of DMRRS-related genes.

### Tumor immune single cell hub database

The Tumor Immune Single-Cell Hub (TISCH) (<http://tisch.comp-genomics.org>) provided a detailed characterization of the immune system heterogeneity in tumors at the single-cell level [34]. Our study used the TISCH database to evaluate the expression of DMRRS-related genes in different immune cells.

### Estimation of drug sensitivity

The pRRophetic R package was used to estimate drug sensitivity, determined by the IC50 of compounds [11].

### Assessment of immunophenoscore (IPS)

We downloaded the immunophenoscore (IPS) from The Cancer Immunome Atlas (TCIA) to validate the predicting roles of DMRRS in the immune checkpoint inhibitor (ICI) treatment. A higher IPS score indicates a better immunotherapy response [2].

### Clinical specimens and immunohistochemistry (IHC)

We retrospectively collected 18 paraffin-embedded LUAD specimens and 18 adjacent normal tissues from the National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital (Shenzhen, China) to compare the differential expression of UHRF1 and MECP2. For survival analysis, we further enrolled 100 tumors for IHC and collected clinical information including age, gender, TNM stage, smoking or not, and disease-free survival (DFS). The cutoff value was generated using the surv-cutpoint R function. Informed consent was obtained from all patients. This study was approved by the Ethics and Research Committees of the National Cancer Center/Cancer Hospital & Shenzhen Hospital, the Chinese Academy of Medical Sciences, and Peking Union Medical College.

After deparaffinization, rehydration and antigen retrieval, the samples were incubated with primary antibodies against MECP2 (Proteintech, 21402-1-AP, 1: 100) and UHRF1 (Proteintech, 10861-1-AP, 1: 100) at 4 °C overnight. The slides were then incubated with anti-rabbit secondary antibody and followed by chromogen

DAB staining and haematoxylin counterstaining. The expression of selected markers was scored by percent of positive cells (1–100%) and staining intensity (1=weak, 2=moderate, 3=strong). A final histoscore (H-score) was derived as:  $H\text{-score} = (\text{percentage of weak intensity} \times 1) + (\text{percentage of moderate intensity} \times 2) + (\text{percentage of strong intensity} \times 3)$ , yielding a range of possible H-scores of 0 to 300 [6]. Two pathologists who were blind to the information of patients independently validated the IHC results.

### Statistical analysis

All statistical analyses were performed using SPSS 26.0 (Chicago, IL, USA) and R software version 4.0.5 (<http://www.r-project.org>). Continuous parameters were expressed as mean value  $\pm$  standard deviation (SD) and were compared by the Mann-Whitney U test and Kruskal–Wallis test between different groups. For categorical parameters, proportions were compared by Pearson's Chi-square test. Pearson's correlations were employed to assess associations between two variables. Survival curves were plotted by the Kaplan–Meier method and were compared by log-rank test. A *P* value less than 0.05 was considered statistically significant.

## Results

### Multi-omic characteristics of DNA methylation regulators in TCGA-LUAD

The study flowchart is depicted in Fig. 1. In the TCGA-LUAD cohort, several important oncogenes like TP53 and KRAS were among the top mutated genes, with missense mutations being the most frequent mutation event (Fig. 2A and B). Despite the lower mutation rates, the mutant of DNMT3A, TET3, UHRF1, and NTHL1 showed worse overall survival (OS) (Fig. 2C and D). Several co-occurrence relationships among the DNA methylation regulators were revealed, such as DNMT3A and NTHL1, DNMT1 and ZBTB4, and so on (Fig. 2E). As for CNV, readers like UHRF1-2, MBD1-3, and ZBTB4 displayed the highest frequency of copy number losses, while writers and erasers tend to have prevalent copy number gains (Fig. 2F).

To investigate the expression features, we systematically analyzed 57 LUAD tissues and adjacent normal tissues from TCGA. The results showed that most DNA regulators were differently expressed between LUAD and normal tissues (Fig. 2G). The expression levels of three writers (DNMT1, DNMT3A, DNMT3B), two erasers (TET1, TET3), and eight readers (MBD4, ZBTB33, UHRF1, UHRF2, UNG, TDG, NTHL1, SMUG1) were dramatically higher in tumor tissues ( $p < 0.05$ ). While MBD3, ZBTB4, and MECP2 were markedly lower in LUAD tissues than in normal tissues ( $p < 0.05$ ). No

statistically significant difference was evident regarding the expression level of TET2, MBD1, MBD2, and ZBTB38.

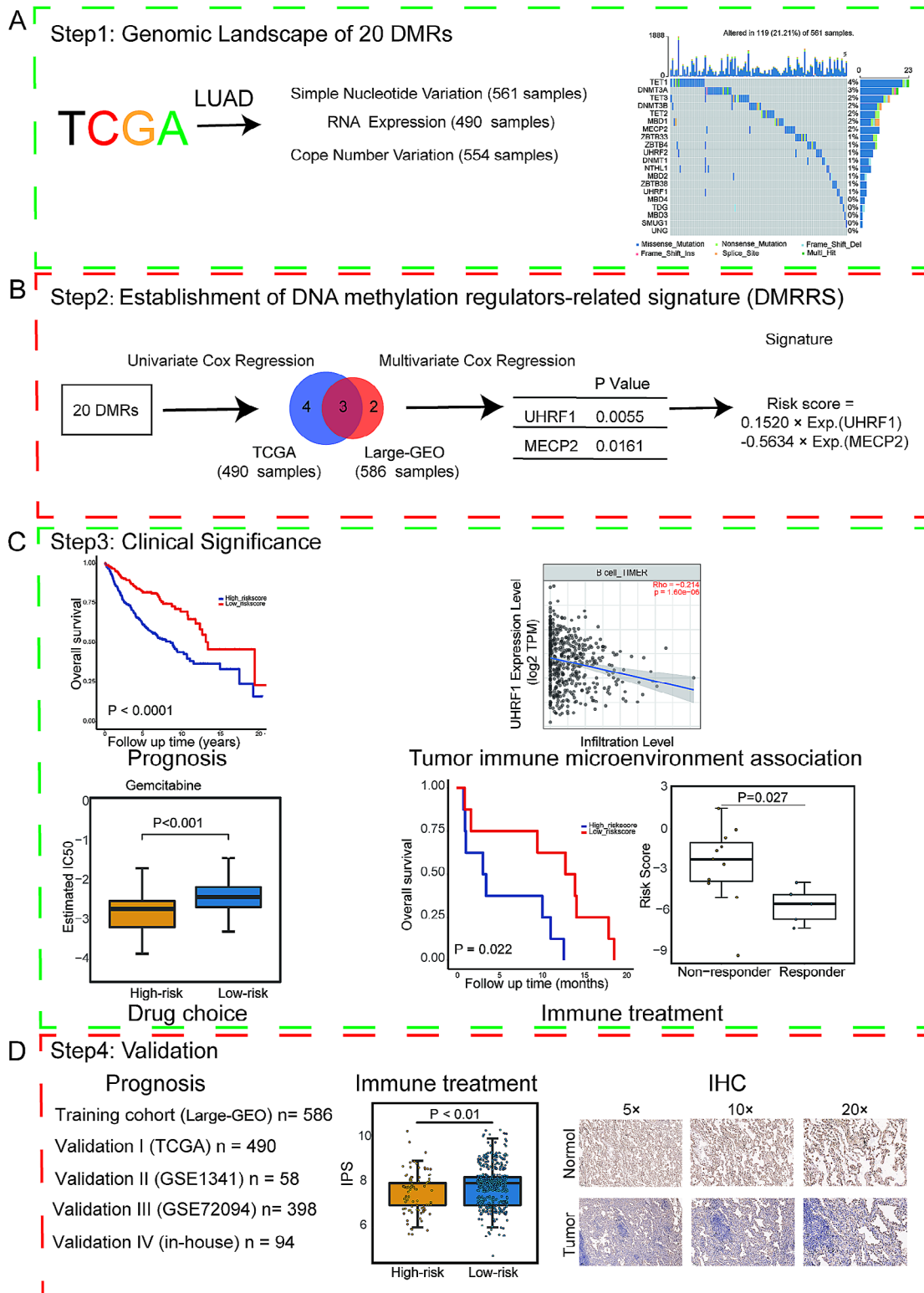
### Construction of DNA methylation regulators-related signature (DMRRS)

