

Molecular subtypes, prognostic and immunotherapeutic relevant gene signatures mediated by DNA methylation regulators in hepatocellular carcinoma

Rongfeng Shi^{1,*}, Hui Zhao^{1,*}, Suming Zhao¹, Hongxin Yuan¹

¹Department of Interventional Radiology, Affiliated Hospital of Nantong University, Nantong 226001, Jiangsu, P.R. China

*Equal contribution

Correspondence to: Suming Zhao, Hongxin Yuan; **email:** sumingzhaontu@163.com, <https://orcid.org/0000-0002-9944-0692>; yuanhongxin1880@aliyun.com, <https://orcid.org/0000-0003-2025-8810>

Keywords: DNA methylation, immunotherapy, prognosis, biomarker, hepatocellular carcinoma

Received: March 25, 2022

Accepted: June 14, 2022

Published: June 30, 2022

Copyright: © 2022 Shi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Growing evidence has revealed the crucial role of epigenetics in tumor progression and immune response. However, the molecular subtypes and their microenvironment characterization mediated by DNA methylation regulators in hepatocellular carcinoma remain little known. In this study, we comprehensively integrated the transcriptome profiling of twenty DNA methylation regulators in hepatocellular carcinoma. Consensus clustering was used to identify distinct methylation regulator-related molecular subtypes. The prognostic DMS signature was constructed using principal components analysis. Most regulators experienced a low genomic variation, but we found a remarkably difference in mRNA expression of these regulators between normal and tumor tissues. Three distinct methylation regulator-related molecular subtypes were successfully identified according to the expression of 20 regulators, which had substantially different biological characteristics and prognosis. The classic carcinogenic pathways and stromal activity including TGF-beta, p53 and WNT signaling pathway were significantly activated in subtype B, leading to a survival inferiority in subtype B compared to other two subtypes. Further analysis demonstrated the constructed DMS signature was an independent predictive biomarker in patient prognosis. Two anti-checkpoint immunotherapy cohorts demonstrated patients with high DMS presented significantly improved treatment advantages and enhanced responses especially the survival prolonged. Generally, the high DMS groups improved more than 15% clinical response to immunotherapy than low DMS groups. In conclusion, this study identified three DNA methylation regulator-related subtypes with distinct clinical, molecular and biological characteristics, and constructed a prognostic and immunotherapeutic relevant gene signature. It might help to promote individualized immunotherapy for hepatocellular carcinoma from the perspective DNA methylation regulators.

INTRODUCTION

DNA methylation, as one of the epigenetic modifications most well characterized, was involved in multiple biological processes in mammals. The main form of DNA methylation modification was 5mC, indicating that DNA methylation occurs on the fifth carbon atom of CpG dinucleotide cytosine residues

[1–3]. Although technological advancements and novel mechanistic insights have altered strategies for treating hepatocellular carcinoma over the past decade, these improvements can only benefit a small number of patients, with five-year survival rates less than 50% [4–6]. However, a small group of patients with robust responses achieved astounding survival benefits from anti-checkpoint immunotherapy, which was represented

by the anti-PD-1/PD-L1 antibodies. Unfortunately, for the great majority of patients, the benefits are either minimal or nonexistent, far from meeting a clinical need. Furthermore, clinical response rates differ both within and between tumor types, indicating there existed intrinsic and adaptive immune resistance to anti-checkpoint immunotherapy [7–9]. It was reported that several genes, regulated by hypomethylation or hypermethylation of the promoter, were significantly correlated with the initiation and progression of hepatocellular carcinoma. Qian et al. reported that the stemness and tumorigenicity in hepatocellular carcinoma were regulated by DNMT1-mediated methylation of BEX1 [10]. Luming et al. revealed that the enforced HOXD3 promoter methylation mediated by MeCP2 was involved in hepatocellular carcinoma progression via HB-EGF/EGFR pathway [11]. However, the molecular subtypes mediated by DNA methylation regulators and their roles in patient survival and immunotherapeutic efficacy remain unknown. In this study, we comprehensively analyzed the genomic characterization of 20 DNA methylation regulators, and their mediated molecular subtypes in hepatocellular carcinoma. We successfully defined three DNA methylation regulator-related molecular subtypes with distinct prognostic features and biological functions in hepatocellular carcinoma based on 369 TCGA samples. We also constructed DNA methylation related signature (DMS) based on the overlapping differentially expressed genes across three subtypes, which was confirmed to be significantly related to the patient's prognosis and efficacy of anti-checkpoint immunotherapy.

RESULTS

Genomic characterization of DNA methylation regulators in hepatocellular carcinoma

We extracted a total of 20 DNA methylation regulators from TCGA-LIHC and GSE14520 datasets, including 14 readers (SMUG1, NTHL1, TDG, UNG, MECP2, UHRF1, UHRF2, ZBTB4, ZBTB33, ZBTB38, MBD1, MBD2, MBD3, MBD4), 3 erasers (TET1, TET2, TET3) and 3 writers (DNMT1, DNMT3A, DNMT3B) [12]. The analysis of mutational landscape for 20 DNA methylation regulators showed that in hepatocellular carcinoma, only 33 samples of 364 samples experienced at least one mutation, accounting for 9.07% (Figure 1A). The CNV analyses indicated these regulators also a relatively low CNV alteration frequency in hepatocellular carcinoma. TET2, MBD3, UHRF1, DNMT1, MBD2 and MBD2 mainly focused on the frequency of copy number deletion, while DNMT3A, ZBTB33, MECP2 and NTHL1 exhibited copy number amplifications (Figure 1B). The position

of 20 regulators in chromosome was shown in Figure 1C. Based on the TCGA-LIHC cohort, we found the expression of DNA methylation regulators between normal tissues and tumor tissues showed significant difference. TET2, ZBTB4, MBD4, ZBTB38 and NTHL1 was significantly down-regulated in tumor tissues, while TET3, SMUG1, MBD1, TET1, UNG, DNMT3A, MECP2, DNMT3B, DNMT1 and UHRF1 was significantly up-regulated in tumor tissues (Figure 1D). Survival analyses with the univariate Cox regression model revealed their crucial roles in the patient outcomes, of these, DNMT3A, UHRF1, DNMT1, DNMT3B and TET1 the risk factors for hepatocellular carcinoma. However, other regulators did not show a significant effect on prognosis (Figure 1E). Additionally, we found tumors with the up-regulated eraser genes showed a high expression of writer genes, except for TET2 eraser gene (Supplementary Figure 1A–1C).

Consensus clustering for identifying methylation regulator-related molecular subtypes

Based on the protein-protein interaction (PPI) network constructed by the STRING, we found these regulators showed a widespread protein interaction (Figure 2A). Further analysis showed that these regulators exhibited significant positive correlations at mRNA expression levels (Figure 2B). A network was used to visualize the prognostic significance and interaction among these regulators.(Figure 2B) Previous studies showed that DNA methylation play anti-tumor immune effects by regulating cell infiltration in the tumor microenvironment [13, 14]. We therefore explored the correlation between tumor environment cell infiltration and DNA methylation regulators. We found ZBTB4, MBD2 and DNMT1 present a significant correlation with most cell infiltration abundance (Figure 2C). The consensus clustering was used to identify the distinct molecular subtypes mediated by DNA methylation regulators in hepatocellular carcinoma. According to the expression of 20 regulators, all tumor samples were clustered into three distinct subtypes. We termed these subtypes as methylation regulator-related molecular subtype A, B and C, respectively (Supplementary Figure 2). Survival analysis revealed a significant difference on prognosis between three subtypes, with particularly prominent survival advantage in subtype A and C, and the survival inferiority in subtype B (Figure 2E). NTHL1, MBD3 and SMUG1 were characteristically expressed in subtype A, DNMT3A, TET1, DNMT3B, TET3, UHRF2, DNMT1, UHRF1, TDG and UNG characteristically expressed in subtype B, while ZBTB4, MBD4, ZBTB33, TET2 and ZBTB38 characteristically expressed in subtype C (Figure 2D–2F).

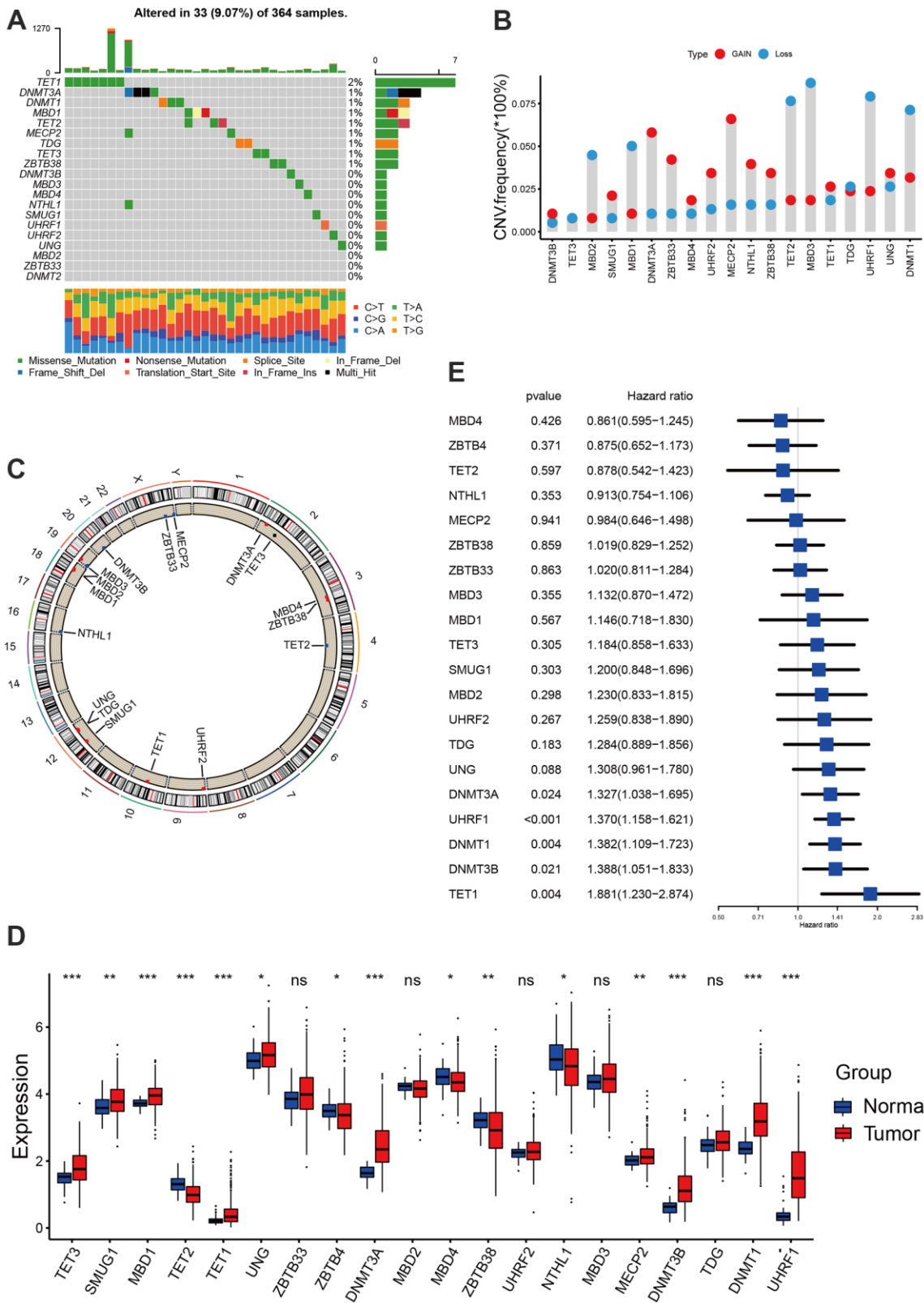


Figure 1. Landscape of DNA methylation regulators in hepatocellular carcinoma. (A) The mutation landscape of 20 DNA methylation regulators in TCGA-LIHC cohort. (B) The Copy number variation frequency of 20 DNA methylation regulators. (C) The position of the 20 regulators in the chromosome. (D) Expression of 20 regulators in tumor and normal samples based on TCGA-cohort. (E) Survival analyses for the 20 regulator genes using univariate Cox regression model. All data analyses were based on the TCGA-LIHC cohort. CNV, copy number variation. Ns, not significant. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

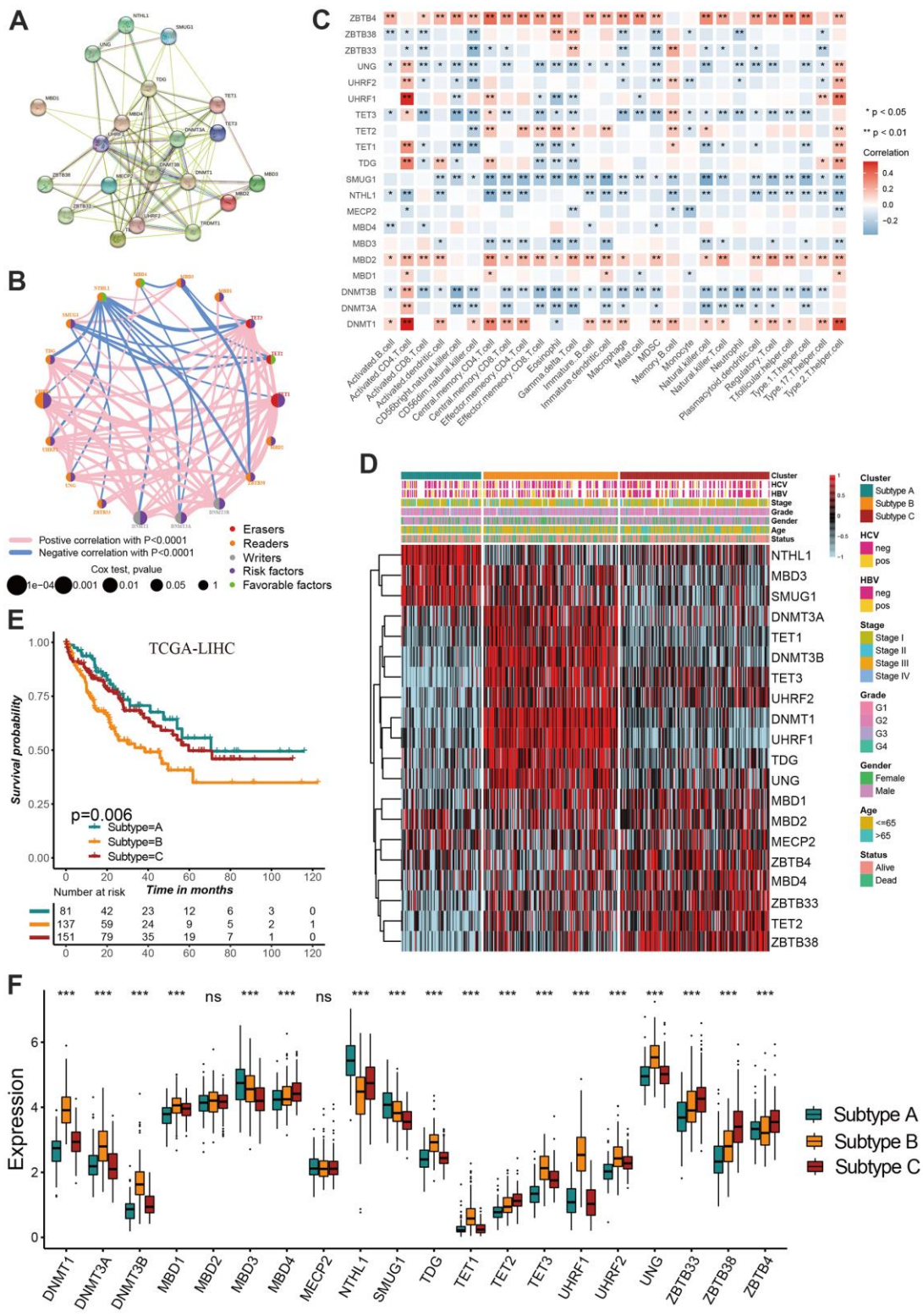


Figure 2. Identification of DNA methylation regulator-related molecular subtypes in hepatocellular carcinoma. (A) The protein-protein interactions (PPI) network between DNA methylation regulators using STRING database. (B) A network was used to visualize the prognostic significance and expression correlation among these regulators. (C) The correlation between DNA methylation regulator expression and immune cell infiltration levels. (D) The hierarchical clustering of 20 DNA methylation regulators among three molecular subtypes. (E) Survival analyses for three distinct DNA methylation regulator-related molecular subtypes. (F) DNA methylation regulators expressed in the three molecular subtypes. All data analyses were based on the TCGA-LIHC cohort. Neg, negative; Pos, positive; Ns, not significant. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Characteristics of distinct methylation regulator-related molecular subtypes

Considering the association between tumor microenvironment and DNA methylation, we investigate the difference in immune infiltration abundance among three subtypes. The infiltration level of activated CD8 T cell, activated CD4 T cell, CD56bright natural killer cell, gamma delta T cell, neutrophil and Type 2 T helper cell were significantly different between the three molecular subtypes (Figure 3A). Although subtype B showed a relatively immune cell infiltration, numerous classic carcinogenic pathways and stromal activity were significantly activated in subtype B including WNT signaling pathway, p53 signaling pathway and TGF-beta signaling pathway, leading to a survival inferiority compared to other two subtypes (Figure 3B, 3C). The transcriptome difference between three subtypes was investigated to further reveal the underlying biological characteristics in these three molecular subtypes, and a total of 1037 overlapping differentially expressed genes (DEGs) were obtained (Figure 3D). The GO enrichment analysis demonstrated the DEGs were prominently associated with the DNA methylation related biological pathways (Figure 3E).