Five regulators in Large-GEO data (MECP2, ZBTB4, NTHL1, UHRF1, and MBD4) and seven regulators (TET2, MECP2, ZBTB4, UNG, SMUG1, TDG, and UHRF1) in TCGA were significantly related to the OS after univariate Cox regression analysis (Supplementary Tables 2 and 3). Three genes were linked to patients' survival in both cohorts. So, we next used the three intersected genes to build the prognostic signature by the multivariate Cox regression analysis in the Large-GEO cohort with the backward stepwise method (Fig. 3A). Only two regulators, UHRF1 and MECP2, were included in the DMRRS, and the equation was as follows:  $(\text{Risk score} = 0.1520 \times \text{UHRF1 expression level}) - (0.5634 \times \text{MECP2 expression level})$ . Afterward, all LUAD patients were divided into low-risk and high-risk groups based on the optimal cutoff value of the risk score. Kaplan–Meier survival analysis revealed that the OS of the high-risk group was lower than that of the low-risk group in the Large-GEO cohort (Fig. 3B). Besides, in the Large-GEO dataset, the disease-free survival (DFS) of the low-risk group was also significantly higher than the other group (Fig. 3C). The Receiver operating characteristic (ROC) curve was used to assess the prognostic accuracy of identified risk signatures, with the 1-, 3- and 5-year AUC values for the two risk signatures being 0.620, 0.622, and 0.625, respectively (Supplementary Fig. 1A).

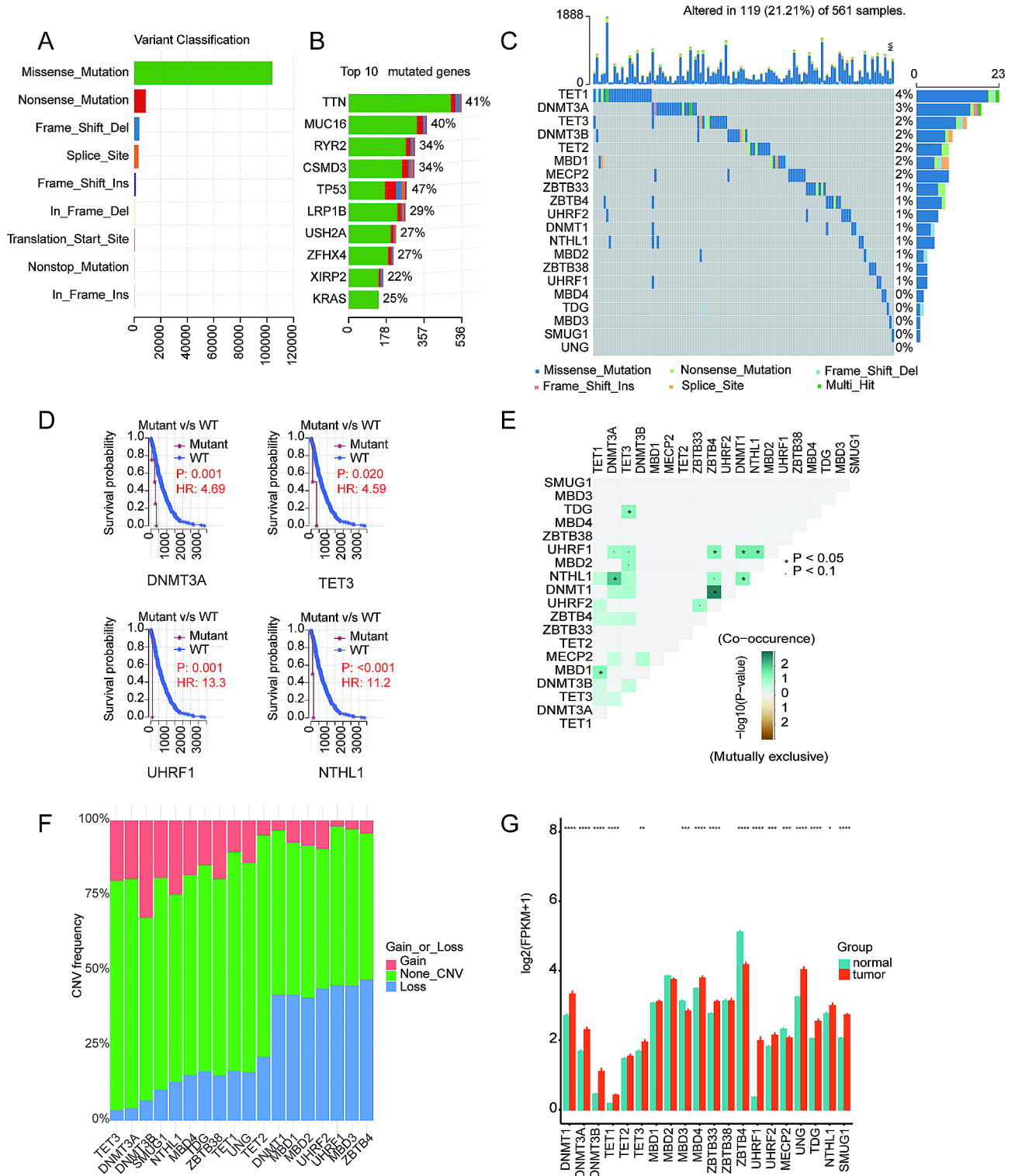
### Clinical characteristics associated with DMRRS

The clinical characteristics of patients in different risk groups were further evaluated in the TCGA cohort. The low-risk patients showed better clinical outcomes as expected in the cohort (Fig. 3D and E).

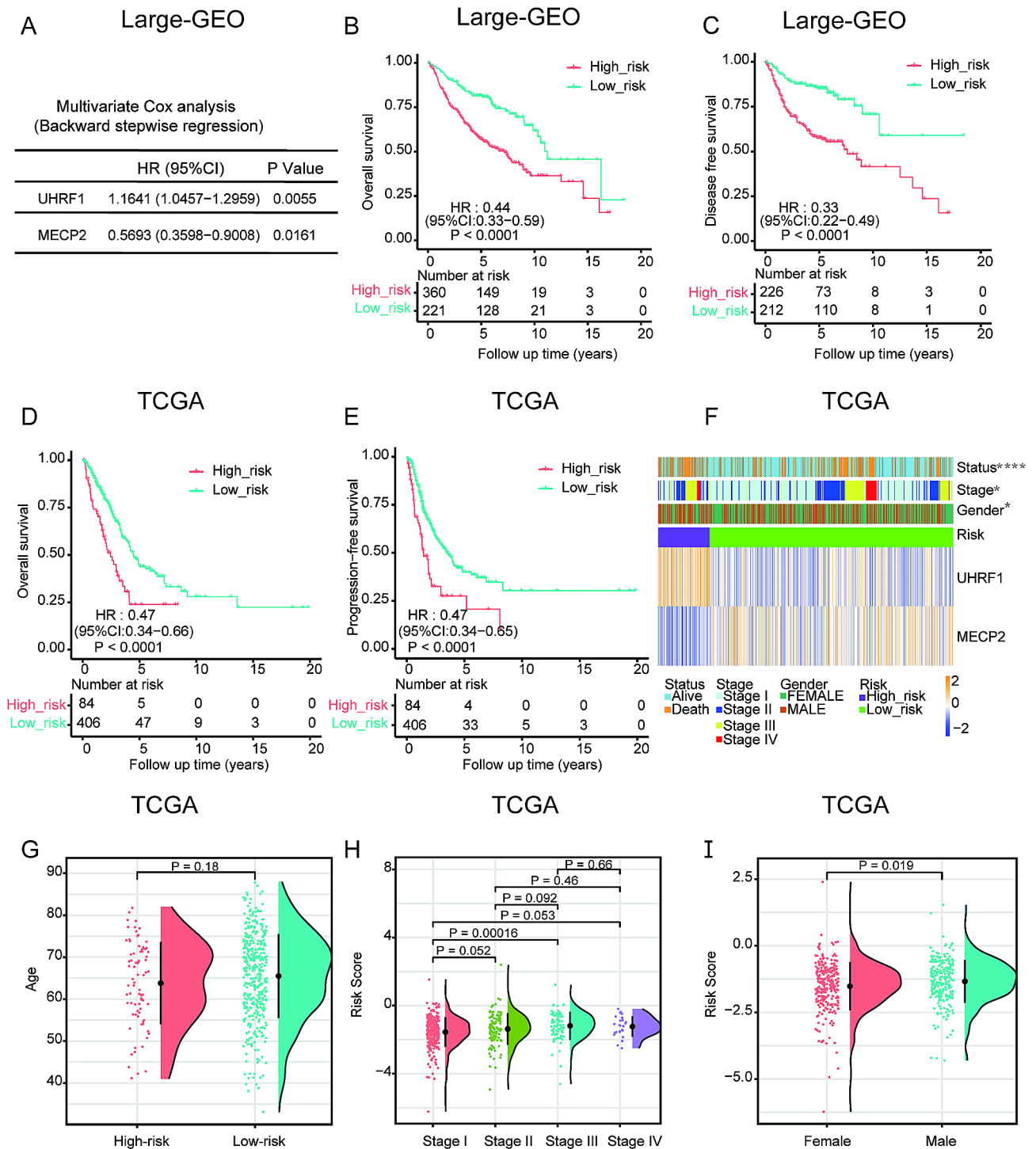
The heatmap demonstrated the expression levels of two DNA methylation regulators between high- and low-risk groups in TCGA-LUAD (Fig. 3F). The expression levels of UHRF1 were typically higher in the high-risk group than those of the low-risk group, while MECP2 expression was lower in the high-risk group. The difference in gender and TNM stages between the two groups was significant. As shown in Fig. 3G–I, male patients had an increased risk score level than female patients. Patients in stage I tended to have a lower score when compared with stage II–IV. We then performed OS analysis between high- and low-risk subgroups regarding different clinical features, including age, gender, and stages. The low-risk group had obviously better OS than the high-risk group whatever the age or the gender (Supplementary Fig. 1B and 1C). As for the stage, Patients with stage I tend to