Construction of prognostic signature based on overlapping DEGs

We have showed that these 1037 genes were closely related to the three molecular subtypes we have identified. Therefore, we again performed consensus clustering on the expression of these 1037 genes to determine the ability to reproduce these three subtypes in hepatocellular carcinoma. The results confirmed that, based on the expression of 1037 genes, we could still identify three distinct subtypes in hepatocellular carcinoma, we termed these subtypes derived from these 1037 gene expression as Gene cluster A, B and C (Supplementary Figure 3 and Figure 4A). Survival analysis for these three groups showed a significant difference, prominent survival advantage in Gene.cluster A and C, and the survival inferiority in Gene.cluster B (Figure 4B). It was found the three gene clusters exhibited specific DNA methylation-related transcriptome characterization, respectively (Figure 4C). Further analyses indicated DNMT1, DNMT3A and DNMT3B showed a significantly high expression in Gene.cluster B compared to other two clusters, while ZBTB33, ZBTB38, ZBTB4 were remarkably up-regulated in Gene.cluster C (Figure 4D). Considering the crucial role of methylation regulator mediated molecular subtypes in prognosis, we constructed the prognostic signature based on prognosis-related overlapping

DEGs using the PCA algorithm, which we termed as DMS signature (Supplementary Table 1). The alluvial diagram also revealed the changes of sample attributes including methylation regulator-related molecular subtypes, Gene.cluster, survival status and DMS (Figure 4E). We performed the correlation analysis between DMS score and 20 methylation regulator expression based on TCGA cohort, and found DMS score had the most significant positive correlation with NTHL1 ($cor=0.58$, Supplementary Figure 4), and the most significant negative correlation with DNMT1 ($cor=-0.75$, Supplementary Figure 4). Additionally, a significant distinction on DMS between methylation regulator-related molecular subtypes as well as between Gene.clusters was observed. Subtype A and Gene.cluster A displayed a highest median DMS, while the subtype B and Gene.cluster B displayed a lowest median DMS (Figure 4F, 4G). Hepatitis B virus positive patients presented a lower DMS (Figure 5A), while hepatitis C virus positive patients presented a higher DMS (Figure 5B). We did not observe the significant difference on DMS between low and high ALT patients (Figure 5C). Based on the optimal cut-point at -1.88 acquired from MaxStat algorithm, patients were classified as high and low DMS group (Figure 5D). Patients in the high DMS group experienced a remarkable survival benefit compared to low DMS group (Figure 5E). We then used the external validation set GSE14520 to validate the prognostic value of DMS in hepatocellular carcinoma. We still observed a significant survival advantage in high DMS group than low DMS group (Figure 5F). Although the statistical P value was insignificant, patients with high DMS still experienced an advantage trend in recurrence-free survival compared to those with low DMS. The landscape of somatic mutation in the low and high DMS groups was then further investigated. The tumor mutation burden did not show a significant difference across low and high DMS groups (Figures 5H, 5I, 6A, 6B). We used the waterfall plot to summarize the distinction on tumor mutation burden between low and high DMS groups (Figure 5H, 5I). We used the multivariate Cox regression model to further reveal the value of DMS in predicting patient prognosis, and found DMS signature was as an independent biomarker in predicting patient outcomes (Figure 6C). However, given the range of hazard ratio 95% CI (0.96 to 0.99, $p = 0.008$), we still needed to be cautious about this result. We then perform GSEA enrichment analysis, and found the classic carcinogenic signaling pathways were significantly activated in the low DMS group such as MAPK, MTOR, P53, TGF-beta and WNT signaling pathways, which could lead to the worse prognosis in patients with low DMS (Figure 6D).



Figure 3. Characteristics of distinct DNA methylation regulator-related molecular subtypes. (A) The abundance of 28 tumor microenvironment cell infiltration among three molecular subtypes. (B, C) GSEA enrichment showing the activation states of biological pathways in distinct molecular subtypes. (D) The Venn diagram showing 1037 overlapping differentially expressed genes (DEGs) between three DNA methylation regulator-related molecular subtypes. (E) GO functional enrichment analyses for 1037 overlapping differentially expressed genes. All data analyses were based on the TCGA-LIHC cohort. Neg, negative; Pos, positive; Ns, not significant. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

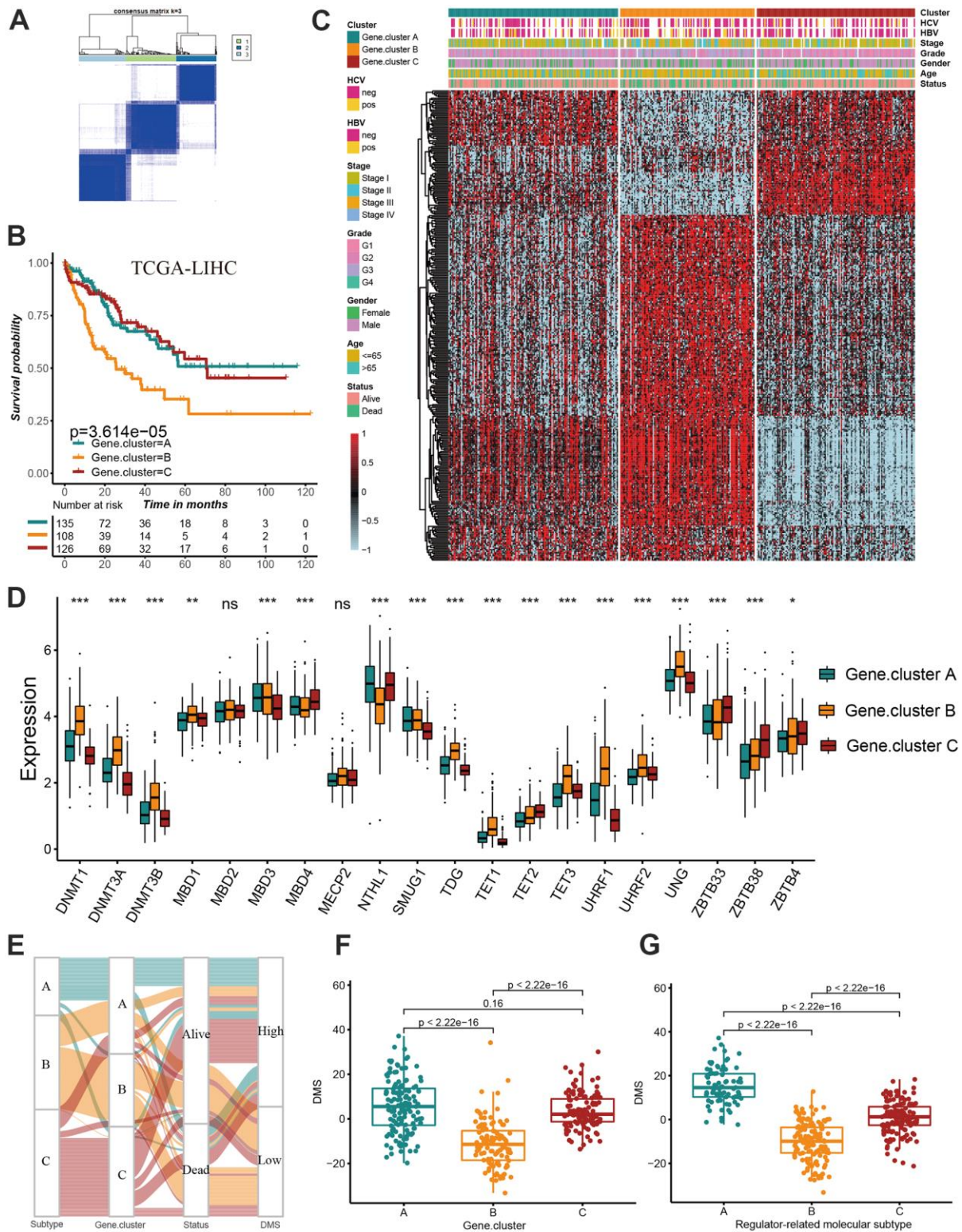


Figure 4. Construction of prognosis-related DMS signature. (A) Consensus matrices of DNA methylation subtype-related genes. for k=3. (B) Survival analyses for three distinct Gene.clusters. (C) The hierarchical clustering of 1037 overlapping differentially expressed genes among three Gene.clusters. (D) Difference in the 20 DNA methylation regulator expression among three Gene.clusters. (E) The alluvial diagram showing the changes of sample attributes including methylation regulator-related molecular subtypes, Gene.cluster, survival status and DMS. (F) Differences in DMS score across three Gene.clusters. (G) Differences in DMS score across three distinct methylation regulator-related molecular subtypes. All data analyses were based on the TCGA-LIHC cohort. Neg, negative; Pos, positive; Ns, not significant. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