**Fig. 1** Study Design and Workflow of The Study. **(A)** Step 1: Genomic landscape of 20 DMRs, including simple nucleotide variation, copy number variation and RNA expression analysis. **(B)** Step 2: Establishment of DMRRS. **(C)** Step 3: Clinical significance of DMRRS. The analysis of prognostic value, tumor microenvironment association, chemotherapy and targeted therapy drug sensitivity, and immune response in DMRRS-defined subgroups. **(D)** Step 4: Validation of prognostic value and predicting potential of immune response of DMRRS. IHC validation of the abnormal expression of DMRRS' two genes between normal and tumor tissues. DMRs: DNA methylation regulators. DMRRS: DNA methylation regulators-related signature



**Fig. 2** Multi-Omic Landscape of DNA Methylation Regulators in TCGA-LUAD Cohort. **(A)** Classification of mutations of different DNA methylation regulators genes. **(B)** the top 10 mutated genes. **(C)** Mutations of the 20 DNA regulators genes. **(D)** The survival curve of mutated regulators genes. **(E)** The mutation co-occurrence features of DNA methylation regulators. **(F)** The CNV features of DNA methylation regulators. **(G)** Expression levels of 20 DNA methylation regulators in 57 tumor and adjacent normal pairs. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  and \*\*\*\* $p < 0.0001$ . HR: hazard ratio



**Fig. 3** Construction of DMRRs in Large-GEO and Clinical Features of DMRRS-Defined Subgroups in TCGA Cohort. **(A)** Using multivariate Cox regression, two DNA methylation regulators were selected for risk coefficient calculation. **(B)** Kaplan-Meier curves of OS for patients with LUAD based on the risk score in the Large-GEO cohort. **(C)** Kaplan-Meier curves of DFS for patients with LUAD based on the risk score in the Large-GEO cohort. **(D)** Kaplan-Meier curves of OS for patients with LUAD based on the risk score in the TCGA cohort. **(E)** Kaplan-Meier curves of PFS for patients with LUAD based on the risk score in the TCGA cohort. **(F)** Heatmap and clinical features of high- and low-risk groups in TCGA cohort. **(G-I)** Relationship between risk scores and **(G)** Age, **(H)** Stage, and **(I)** Gender in TCGA cohort. OS: overall survival; DFS: disease-free survival; PFS: progression-free survival. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  and \*\*\*\* $p < 0.0001$

have more influence on the OS than the lower scores. However, the trend of better OS in the low-risk group still holds (Supplementary Fig. 1D).

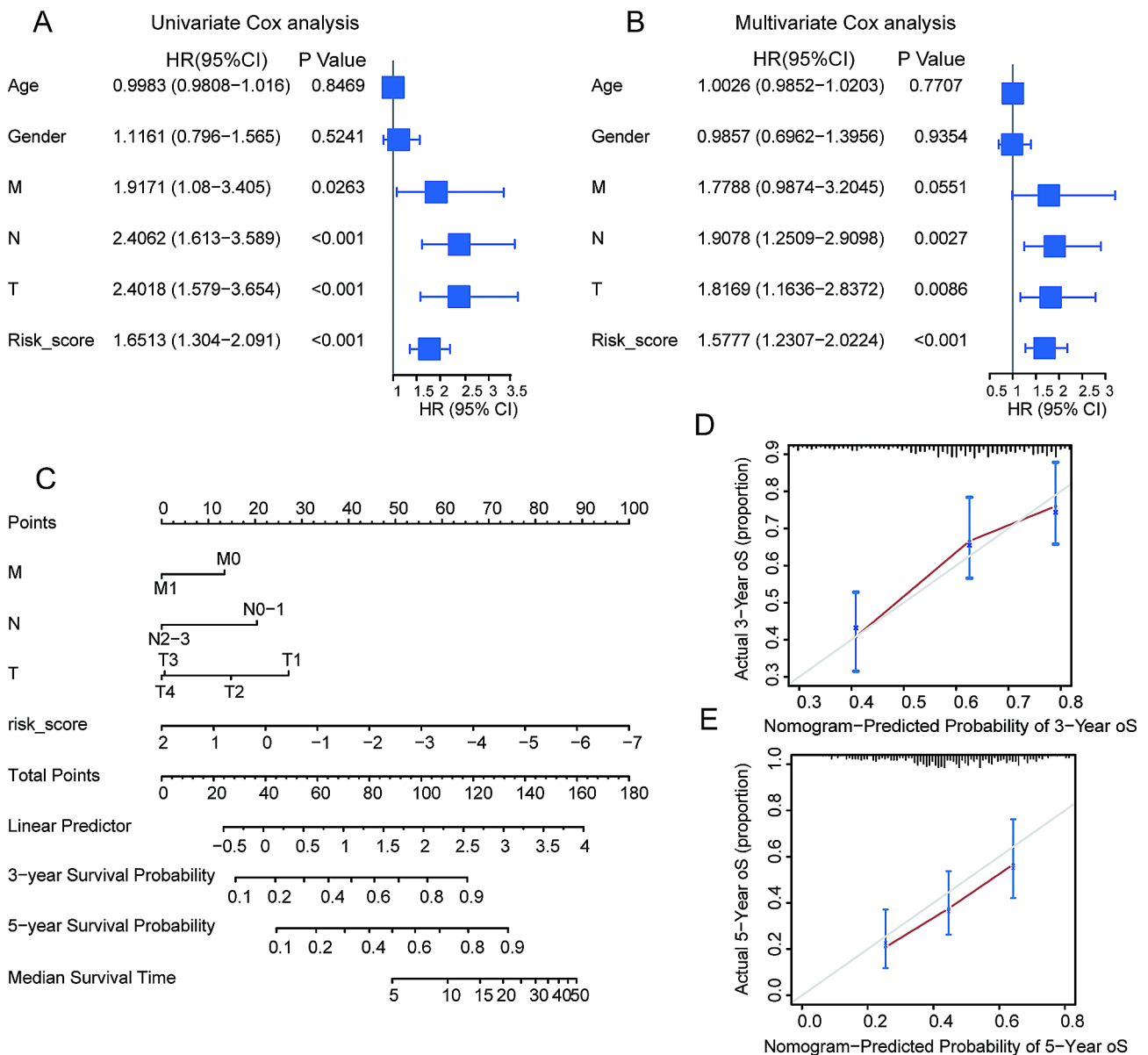
**Construction and evaluation of the prognostic nomogram**

To illustrate the prognostic benefits of the DMRRS-defined low-risk patients weren't just because of lower tumor stages, we performed the univariate and multivariate Cox regression analyses in the TCGA dataset. The results demonstrated that DMRRS was an independent prognostic factor (Fig. 4A and B). Thus, we fabricated a nomogram based on clinical risk characteristics,

including TNM stages and DMRRS-defined risk scores, to predict 3-and 5-year OS probability. The risk model showed predominant predictive ability (Fig. 4C). The C index of the nomogram model was 0.69, and correlation charts displayed ideal consistency (Fig. 4D and E). Collectively, those results above indicated a promising prognostic significance of DMRRS.

**Correlation between the tumor microenvironment and DMRRS**

To reveal the different biological behaviors of risk clusters in LUAD, we then compared GSVA scores in high- and



**Fig. 4** Construction and Validation of Nomogram. (A) Forest plot of univariate Cox analysis results. (B) Forest plot of multivariate Cox analysis results. (C) Nomogram for predicting 3- and 5-year overall survival of OC patients. (D-E) Calibration curves of the nomogram prediction of 3 years (D) and 5 years (E) OS of patients



low-risk groups using the limma package [28]. Intriguingly, many cancer-related pathways like cell cycle, DNA replication, and p53 signaling pathways were excessively activated in the high-risk group. In contrast, several immune-related pathways were activated in the low-risk group, such as the B/T cell receptor signaling pathway, natural killer cell-mediated cytotoxicity, and complement and coagulation cascades (Fig. 5A).

Next, we analyzed the relationship between immune scores, infiltration levels of immune cell types, and DMRRS-defined subgroups. The results showed that the high-risk patients had remarkably decreased immune scores (Fig. 5B). In addition, patients in the high-risk group were closely related to the macrophage M1 cell and activated Mast cells (Supplementary Fig. 1E), while low-risk patients tended to associate with adaptive immune cells such as activated CD4 T cells and CD8 T cells (Fig. 5C). It was rational to speculate that the DMRRS-defined low-risk group might be highly associated with the active tumor immune microenvironment.

We then analyzed the correlation of the two components of DMRRS: UHRF1 and MECP2, with immune cell infiltration in the LUAD tumor microenvironment. Interestingly, the infiltration level of certain immune cells was closely related to the expression of both two genes. As shown in Fig. 5D, though in low correlations, the B cell infiltrate was negatively associated with UHRF1 expression, while CD8+T cells, macrophages, and neutrophil cells were positively correlated with UHRF1 expression. All six immune cells except myeloid cells were positively correlated with MECP2 expression. Besides, the patients with mutated MECP2 showed higher CD8+T cells and neutrophil cells contents (Fig. 5E). Subsequently, GSEA was performed to identify the abnormally activated signaling pathways due to UHRF1 and MECP2 abnormal expression in LUAD. The results exhibited that some immune-related pathways like B and T cell receptor signaling pathways were closely related to both genes (Fig. 5F).

#### Predictive value of DMRRS in anti-PD-1 immunotherapy

Based on the GSE126044 immunotherapy cohort, consisting of 16 NSCLC patients with the intervention of anti-PD-1 antibody Nivolumab, we evaluated the ability of the DMRRS in predicting anti-PD-1 treatment response. The risk score of responders was significantly lower than non-responders, and the proportion of patients who responded to anti-PD-1 treatment in the low-risk group was 62.5% versus 0% in the high-risk group (Fig. 6A and B). Moreover, patients with lower risk scores exhibited a markedly prolonged survival, including OS and PFS (Fig. 6C and D).

#### Chemotherapy and targeted therapy sensitivity between DMRRS-defined subgroups

For LUAD treatment, chemotherapy and targeted therapy play indispensable roles and are widely used nowadays. It's thus crucial to identify subgroup patients who could be more sensitive to some specific drugs. Therefore, we estimated the therapeutic response to several commonly used drugs in high- and low-risk group patients. The low-risk group was more sensitive to Lapatinib, while the high-risk group was more sensitive to Cisplatin, Paclitaxel, Vinblastine, and Gemcitabine in both TCGA and Large-GEO cohorts (Fig. 6E and F). As lapatinib is not routinely used in lung cancer patients, we further compared the EGFR mutation in different risk groups. Consistent with the therapeutic response, the EGFR-Mut patients had significantly lower risk scores than the EGFR-Wild patients (Fig. 6F). Altogether, the DMRRS could help us determine the personalized treatment for LUAD patients.

#### Validation of predictive potential of DMRRS in prognosis and immunotherapy response

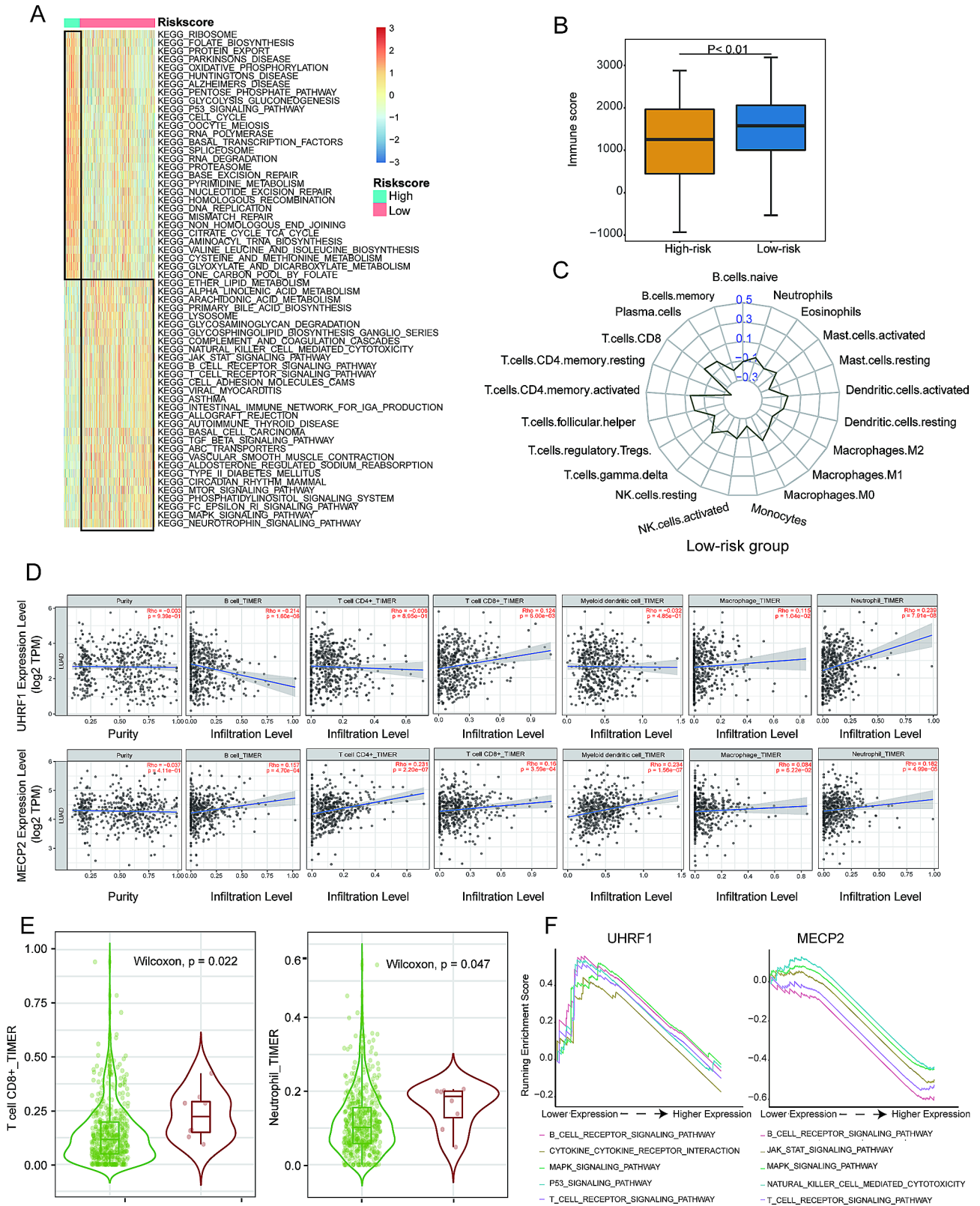
As independent validation sets, GSE3141 and GSE72094 cohorts were utilized to test the prognostic value of DMRRS further. Consistent with large-GEO and TCGA results, the low-risk groups showed longer OS than patients in high-risk (Fig. 7A and B). ROC curve exhibited that the 1- and 3- year AUC values of GSE3141 were 0.656 and 0.759, with 0.625 and 0.574 for 1- and 3- year respectively in GSE72094 (Figs. 7C and D).