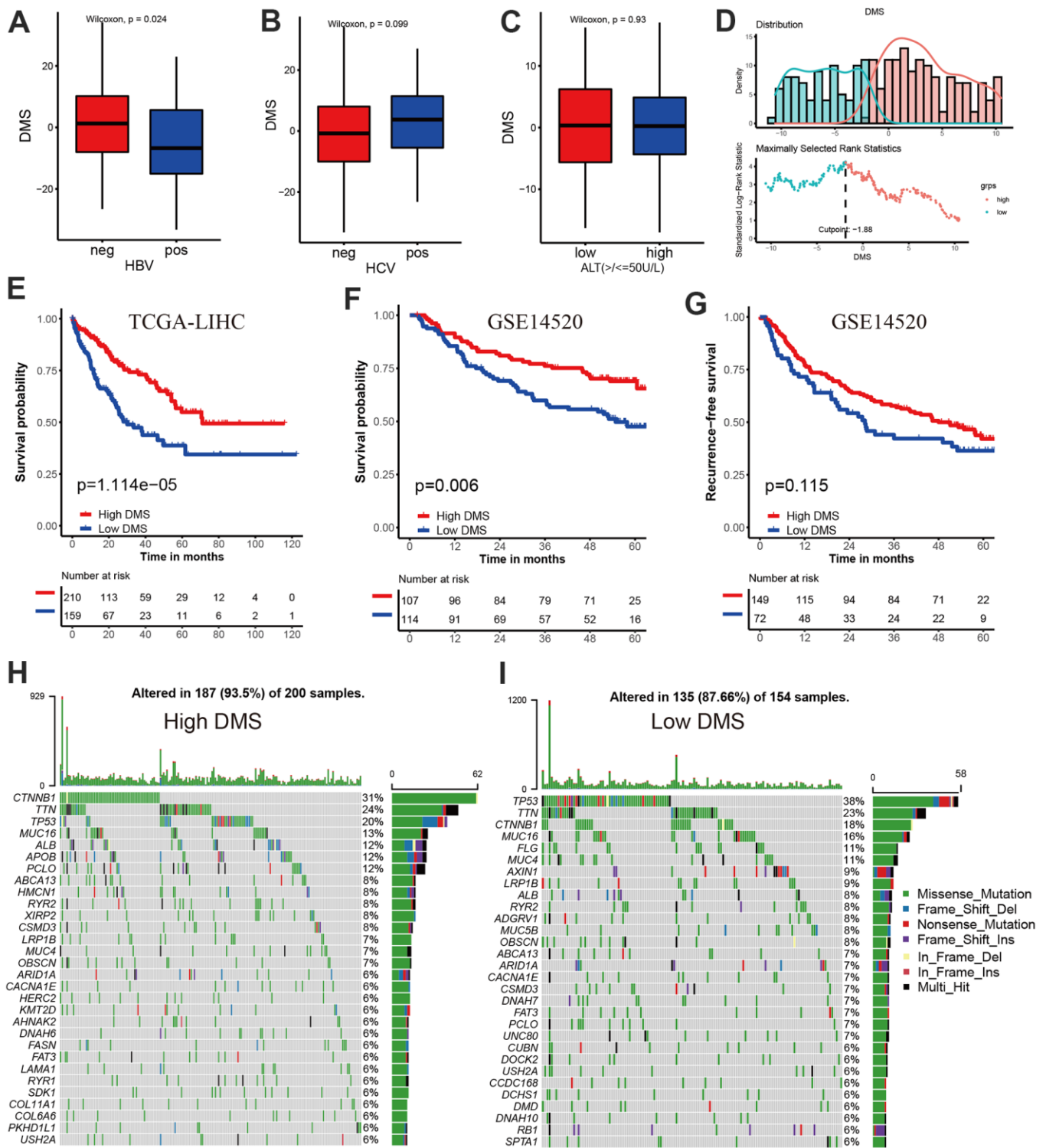


Figure 5. Prognostic value and external validation of DMS signature. (A) Differences in DMS score between hepatitis B virus positive and negative group in TCGA-LIHC cohort. (B) Differences in DMS score between hepatitis C virus positive and negative group in TCGA-LIHC cohort. (C) Differences in DMS score between ALT high and low group in GSE14520 cohort. (D) The MaxStat R package identified the optimal cut-off point to dichotomize DMS. (E) Kaplan-Meier curves showing the survival difference between the low and high DMS groups in TCGA-LIHC cohort. (F) External validation the value of DMS in predicting patient prognosis in GSE14520 cohort. (G) Kaplan-Meier curves showing the recurrence-free survival difference between the low and high DMS groups in GSE14520 cohort. (H, I) The waterfall plot showing the differences of TMB landscape between low and high DMS groups in TCGA-LIHC cohort. (H) High DMS group. (I) Low DMS group. Neg, negative; Pos, positive.

Role of DMS in predicting efficacy of immunotherapy

The breakthrough of immunotherapy represented by immune checkpoints in cancer treatment has brought encouraging and instructive strategies for the successful cure of cancer. To further revealed the predictive values of DMS signature in patients treated with anti-checkpoint immunotherapy, we collected two

immunotherapy cohorts with completed survival information including IMvigor210 cohort with the intervention of PD-L1 antibody and TCGA-SKCM cohort with the intervention of PD-1 and CTLA-4 antibody. In the IMvigor210 cohort, compared with patients with low DMS score, patients with high DMS score had a prominent clinical benefit and therapeutic advantage. The survival in the high-DMS group was

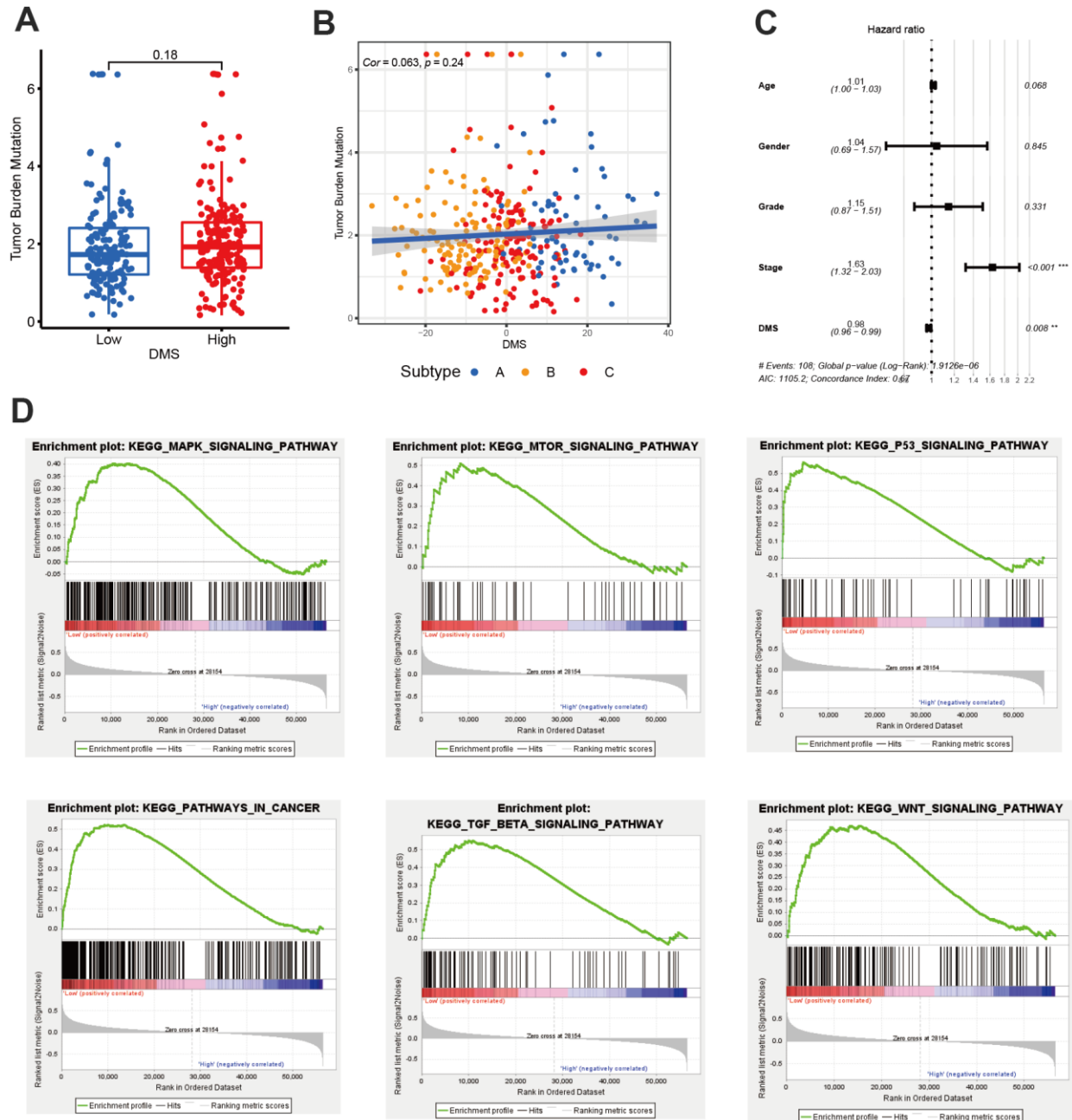


Figure 6. The value of DMS in predicting clinical outcomes in patients with hepatocellular carcinoma. (A) Differences in tumor burden mutation between low and high DMS groups. (B) The correlation between tumor burden mutation and DMS score. (C) Multivariate cox regression analysis for DMS in predicting patient's survival in TCGA-LIHC cohort. (D) GSEA enrichment analysis showing the activated biological pathways in patients with low DMS. All data analyses were based on the TCGA-LIHC cohort. Cor, correlation.

significantly prolonged (Figure 7A–7C). In the TCGA-SKCM cohort, we still observed the treatment advantages in the high-DMS group than low-DMS group, emphasizing the predictive value of DMS signature in patient receiving immune checkpoint blockade therapy (Figure 7D–7G). In addition, when stratifying patients by DMS and neoantigen mutational burden, we found that patients with neoantigen burden as well as high DMS features had significantly improved survival when receiving immunotherapy (Figure 7H).