Since no other immunotherapy cohort of lung cancer patients was available, we tried to validate the predictive potential of DMRRS of ICIs by exploring the correlation between DMRRS and IPS score, one recognized immunotherapy predictor. As depicted in Fig. 7E and F, low-risk patients had significantly increased IPS score and higher response rate in the anti-CTLA4 treatment.

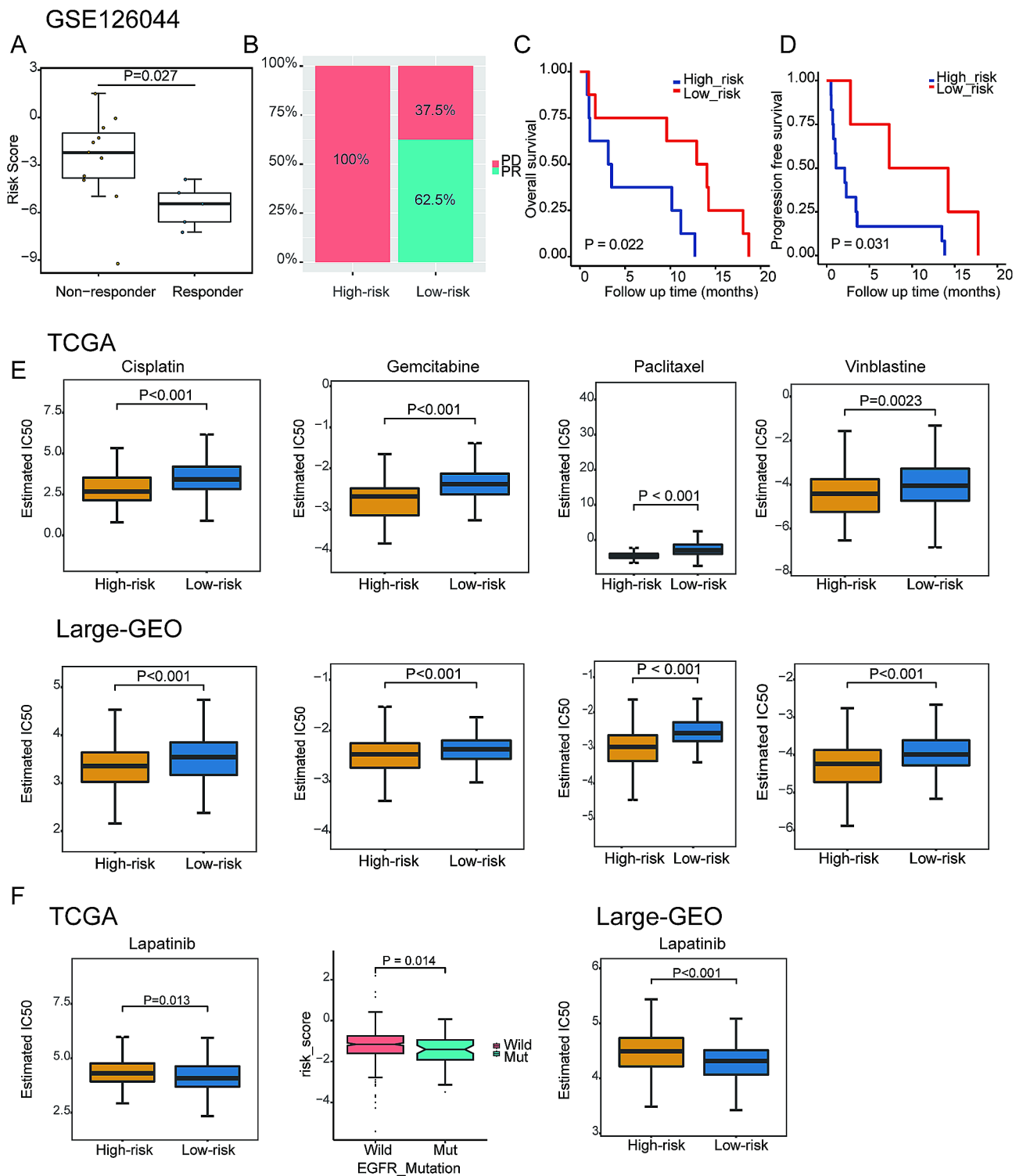
#### Validation of DMRRS-related gene expression in LUAD tissues

Both UHRF1 and MECP2 genes were differentially expressed in normal and tumor lung samples, as indicated in Fig. 2G. We quantified those two genes by IHC in 18 LUAD tumors and adjacent normal tissues. The IHC staining revealed significantly lower expression of MECP2 in tumors ( $P=0.022$ ). In contrast, the UHRF1 expression was relatively elevated in tumors with a  $P$  value of 0.064 due to the limited test samples (Fig. 8A-C).

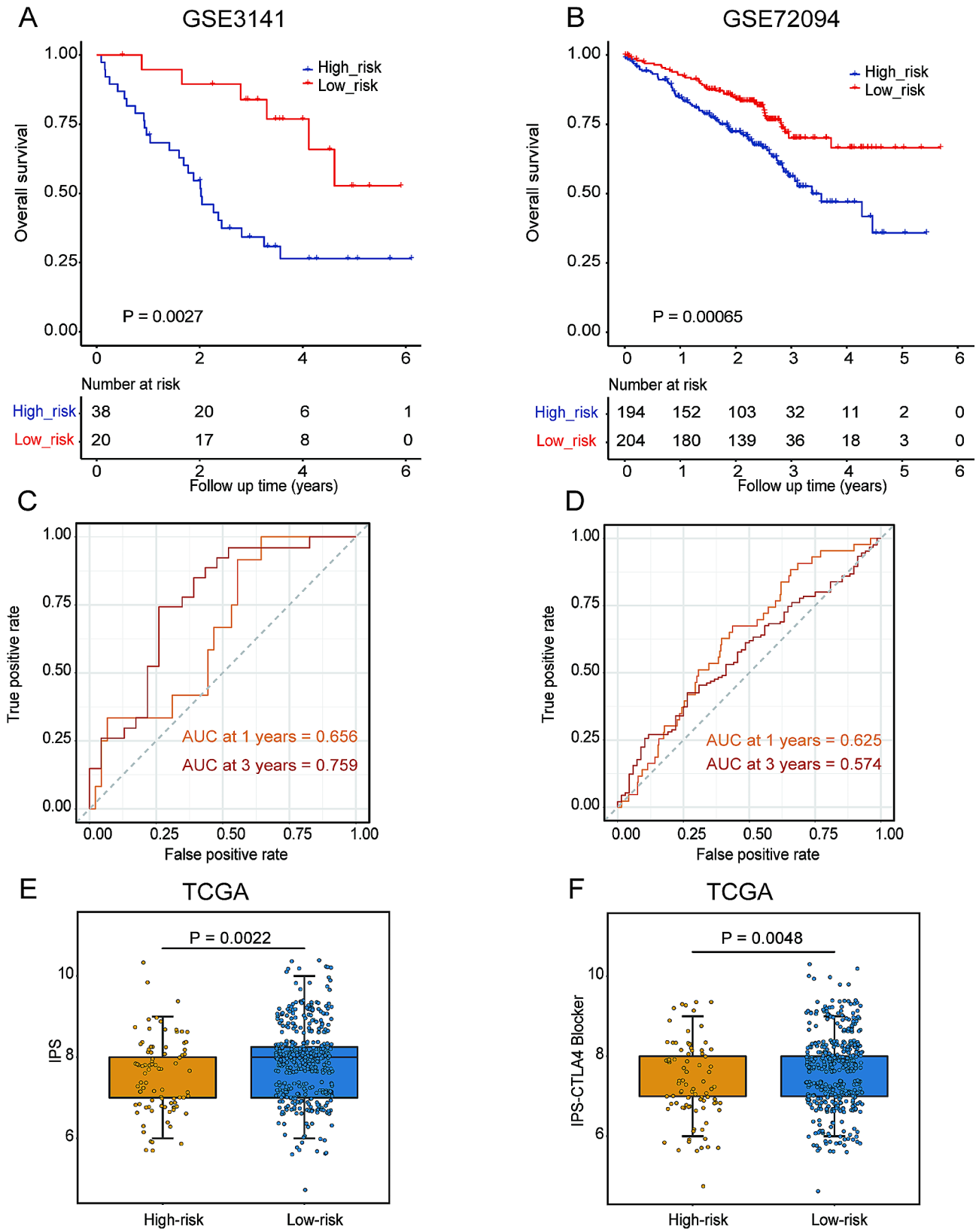
For survival analysis, 94 patients with survival data were included. The results showed that patients with higher expression of MECP2 or lower expression of UHRF1 had longer disease-free time compared with the other group (Fig. 8D). The results were consistent with the survival analysis in the TCGA data (Supplementary