DISCUSSION

With the in-depth understanding of the heterogeneity and complexity of tumor microenvironment, increasing evidence highlights the critical role of DNA methylation in inducing immune escape and immunotherapeutic drug resistance [15–18]. Nevertheless, DNA methylation regulator mediated molecular subtypes, microenvironmental characteristics under these subtypes as well as the impacts on immunotherapeutic efficacy still remain little known. Identifying the distinct molecular subtypes in order for classifying patients will promote the development of individualized treatment of cancer. Additionally, identification of the DNA methylation regulator mediated molecular subtypes may help reveal potential biomarkers significantly related to clinical response to immunotherapy and potentially uncover novel immunotherapeutic targets [19, 20].

In this research, we comprehensively integrated the transcriptome data of 20 DNA methylation regulators. Although all the regulators experienced a low genomic variation, there existed a remarkably difference in mRNA expression between normal and tumor tissues. According to the expression of 20 regulators, we successfully revealed three distinct DNA methylation regulator-related molecular subtypes in hepatocellular carcinoma, which had substantially different biological characteristics and prognosis. Subtype A was characterized by the higher expression of NTHL1, MBD3 and SMUG1, with a particularly prominent survival advantage; Subtype B characterized by the up-regulation of DNMT3A, TET1, DNMT3B, TET3, UHRF2, DNMT1, UHRF1 and TDG, with a significantly survival inferiority; while ZBTB4, MBD4, ZBTB33, TET2 and ZBTB38 characteristically expressed in subtype C whose prognosis was similar with subtype A. The immune infiltration analysis for the three molecular subtypes showed the activated CD4 T cell and activated CD8 T cell presented a higher infiltration level in the subtype B whose survival was bad. We then further analyzed the characteristics of the biological function of each molecular subtype to clarify the reasons for the

poor prognosis of subtype B with high levels of immune infiltration. We found numerous classic carcinogenic signaling pathways and stromal activity including WNT, p53 and TGF-beta signaling pathway were significantly activated in subtype B, which could result in a survival inferiority in subtype B. Subtype B exhibits characteristics of stromal activation, which may mediate immune escape of this subtype. Previous studies have classified tumors into three immune subtypes, including immune-inflamed, immune-excluded, and immune-desert. Although the immune-excluded subtype shows similar immune activation characteristics to the immune-inflamed subtype, the activated immune cells are localized around tumor cell nests without infiltrating into the tumor parenchyma, which leads to the false immune activation. Consistent with the characteristics of the immune-excluded subtype, immunotherapy can activate immune cells around the tumor, but cannot stimulate their infiltration into the tumor interior, resulting in a low response rate of such tumors to immunotherapy. [21–25]. TGF-beta promotes the growth, infiltration and metastasis of tumor cells by inducing immune escape, promoting blood vessel formation, and inducing epithelial-mesenchymal transition [26, 27]. In the previous studies, Zhang et al. also revealed three subtypes with distinct molecular, tumor microenvironment and clinical characterization in gastric cancer based on the expression of 21 m⁶A RNA methylation regulators, of these, the tumor microenvironment characterization under m⁶A cluster B subtype was highly consistent with immune-excluded phenotype [28]. Shen et al. reported three subtypes with distinct metabolic characteristics using the expression of 23 m⁶A RNA methylation regulators in hepatocellular carcinoma, and constructed m⁶A score signature to predict the prognosis and treatment response [29]. The above suggested that tumor heterogeneity could be further revealed by molecular classification of tumors by a specific gene set. At present, the transformation of cold tumors into hot tumors by targeting the immunophenotype of the tumor has become a hot topic in the field of cancer research. Previous studies have shown that MYC amplification mediated CCL23, CCL5, PD-L1, CD47 and IL1 β expression decreased could induce macrophages and DCs inactivation, and also limit the recruitment of T cells, B cells and natural killer cells [30, 31]. Considering the high expression of DNMT3A, TET1, DNMT3B, TET3, UHRF2, DNMT1, UHRF1 and TDG in Subtype B, changing the tumor microenvironment cell infiltration characteristics by reversing expression of these DNA methylation regulators may be more clinically practical. Based on the overlapping differentially expressed genes, we constructed DMS signature to further reveal the value of these molecular subtypes in evaluating patient's prognosis and efficacy of immunotherapy. Subtype A

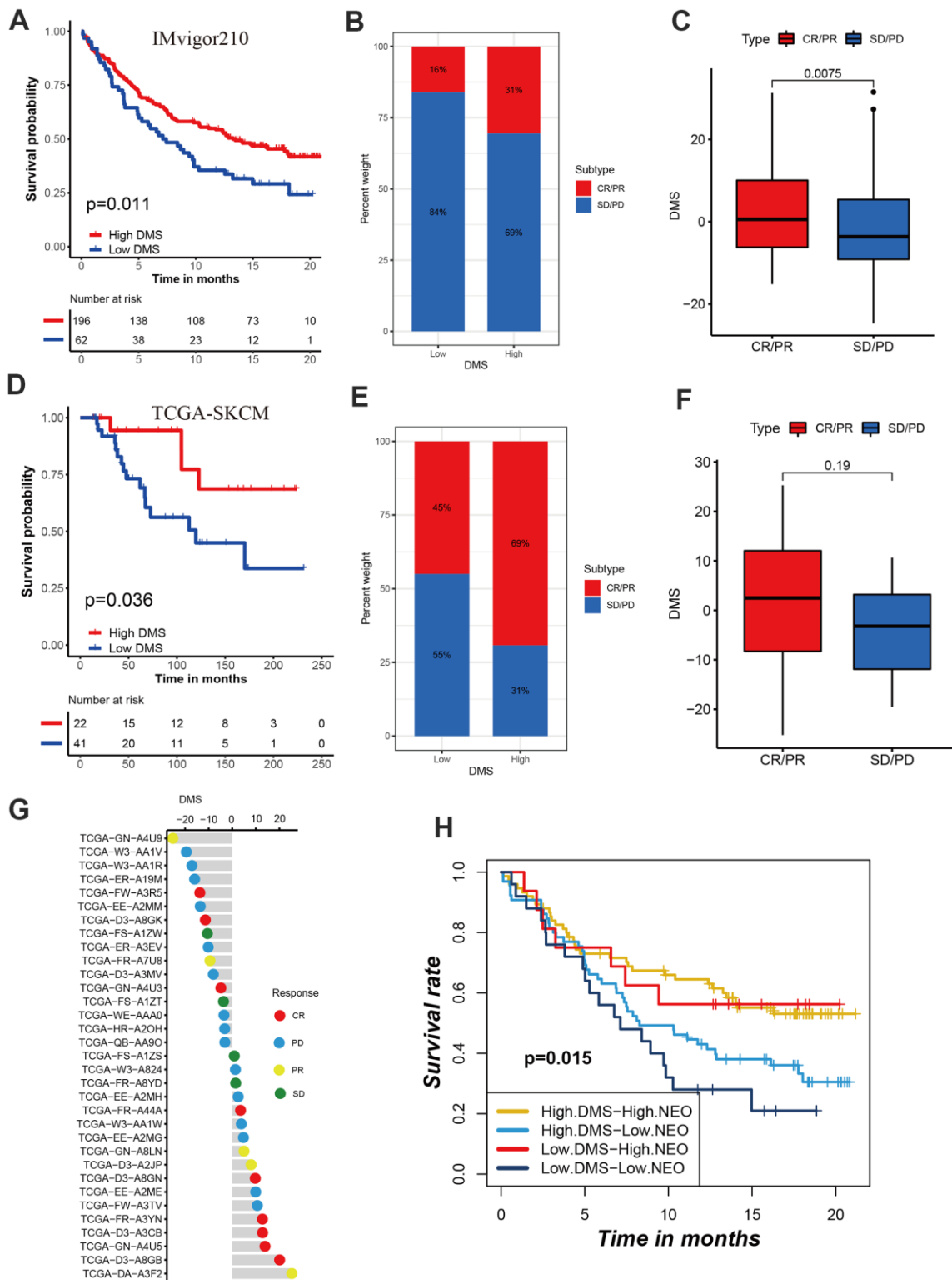


Figure 7. Role of DMS in predicting efficacy of immunotherapy. (A) Kaplan-Meier curves displaying the survival difference of high and low DMS groups in IMvigor210 cohort. (B) The ratio of clinical response types in high DMS and low DMS groups in the IMvigor210 cohort when treated with anti-PD-1 immunotherapy. (C) Differences in DMS score between different clinical response types in the IMvigor210 cohort. (D) Survival analyses for DMS in TCGA-SKCM cohort. (E) The ratio of clinical response types in each group in the TCGA-SKCM cohort. (F, G) Differences in DMS score between different clinical response types in the TCGA-SKCM cohort. (H) Survival analyses for patients receiving anti-PD-L1 immunotherapy stratified by both neoantigen burden and DMS signature. (I) Differences in DMS score among different immune phenotypes. CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease. NEO, neoantigen burden.

exhibited a high DMS and subtype B showed a low DMS. DMS signature was proved to be an independent biomarker to predict the prognosis of patients. Additionally, DMS was correlated with hepatitis B and C. Similar to the subtype B, low DMS group was significantly enriched in the classic carcinogenic signaling pathways such as MAPK, MTOR, P53, TGF-beta and WNT signaling pathways. Using two anti-PD-1/L1 and anti-CTLA immunotherapeutic cohorts, we found the DMS signature could predict the patient response to immunotherapy. The checkpoint immunotherapy significantly improved the clinical response and prolonged the survival in patients with high DMS score compared with those with low DMS score. Generally, the high DMS groups improved more than 15% clinical response to immunotherapy than low DMS groups. This suggested that the DNA methylation regulator related gene signature could predict the efficacy and clinical responses in patients treated with anti-checkpoint immunotherapy.