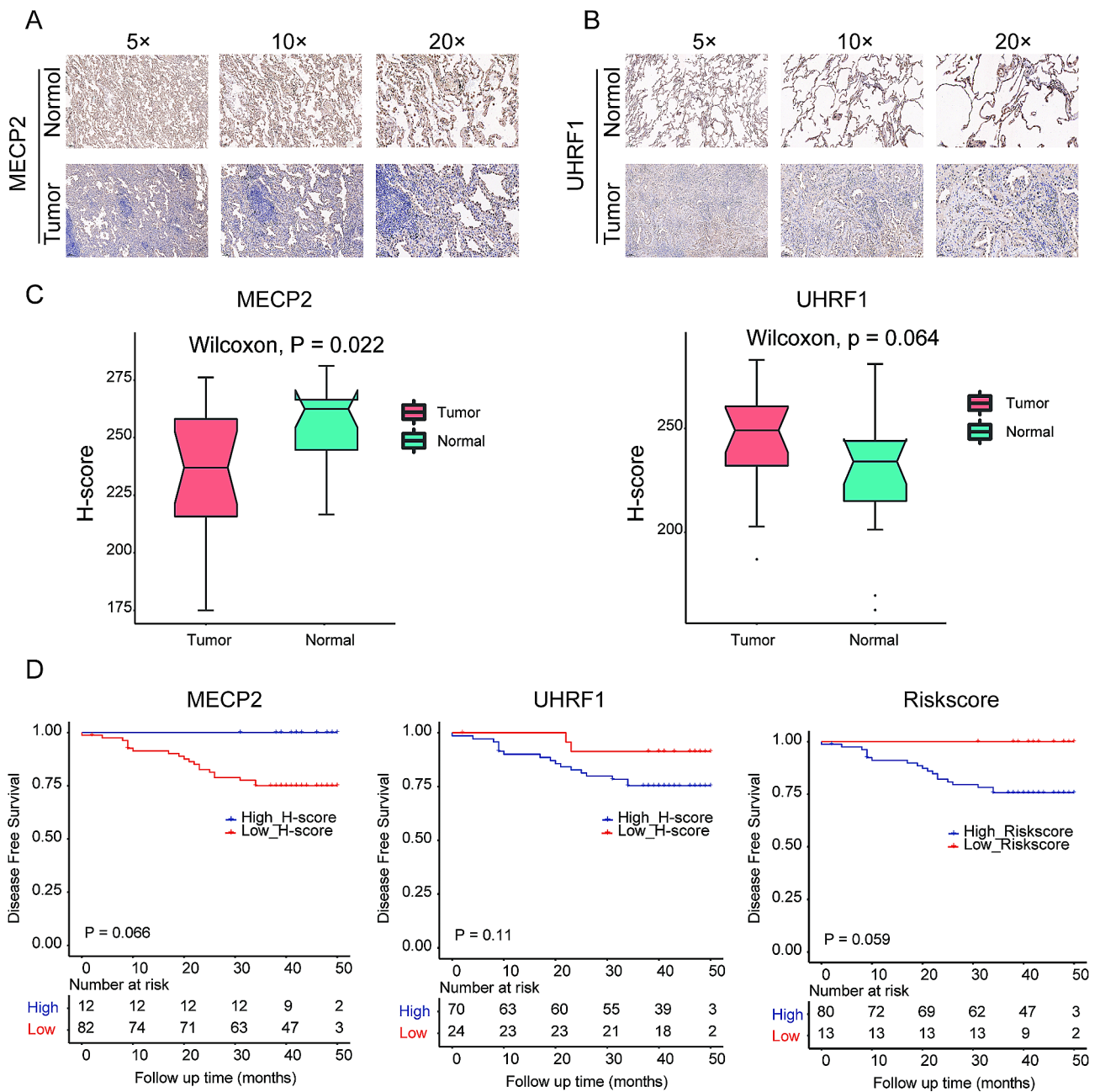
**Fig. 5** Correlation between DMRRS and Tumor Microenvironment. **(A)** GSEA results to distinguish the pathways between high- and low-risk groups. **(B)** Immune scores difference between two risk groups. **(C)** Correlation between low-risk groups and 22 immune cells. **(D)** Correlation analysis of UHRF1 and MECP2 with the infiltration level of the six main immune cells after adjusting for the purity. **(E)** The significant difference in CD8+ T cells and neutrophil infiltration levels between mutant and wild type. **(F)** The enrichment pathways of GSEA varied with different expression of UHRF1 or MECP2



**Fig. 6** The Role of DMRRS in Anti-PD-1 Immunotherapy and Drug Sensitivity. **(A)** The risk score difference in non-responders versus responders. **(B)** The proportion of patients with response to PD-L1 blockade immunotherapy in low- or high-score groups. **(C)** OS analyses for low (8 cases) and high (8 cases) risk patients' groups in the anti-PD-L1 immunotherapy cohort (GSE126044 cohort). **(D)** PFS analyses for low (8 cases) and high (8 cases) risk patients' groups in the anti-PD-L1 immunotherapy cohort (GSE126044 cohort). **(E)** Distribution of the estimated IC50 of Cisplatin, Gemcitabine, Paclitaxel, and Vinblastine between high- and low-score groups in the TCGA and the Large-GEO cohort. **(F)** Distribution of the estimated IC50 of Lapatinib between high- and low-score groups in the TCGA and the Large-GEO cohort. And distribution of the risk scores in different EGFR mutation statuses in the TCGA cohort. PD, progressive disease. PR: partial response. OS: overall survival. PFS: progression-free survival



**Fig. 7** Validation of Prognosis and Predictive Potential in Immunotherapeutic Benefits of DMRRS. **(A-B)** Kaplan-Meier curves of OS in high- and low-risk groups in the GSE3141 and GSE72094 datasets. **(C-D)** ROC curve analyses in the GSE3141 and GSE72094 datasets. **(E-F)** The distribution of IPS in the high-risk and low-risk groups in the TCGA dataset



**Fig. 8** IHC validation of DMRRS Gene Expressions and survival analysis in LUAD Tissues. (A-B) Quantification of protein expression of MECP2 (A) and UHRF1 (B). (C) Comparison of protein expression of MECP2 and UHRF1. (D) Kaplan-Meier curves of DFS in different MECP2 and UHRF1 protein expression groups and for patients with different risk scores in our in-house dataset. DFS, disease-free survival

Fig. 2). Though the signature was constructed with RNA expression, we were curious about the survival value of the same signature formula using the IHC data. The group with higher risk scores exhibited worse DFS (Fig. 8D,  $P=0.059$ ). Interestingly, the signature performed better than both genes alone from the view of  $P$  value. The clinical characteristics of 94 patients are shown in **Supplementary Table 4**. Most patients were at an early stage with fairly good clinical outcomes, which

may contribute to the non-significant  $P$  value in our in-house data survival analysis.

### Discussion

DNA methylation has been widely revealed as a promising target for the development of robust prognostic and antitumor immunity biomarkers [17, 18]. However, since most studies have focused on the role of DNA methylation location, further research focusing on DNA

methylation regulators is warranted to demonstrate the potential regulatory mechanism of DNA methylation.

This study comprehensively analyzed the prognostic effects, tumor microenvironment association, chemotherapy and targeted therapy drug sensitivity, and immune response of DNA methylation regulators. Most regulators were significantly abnormally expressed in tumor tissue than in the adjacent normal tissues. Then we nominated the DMRRS for reliable use as an independent prognostic factor. Concurrently, the signature was also closely related to immune scores and immune cell infiltration levels. In the immune cohort GSE126044, the DMRRS showed an impressive success in selecting responders who could benefit from ICIs. More critically, the use of the DMRRS in those settings, including predicting prognosis and immune response, had been carefully vetted and validated.

The DMRRS consists of UHRF1 and MECP2. Ubiquitin-like with PHD and ring finger domains 1 (UHRF1) is a chromatin modifier, which participates in DNA methylation and can contribute to cancer progression in many tumors, including lung cancer [9, 12, 26]. Researchers also found that by affecting the cell cycle and inducing cell apoptosis, UHRF1 could contribute to the poor prognosis in LUAD recently [36]. Methyl-CpG Binding Protein 2 (MECP2) is a member of the methyl-CpG-binding domain (MBD) family and plays a vital role in chromatin organization [15]. Accumulating studies have proved the potential roles of MECP2 in tumor progression [22, 37]. Thus, it's reasonable to speculate that the DMRRS influences the prognosis of LUAD patients. In our study, the DMRRS did show excellent prognostic value in the Large-GEO training dataset and three validation cohorts, including TCGA, GSE3141, and GSE72094.