CONCLUSIONS

This study identified three DNA methylation regulator mediated subtypes with distinct clinical, molecular and biological characteristics in hepatocellular carcinoma, and constructed DMS signature, which could serve as an independent predictive biomarker in patient survival and response to immunotherapy. It may help promote individualized immunotherapy for hepatocellular carcinoma from the perspective DNA methylation regulators.

MATERIALS AND METHODS

Sample datasets collection and processing

In total, we collected 20 DNA methylation regulators based on existed published studies [12]. For training cohorts, we downloaded RNA sequencing data of TCGA-LIHC with FPKM types from The Cancer Genome Atlas (TCGA) Genomic Data Commons (GDC) Data Portal (<https://portal.gdc.cancer.gov/>) via TCGAbiolinks R package [32]. Then, we transformed the FPKM value into TPM values [33]. The copy number variation (CNV) data and somatic mutation were acquired from UCSC Xena public data hubs (<http://xena.ucsc.edu/>). The corresponding clinical information were curated from the TCGA GDC. The GSE14520 from Gene Expression Omnibus (GEO) database served as validation set [34]. The GSE14520, which was based on the Affymetrix Human Genome U133A 2.0 Array (GPL571), investigated the gene expression subtypes in tumor and paired non-tumor tissue of HCC patients as well as healthy donor liver. We used the affy R package to perform data

preprocessing [35]. A total of 369 patients in TCGA-LIHC cohort and 221 patients in GSE14520 cohort with completed survival information were selected for further analysis. The median age was 59 years and 51 years in TCGA-LIHC and GSE14520 cohort, respectively. 255 patients and 170 patients were diagnosed with stage I/II in TCGA-LIHC and GSE14520 cohort, respectively. The baseline characteristics of patients in the TCGA-LIHC and GSE14520 cohorts was presented in Table 1. We included two immunotherapy cohorts after a systematical publicly search: The IMvigor210 cohort, which investigated the PD-L1 antibody in advanced urothelial cancer, was acquired from <http://research-pub.Gene.com/imvigor300corebiologies>. The raw count data was also transformed into TPM value. The TCGA-SKCM cohort, which investigated PD-1 and CTLA-4 antibody in advanced melanoma, was acquired from TCGA GDC data portal. The FPKM data was downloaded and then converted to TPM value.

Crosstalk among DNA methylation regulators

We constructed an expression-survival network to reveal a relationship of DNA methylation regulator connection, interactions and prognosis in hepatocellular carcinoma. We identified the protein-protein interactions (PPI) between DNA methylation regulators using the STRING database [36].

Consensus clustering for twenty DNA methylation regulators

In order to identify distinct DNA methylation regulator mediated molecular subtypes in hepatocellular carcinoma, we performed consensus molecular clustering according to the mRNA expression of twenty DNA methylation regulators. The R package ConsensuClusterPlus was used and 1,000 times for subtype clustering were repeated in order for classification stability [37].

Gene set variation analysis (GSVA)

The GSVA enrichment analysis was used to reveal the specified activated biological pathways among distinct DNA methylation subtypes. The enrichment score represented the relative activity of each biological pathway. The hallmark and 'c2.cp.kegg.v6.2.symbols' get sets were acquired from the Molecular Signatures Database (MSigDB) for GSVA. We also estimated the tumor microenvironment abundance of immune cell infiltration including activated CD4 T cell, CD8 T cell, B cell and other cell types. The gene sets for estimating infiltration abundances were obtained from the published study [28, 38–40].

Table 1. Baseline characteristics of patients in the TCGA-LIHC and GSE14520 cohorts.

Cohorts	TCGA-LIHC (n=369)	GSE14520 (n=221)
Age, years	59 (16 - 94)	51 (21 - 77)
Sex		
Male	248	191
Female	121	30
Status		
Alive	239	136
Dead	130	85
Stage		
Stage I/II	255	170
Stage III/ IV	90	49
Unknown	24	2
Hepatitis B virus infection		
Positive	44	212
Negative	147	6
Unknown	178	3

Identification of differentially expressed genes and functional annotation

We identified the differences in mRNA transcriptome between the three molecular subtypes in hepatocellular carcinoma with the Limma R package [41]. The p value less than 0.001 was the criterion for screening the differentially expressed genes. Then we applied the R package clusterProfiler to perform the functional annotation for these differentially expressed genes [42]. The term BP (biological process) was selected for revealing the biological function of these genes.

Construction of prognosis related DMS signature

First, we adopted the univariate Cox regression model to reveal the prognostic value for these differentially expressed genes in hepatocellular carcinoma. Then the principal component analysis (PCA) was performed for the expression of genes with the prognosis P value <0.05. The signature scores were composed of principal components 1 (PC1) and 2 (PC2) [28, 43]. The DMS signature was defined as follows:

$$DMS = \Sigma(PC1_i + PC2_i)$$

where i is the expression of subtypes-related genes with a significant prognosis.

Statistical analysis

The Kruskal-Wallis and One-way ANOVA tests was used to execute the difference significance test for three groups or more [44]. The difference analyses between

the two groups was based on Wilcoxon test. The survival curves with the basis of log-rank tests and the Kaplan-Meier method were generated with the survminer R package. We classified patients into low and high DMS groups through the optimal cut-off point obtained from the MaxStat R package [45]. All statistical P-values were two-sided, and a p < 0.05 was statistically significant. All data was processed through the software R 4.0.5.

Data availability statement

All data related to this work can be acquired from the Gene-Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo>) and the GDC portal (<https://portal.gdc.cancer.gov/>).

AUTHOR CONTRIBUTIONS

HXY, SMZ, RFS and HZ designed this study. RFS and HZ were responsible for the integration and analyses of the data. RFS and HZ wrote this manuscript. HXY, SMZ and RFS revised the manuscript. All authors approved this manuscript.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (grant number 82102167),

Natural Science Foundation of Jiangsu Province of China (grant number BK20180946), the Scientific Research Project of Nantong Municipal Health Commission (grant number QA2019060), the Science Foundation of China Postdoctoral (grant number 2019M651928), the Natural Science Foundation of Nantong City (grant number MS12020028).