Growing evidence has shown that the aberrant expression of DNA methylation regulators could trigger downstream metabolism disorder and is involved in antitumor immunity regulation. For example, TET2 acted as a tumor suppressor in solid tumors and could predict patient response and the efficacy of anti-PD-1/PD-L1 therapy [39]. Based on 21 methylase-related regulators, two DNA methylation modification patterns were identified and correlated with tumor immune-infiltrating microenvironment well in lung cancer [42]. However, a robust DNA methylation regulators-related signature in immunotherapy response remained unexplored. Based on 20 DNA methylation regulators, we constructed the DMRRS that was closely related to CD8+T cells and could effectively predict the immune response in LUAD. Since the higher IPS scores, the better immunotherapy response, we validate that the low-risk patients had higher IPS scores.

The immune checkpoint blockade therapy has revolutionized the treatment of lung cancer. Considering the

good predicting potential of DMRRS in LUAD immune response, we were interested in unraveling why DMRRS can predict the immune response in LUAD. Using the TIMER databases, we revealed that UHRF1 and MECP2 were closely associated with high immune cell infiltration levels in the tumor microenvironment. Also, we performed the single-cell analysis using the TISCH database (Supplementary Fig. 3). The GSE131907 datasets were utilized, including 58 LUAD tissues, and divided into 12 types of cells. In the database, UHRF1 had the highest infiltration level in plasma though the expression was low, and MECP2 had the highest infiltration degree in CD8+T cells. Besides, UHRF1 was the top gene in the dendritic and plasma cells of the GSE131907 cohort. Consistent with our results, UHRF1 regulation of the Akt-mTOR pathway was essential for invariant natural killer T (iNKT) cell survival [4]. In systemic lupus erythematosus (SLE), downregulation of UHRF1 led to T follicular helper (Tfh) cell differentiation both in vitro and in vivo [21]. MECP2 was explicitly bound to the *foxp3* locus, thus promoting *Foxp3* expression and Tregs' resilience [20]. These findings implied that DMRRS might shape the tumor microenvironment for immunotherapy treatment by recruiting immune cells. In fact, targeting UHRF1 in combinational immunotherapy of lung cancer has already been undergone.

In view of the clinical significance, we constructed a tool, namely DMRRS, with an excellent ability to identify LUAD patients who may live longer and benefit from immunotherapy. The predictive value was validated. We also found that the DMRRS may regulate the tumor immunity of LUAD through interacting with the tumor microenvironment immune cells.

Undeniably, there are several limitations in this study. First, there were relatively few stage IV LUAD patients in the TCGA database, which may lead to biases in subsequent studies on DNA methylation regulators and the TNM stage. Second, the samples for IHC staining weren't enough due to a lack of budget and time. Finally, the regulatory mechanism of DNA methylation regulators in LUAD progression and tumor microenvironment was warranted to be further investigated.

## Conclusions

In conclusion, the DMRRS tool or DMRRS-related genes can be a robust biomarker for clinical outcomes and immunotherapy response.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12931-024-02695-4>.

**Supplementary Table 1.** The detailed clinical information of the TCGA-LUAD, GSE19188, GSE30219, GSE31210, GSE3141, GSE37745, GSE50081, and GSE72094 cohorts.

**Supplementary Table 2.** Univariate Cox Analysis Results of 20 DNA Methylation Regulators in TCGA.

**Supplementary Table 3.** Univariate Cox Analysis Results of 20 DNA Methylation Regulators in GEO.

**Supplementary Table 4.** Patient characteristics of our in-house data.

**Supplementary Fig. 1.** ROC curve in the GEO cohorts and survival analysis in the TCGA cohorts regarding different clinical features. (A) ROC curve of 1-, 3- and 5-years. (B-D) OS analysis between high- and low-risk subgroups in regard of different clinical features, including age, gender, and stages. (E) Correlation between high-risk groups and immune cells. ROC: Receiver operating characteristic. DFS: Disease-free survival.

**Supplementary Fig. 2.** Survival analysis of different expression groups of MECP2 and UHRF1 in the TCGA cohort. (A) OS analysis between high- and low-expression subgroups of MECP2 and (B) UHRF1 in TCGA-LUAD cohort. OS: Overall survival.

**Supplementary Fig. 3.** The single-cell analysis using the TISCH database. (A) The cell types and their distribution in the GSE131907 of the TISCH database. (B) The pie chart showed different cell types and their distribution in the GSE131907 cohort. (C and D) The expression of UHRF1 and MECP2 in different immune cells in the GSE131907 dataset. (E and F) The distribution of UHRF1 and MECP2 in different cell types in the GSE131907. TME: tumor microenvironment; TISCH: Tumor Immune Single Cell Hub.

#### Acknowledgements

We acknowledge the TCGA and GEO databases for providing their platforms and contributors for uploading their meaningful datasets.

#### Author contributions

Songhua Cai and Jian Zeng designed the study. Jing Huang, Chujian Huang and Can Huang collected the data. Zichang Xiang and Yao Ni performed the data analysis and interpreted the data. Jing Huang, Chujian Huang and Can Huang drafted the manuscript. Songhua Cai and Jian Zeng revised the manuscript. All authors read and approved the final manuscript.

#### Funding

This study was supported by Shenzhen High-level Hospital Construction Fund, Shenzhen Key Medical Discipline Construction Fund (No. SZXK075); Sanming Project of Medicine in Shenzhen (No. SZSM201612097); and Cancer Hospital, Chinese Academy of Medical Sciences, Shenzhen Center/Shenzhen Cancer Hospital Research Project (No.SZ2020QN009 and No.SZ2020ZD010).

#### Data availability

Data used in this study from the TCGA and GEO databases can be accessed without restriction. In-house immunohistochemistry (IHC) images can be accessible through email to the corresponding author Songhua Cai on reasonable request.

#### Declarations

#### Ethical approval

This study was performed in accordance with the Declaration of Helsinki and approved by the Ethics and Research Committees of the National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College.

#### Consent for publication

Written informed consent for publication was obtained.

#### Conflict of interest

None.