REFERENCES

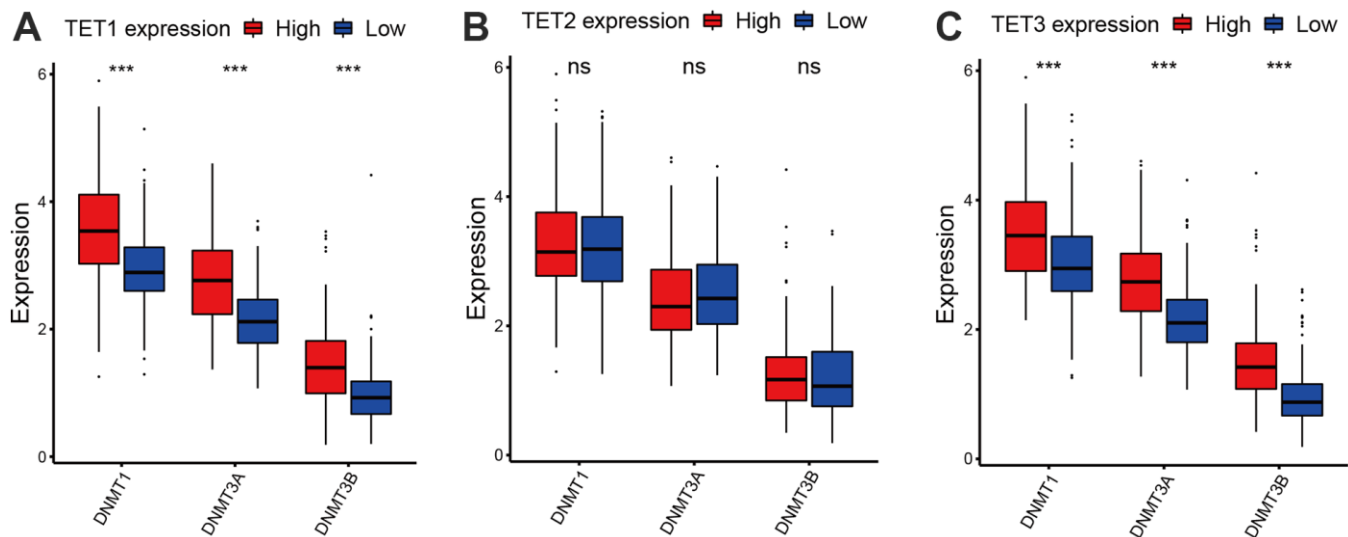
1. Li E, Zhang Y. DNA methylation in mammals. *Cold Spring Harb Perspect Biol.* 2014; 6:a019133. <https://doi.org/10.1101/cshperspect.a019133> PMID:[24789823](https://pubmed.ncbi.nlm.nih.gov/24789823/)
2. Schübeler D. Function and information content of DNA methylation. *Nature.* 2015; 517:321–6. <https://doi.org/10.1038/nature14192> PMID:[25592537](https://pubmed.ncbi.nlm.nih.gov/25592537/)
3. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nat Rev Genet.* 2013; 14:204–20. <https://doi.org/10.1038/nrg3354> PMID:[23400093](https://pubmed.ncbi.nlm.nih.gov/23400093/)
4. Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Abate D, Abbasi N, Abbastabar H, Abd-Allah F, Abdel-Rahman O, Abdelalim A, Abdoli A, Abdollahpour I, Abdulle ASM, Abebe ND, Abraha HN, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol.* 2019; 5:1749–68. <https://doi.org/10.1001/jamaoncol.2019.2996> PMID:[31560378](https://pubmed.ncbi.nlm.nih.gov/31560378/).
Erratum in: *JAMA Oncol.* 2020; 6:444. <https://doi.org/10.1001/jamaoncol.2020.0224> PMID:[32163169](https://pubmed.ncbi.nlm.nih.gov/32163169/).
Erratum in: *JAMA Oncol.* 2020; 6:789. <https://doi.org/10.1001/jamaoncol.2020.0741> PMID:[32271356](https://pubmed.ncbi.nlm.nih.gov/32271356/).
Erratum in: *JAMA Oncol.* 2021; 7:466. <https://doi.org/10.1001/jamaoncol.2020.8307> PMID:[33507210](https://pubmed.ncbi.nlm.nih.gov/33507210/)
5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020; 70:7–30. <https://doi.org/10.3322/caac.21590> PMID:[31912902](https://pubmed.ncbi.nlm.nih.gov/31912902/)
6. Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat Rev Gastroenterol Hepatol.* 2019; 16:589–604. <https://doi.org/10.1038/s41575-019-0186-y> PMID:[31439937](https://pubmed.ncbi.nlm.nih.gov/31439937/)
7. Ali HR, Chlon L, Pharoah PD, Markowitz F, Caldas C. Patterns of Immune Infiltration in Breast Cancer and Their Clinical Implications: A Gene-Expression-Based Retrospective Study. *PLoS Med.* 2016; 13:e1002194. <https://doi.org/10.1371/journal.pmed.1002194> PMID:[27959923](https://pubmed.ncbi.nlm.nih.gov/27959923/)
8. Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM, Waldner MJ, Bindea G, Mlecnik B, Galon J, Trajanoski Z. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. *Genome Biol.* 2015; 16:64. <https://doi.org/10.1186/s13059-015-0620-6> PMID:[25853550](https://pubmed.ncbi.nlm.nih.gov/25853550/)
9. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015; 12:453–7. <https://doi.org/10.1038/nmeth.3337> PMID:[25822800](https://pubmed.ncbi.nlm.nih.gov/25822800/)
10. Wang Q, Liang N, Yang T, Li Y, Li J, Huang Q, Wu C, Sun L, Zhou X, Cheng X, Zhao L, Wang G, Chen Z, et al. DNMT1-mediated methylation of BEX1 regulates stemness and tumorigenicity in liver cancer. *J Hepatol.* 2021; 75:1142–53. <https://doi.org/10.1016/j.jhep.2021.06.025> PMID:[34217777](https://pubmed.ncbi.nlm.nih.gov/34217777/)
11. Wang L, Gao Y, Tong D, Wang X, Guo C, Guo B, Yang Y, Zhao L, Zhang J, Yang J, Qin Y, Liu L, Huang C. MeCP2 drives hepatocellular carcinoma progression via enforcing HOXD3 promoter methylation and expression through the HB-EGF/EGFR pathway. *Mol Oncol.* 2021; 15:3147–63. <https://doi.org/10.1002/1878-0261.13019> PMID:[34028973](https://pubmed.ncbi.nlm.nih.gov/34028973/)
12. Meng Q, Lu YX, Ruan DY, Yu K, Chen YX, Xiao M, Wang Y, Liu ZX, Xu RH, Ju HQ, Qiu MZ. 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> PMID:[33996253](https://pubmed.ncbi.nlm.nih.gov/33996253/)
13. Wu HX, Chen YX, Wang ZX, Zhao Q, He MM, Wang YN, Wang F, Xu RH. Alteration in TET1 as potential biomarker for immune checkpoint blockade in multiple cancers. *J Immunother Cancer.* 2019; 7:264. <https://doi.org/10.1186/s40425-019-0737-3> PMID:[31623662](https://pubmed.ncbi.nlm.nih.gov/31623662/)
14. Xu YP, Lv L, Liu Y, Smith MD, Li WC, Tan XM, Cheng M, Li Z, Bovino M, Aubé J, Xiong Y. Tumor suppressor TET2 promotes cancer immunity and immunotherapy efficacy. *J Clin Invest.* 2019; 129:4316–31. <https://doi.org/10.1172/JCI129317> PMID:[31310587](https://pubmed.ncbi.nlm.nih.gov/31310587/)

15. Cao J, Yan Q. Cancer Epigenetics, Tumor Immunity, and Immunotherapy. *Trends Cancer*. 2020; 6:580–92. <https://doi.org/10.1016/j.trecan.2020.02.003> PMID:[32610068](https://pubmed.ncbi.nlm.nih.gov/32610068/)
16. Jones PA, Issa JP, Baylin S. Targeting the cancer epigenome for therapy. *Nat Rev Genet*. 2016; 17:630–41. <https://doi.org/10.1038/nrg.2016.93> PMID:[27629931](https://pubmed.ncbi.nlm.nih.gov/27629931/)
17. Furlan M, Galeota E, de Pretis S, Caselle M, Pelizzola M. m6A-Dependent RNA Dynamics in T Cell Differentiation. *Genes (Basel)*. 2019; 10:28. <https://doi.org/10.3390/genes10010028> PMID:[30626100](https://pubmed.ncbi.nlm.nih.gov/30626100/)
18. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med*. 2013; 19:1423–37. <https://doi.org/10.1038/nm.3394> PMID:[24202395](https://pubmed.ncbi.nlm.nih.gov/24202395/)
19. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, Coussens LM, Gaborilovich DI, Ostrand-Rosenberg S, Hedrick CC, Vonderheide RH, Pittet MJ, Jain RK, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med*. 2018; 24:541–50. <https://doi.org/10.1038/s41591-018-0014-x> PMID:[29686425](https://pubmed.ncbi.nlm.nih.gov/29686425/)
20. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014; 513:202–9. <https://doi.org/10.1038/nature13480> PMID:[25079317](https://pubmed.ncbi.nlm.nih.gov/25079317/)
21. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature*. 2017; 541:321–30. <https://doi.org/10.1038/nature21349> PMID:[28102259](https://pubmed.ncbi.nlm.nih.gov/28102259/)
22. Hegde PS, Karanikas V, Evers S. The Where, the When, and the How of Immune Monitoring for Cancer Immunotherapies in the Era of Checkpoint Inhibition. *Clin Cancer Res*. 2016; 22:1865–74. <https://doi.org/10.1158/1078-0432.CCR-15-1507> PMID:[27084740](https://pubmed.ncbi.nlm.nih.gov/27084740/)
23. Herbst RS, Soria JC, Kowanetz M, Fine GD, Hamid O, Gordon MS, Sosman JA, McDermott DF, Powderly JD, Gettinger SN, Kohrt HE, Horn L, Lawrence DP, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature*. 2014; 515:563–7. <https://doi.org/10.1038/nature14011> PMID:[25428504](https://pubmed.ncbi.nlm.nih.gov/25428504/)
24. Joyce JA, Fearon DT. T cell exclusion, immune privilege, and the tumor microenvironment. *Science*. 2015; 348:74–80. <https://doi.org/10.1126/science.aaa6204> PMID:[25838376](https://pubmed.ncbi.nlm.nih.gov/25838376/)
25. Salmon H, Franciszkiewicz K, Damotte D, Dieu-Nosjean MC, Validire P, Trautmann A, Mami-Chouaib F, Donnadieu E. Matrix architecture defines the preferential localization and migration of T cells into the stroma of human lung tumors. *J Clin Invest*. 2012; 122:899–910. <https://doi.org/10.1172/JCI45817> PMID:[22293174](https://pubmed.ncbi.nlm.nih.gov/22293174/)
26. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, Kadel EE II, Koepfen H, Astarita JL, Cubas R, Jhunjunwala S, Banchereau R, Yang Y, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature*. 2018; 554:544–8. <https://doi.org/10.1038/nature25501> PMID:[29443960](https://pubmed.ncbi.nlm.nih.gov/29443960/)
27. Tauriello DV, Palomo-Ponce S, Stork D, Berenguer-Llargo A, Badia-Ramentol J, Iglesias M, Sevillano M, Ibiza S, Cañellas A, Hernando-Momblona X, Byrom D, Matarin JA, Calon A, et al. TGFβ drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature*. 2018; 554:538–43. <https://doi.org/10.1038/nature25492> PMID:[29443964](https://pubmed.ncbi.nlm.nih.gov/29443964/)
28. Zhang B, Wu Q, Li B, Wang D, Wang L, Zhou YL. m6A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. *Mol Cancer*. 2020; 19:53. <https://doi.org/10.1186/s12943-020-01170-0> PMID:[32164750](https://pubmed.ncbi.nlm.nih.gov/32164750/)
29. Shen X, Hu B, Xu J, Qin W, Fu Y, Wang S, Dong Q, Qin L. The m6A methylation landscape stratifies hepatocellular carcinoma into 3 subtypes with distinct metabolic characteristics. *Cancer Biol Med*. 2020; 17:937–52. <https://doi.org/10.20892/j.issn.2095-3941.2020.0402> PMID:[33299645](https://pubmed.ncbi.nlm.nih.gov/33299645/)
30. Casey SC, Tong L, Li Y, Do R, Walz S, Fitzgerald KN, Gouw AM, Baylot V, Gütgemann I, Eilers M, Felsner DW. MYC regulates the antitumor immune response through CD47 and PD-L1. *Science*. 2016; 352:227–31. <https://doi.org/10.1126/science.aac9935> PMID:[26966191](https://pubmed.ncbi.nlm.nih.gov/26966191/)
31. Kortlever RM, Sodir NM, Wilson CH, Burkhart DL, Pellegrinet L, Brown Swigart L, Littlewood TD, Evan GI. Myc Cooperates with Ras by Programming Inflammation and Immune Suppression. *Cell*. 2017; 171:1301–15.e14. <https://doi.org/10.1016/j.cell.2017.11.013> PMID:[29195074](https://pubmed.ncbi.nlm.nih.gov/29195074/)
32. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Carolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, Noushmehr H. TCGAAbiolinks: an R/Bioconductor package for