Received: 13 September 2023 / Accepted: 18 January 2024

Published online: 05 February 2024

#### References

- 1 Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WE, Poddubskaya E, et al. Nivolumab versus Docetaxel in Advanced squamous-cell non-small-cell Lung Cancer. *N Engl J Med*. 2015;373(2):123–35. <https://doi.org/10.1056/NEJMoa1504627>.
- 2 Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep*. 2017;18(1):248–62. <https://doi.org/10.1016/j.celrep.2016.12.019>.
- 3 Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Carolini D, et al. TCGA-biolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016;44(8):e71. <https://doi.org/10.1093/nar/gkv1507>.
- 4 Cui Y, Chen X, Zhang J, Sun X, Liu H, Bai L, et al. Uhrf1 controls iNKT cell survival and differentiation through the Akt-mTOR Axis. *Cell Rep*. 2016;15(2):256–63. <https://doi.org/10.1016/j.celrep.2016.03.016>.
- 5 Davis S, Meltzer PS. GEOquery: a bridge between the Gene expression Omnibus (GEO) and BioConductor. *Bioinformatics*. 2007;23(14):1846–7. <https://doi.org/10.1093/bioinformatics/btm254>.
- 6 Fedchenko N, Reifennath J. Different approaches for interpretation and reporting of immunohistochemistry analysis results in the bone tissue - a review. *Diagn Pathol*. 2014;9:221. <https://doi.org/10.1186/s13000-014-0221-9>.
- 7 Ferris RL, Blumenschein G Jr, Fayette J, Guigay J, Colevas AD, Licitra L, et al. Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck. *N Engl J Med*. 2016;375(19):1856–67. <https://doi.org/10.1056/NEJMoa1602252>.
- 8 Gandhi L, Rodriguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, et al. Pembrolizumab plus Chemotherapy in Metastatic Non-small-cell Lung Cancer. *N Engl J Med*. 2018;378(22):2078–92. <https://doi.org/10.1056/NEJMoa1801005>.
- 9 Gao SP, Sun HF, Li LD, Fu WY, Jin W. UHRF1 promotes breast cancer progression by suppressing KLF17 expression by hypermethylating its promoter. *Am J Cancer Res*. 2017;7(7):1554–65.
- 10 Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004;20(3):307–15. <https://doi.org/10.1093/bioinformatics/btg405>.
- 11 Gleeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS ONE*. 2014;9(9):e107468. <https://doi.org/10.1371/journal.pone.0107468>.
- 12 Goto D, Komeda K, Uwatoko N, Nakashima M, Koike M, Kawai K, et al. UHRF1, a Regulator of methylation, as a diagnostic and prognostic marker for Lung Cancer. *Cancer Invest*. 2020;38(4):240–9. <https://doi.org/10.1080/07357907.20.1747483>.
- 13 Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013;14:7. <https://doi.org/10.1186/1471-2105-14-7>.
- 14 Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer*. 2019;19(3):133–50. <https://doi.org/10.1038/s41568-019-0116-x>.
- 15 Hite KC, Adams VH, Hansen JC. Recent advances in MeCP2 structure and function. *Biochem Cell Biol*. 2009;87(1):219–27. <https://doi.org/10.1139/o08-115>.
- 16 Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, et al. Genomic and transcriptomic features of response to Anti-PD-1 therapy in metastatic melanoma. *Cell*. 2017;168(3):542. <https://doi.org/10.1016/j.cell.2017.01.010>.
- 17 Jung H, Kim HS, Kim JY, Sun JM, Ahn JS, Ahn MJ, et al. DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load. *Nat Commun*. 2019;10(1):4278. <https://doi.org/10.1038/s41467-019-12159-9>.
- 18 Koch A, Joosten SC, Feng Z, de Ruijter TC, Draht MX, Melotte V, et al. Analysis of DNA methylation in cancer: location revisited. *Nat Rev Clin Oncol*. 2018;15(7):459–66. <https://doi.org/10.1038/s41571-018-0004-4>.
- 19 Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882–3. <https://doi.org/10.1093/bioinformatics/bts034>.
- 20 Li C, Jiang S, Liu SQ, Lykken E, Zhao LT, Sevilla J, et al. MeCP2 enforces Foxp3 expression to promote regulatory T cells' resilience to inflammation. *Proc Natl Acad Sci U S A*. 2014;111(27):E2807–2816. <https://doi.org/10.1073/pnas.1401505111>.
- 21 Liu L, Hu L, Yang L, Jia S, Du P, Min X, et al. UHRF1 downregulation promotes T follicular helper cell differentiation by increasing BCL6 expression in SLE. *Clin Epigenetics*. 2021;13(1):31. <https://doi.org/10.1186/s13148-021-01007-7>.

- 22 Luo D, Ge W. MeCP2 promotes Colorectal Cancer Metastasis by modulating ZEB1 transcription. *Cancers (Basel)*. 2020;12(3). <https://doi.org/10.3390/cancers12030758>.
- 23 Meng Q, Lu YX, Ruan DY, Yu K, Chen YX, Xiao M, et al. DNA methylation regulator-mediated modification patterns and tumor microenvironment characterization in gastric cancer. *Mol Ther Nucleic Acids*. 2021;24:695–710. <https://doi.org/10.1016/j.omtn.2021.03.023>.
- 24 Morales-Nebreda L, McLafferty FS, Singer BD. DNA methylation as a transcriptional regulator of the immune system. *Transl Res*. 2019;204:1–18. <https://doi.org/10.1016/j.trsl.2018.08.001>.
- 25 Ott PA, Bang YJ, Piha-Paul SA, Razak ARA, Bennouna J, Soria JC, et al. T-Cell-inflamed gene-expression Profile, programmed death Ligand 1 expression, and Tumor Mutational Burden Predict Efficacy in patients treated with Pembrolizumab Across 20 cancers: KEYNOTE-028. *J Clin Oncol*. 2019;37(4):318–27. <https://doi.org/10.1200/jco.2018.78.2276>.
- 26 Reardon ES, Shukla V, Xi S, Gara SK, Liu Y, Straughan D, et al. UHRF1 is a Novel Druggable Epigenetic Target in Malignant Pleural Mesothelioma. *J Thorac Oncol*. 2021;16(1):89–103. <https://doi.org/10.1016/j.jtho.2020.08.024>.
- 27 Riaz N, Havel JJ, Kendall SM, Makarov V, Walsh LA, Desrichard A, et al. Recurrent SERPINB3 and SERPINB4 mutations in patients who respond to anti-CTLA4 immunotherapy. *Nat Genet*. 2016;48(11):1327–9. <https://doi.org/10.1038/ng.3677>.
- 28 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. <https://doi.org/10.1093/nar/gkv007>.
- 29 Schübeler D. Function and information content of DNA methylation. *Nature*. 2015;517(7534):321–6. <https://doi.org/10.1038/nature14192>.
- 30 Shi J, Hua X, Zhu B, Ravichandran S, Wang M, Nguyen C, et al. Somatic Genomics and Clinical features of lung adenocarcinoma: a retrospective study. *PLoS Med*. 2016;13(12):e1002162. <https://doi.org/10.1371/journal.pmed.1002162>.
- 31 Siegel R, Miller K, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(1):7–30. <https://doi.org/10.3322/caac.21590>.
- 32 Soria JC, Ohe Y, Vansteenkiste J, Reungwetwattana T, Chewaskulyong B, Lee KH, et al. Osimertinib in untreated EGFR-Mutated Advanced Non-small-cell Lung Cancer. *N Engl J Med*. 2018;378(2):113–25. <https://doi.org/10.1056/NEJMoa1713137>.
- 33 Suarez-Alvarez B, Rodriguez RM, Fraga MF, López-Larrea C. DNA methylation: a promising landscape for immune system-related diseases. *Trends Genet*. 2012;28(10):506–14. <https://doi.org/10.1016/j.tig.2012.06.005>.
- 34 Sun D, Wang J, Han Y, Dong X, Ge J, Zheng R, et al. TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. *Nucleic Acids Res*. 2021;49(D1):D1420–d1430. <https://doi.org/10.1093/nar/gkaa1020>.
- 35 Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med*. 2012;366(26):2443–54. <https://doi.org/10.1056/NEJMoa1200690>.
- 36 Tu Z, Deng X, Hou S, Feng A, Zhang Q. UHRF1 predicts poor prognosis by triggering cell cycle in lung adenocarcinoma. *J Cell Mol Med*. 2020;24(14):8069–77. <https://doi.org/10.1111/jcmm.15438>.
- 37 Wang L, Gao Y, Tong D, Wang X, Guo C, Guo B, et al. MeCP2 drives hepatocellular carcinoma progression via enforcing HOXD3 promoter methylation and expression through the HB-EGF/EGFR pathway. *Mol Oncol*. 2021;15(11):3147–63. <https://doi.org/10.1002/1878-0261.13019>.
- 38 Wu HX, Chen YX, Wang ZX, Zhao Q, He MM, Wang YN, et al. Alteration in TET1 as potential biomarker for immune checkpoint blockade in multiple cancers. *J Immunother Cancer*. 2019;7(1):264. <https://doi.org/10.1186/s40425-019-0737-3>.
- 39 Xu YP, Lv L, Liu Y, Smith MD, Li WC, Tan XM, et al. Tumor suppressor TET2 promotes cancer immunity and immunotherapy efficacy. *J Clin Invest*. 2019;129(10):4316–31. <https://doi.org/10.1172/jci129317>.
- 40 Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612. <https://doi.org/10.1038/ncomms3612>.
- 41 Yu J, Qin B, Moyer AM, Newshean S, Liu T, Qin S, et al. DNA methyltransferase expression in triple-negative breast cancer predicts sensitivity to decitabine. *J Clin Invest*. 2018;128(6):2376–88. <https://doi.org/10.1172/jci97924>.
- 42 Yuan D, Wei Z, Wang Y, Cheng F, Zeng Y, Yang L, et al. DNA methylation Regulator-Mediated modification patterns define the distinct Tumor Micro-environment in Lung Adenocarcinoma. *Front Oncol*. 2021;11:734873. <https://doi.org/10.3389/fonc.2021.734873>.
- 43 Zhang T, Zhao Y, Zhao Y, Zhou J. Expression and prognosis analysis of TET family in acute myeloid leukemia. *Aging*. 2020;12(6):5031–47. <https://doi.org/10.18632/aging.102928>.

## Publisher's Note

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