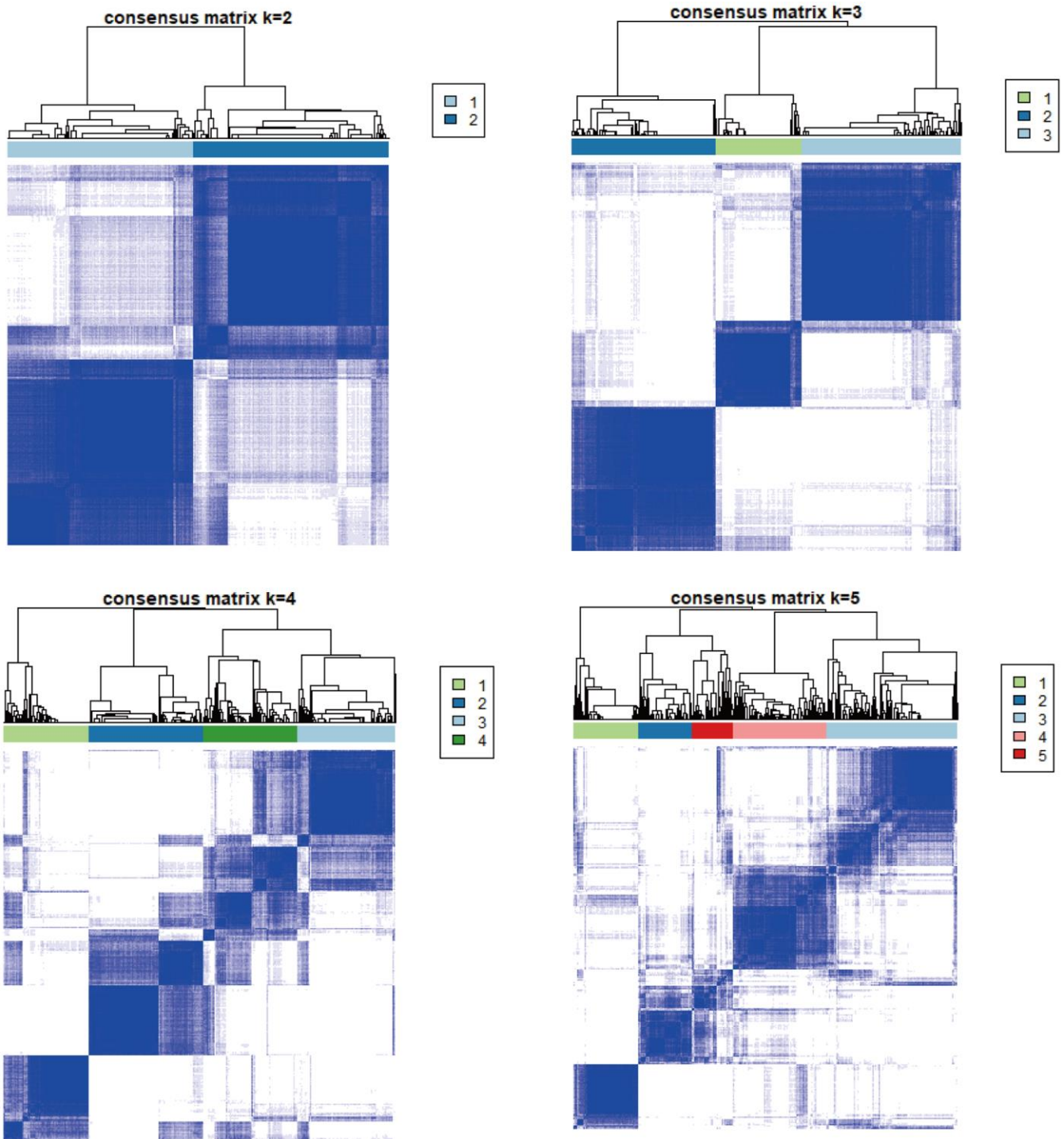
- integrative analysis of TCGA data. *Nucleic Acids Res.* 2016; 44:e71.
<https://doi.org/10.1093/nar/gkv1507> PMID:26704973
33. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 2012; 131:281–5.
<https://doi.org/10.1007/s12064-012-0162-3> PMID:22872506
34. Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, Thorgeirsson SS, Sun Z, Tang ZY, Qin LX, Wang XW. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Res.* 2010; 70:10202–12.
<https://doi.org/10.1158/0008-5472.CAN-10-2607> PMID:21159642
35. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics.* 2004; 20:307–15.
<https://doi.org/10.1093/bioinformatics/btg405> PMID:14960456
36. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019; 47:D607–13.
<https://doi.org/10.1093/nar/gky1131> PMID:30476243
37. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics.* 2010; 26:1572–3.
<https://doi.org/10.1093/bioinformatics/btq170> PMID:20427518
38. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep.* 2017; 18:248–62.
<https://doi.org/10.1016/j.celrep.2016.12.019> PMID:28052254
39. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013; 14:7.
<https://doi.org/10.1186/1471-2105-14-7> PMID:23323831
40. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, Dimitriadoy S, Liu DL, Kantheti HS, Saghaforinia S, Chakravarty D, Daian F, Gao Q, et al, and Cancer Genome Atlas Research Network. Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell.* 2018; 173:321–37.e10.
<https://doi.org/10.1016/j.cell.2018.03.035> PMID:29625050
41. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015; 43:e47.
<https://doi.org/10.1093/nar/gkv007> PMID:25605792
42. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012; 16:284–7.
<https://doi.org/10.1089/omi.2011.0118> PMID:22455463
43. Sotiropoulos C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, Nordgren H, Farmer P, Praz V, Haibe-Kains B, Desmedt C, Lamsimont D, Cardoso F, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst.* 2006; 98:262–72.
<https://doi.org/10.1093/jnci/djj052> PMID:16478745
44. Hazra A, Gogtay N. Biostatistics Series Module 3: Comparing Groups: Numerical Variables. *Indian J Dermatol.* 2016; 61:251–60.
<https://doi.org/10.4103/0019-5154.182416> PMID:27293244
45. Zhu S, Wu Q, Zhang B, Wei H, Li B, Shi W, Fang M, Zhu S, Wang L, Lang Zhou Y, Dong Y. Autophagy-related gene expression classification defines three molecular subtypes with distinct clinical and microenvironment cell infiltration characteristics in colon cancer. *Int Immunopharmacol.* 2020; 87:106757.
<https://doi.org/10.1016/j.intimp.2020.106757> PMID:32769067

SUPPLEMENTARY MATERIALS

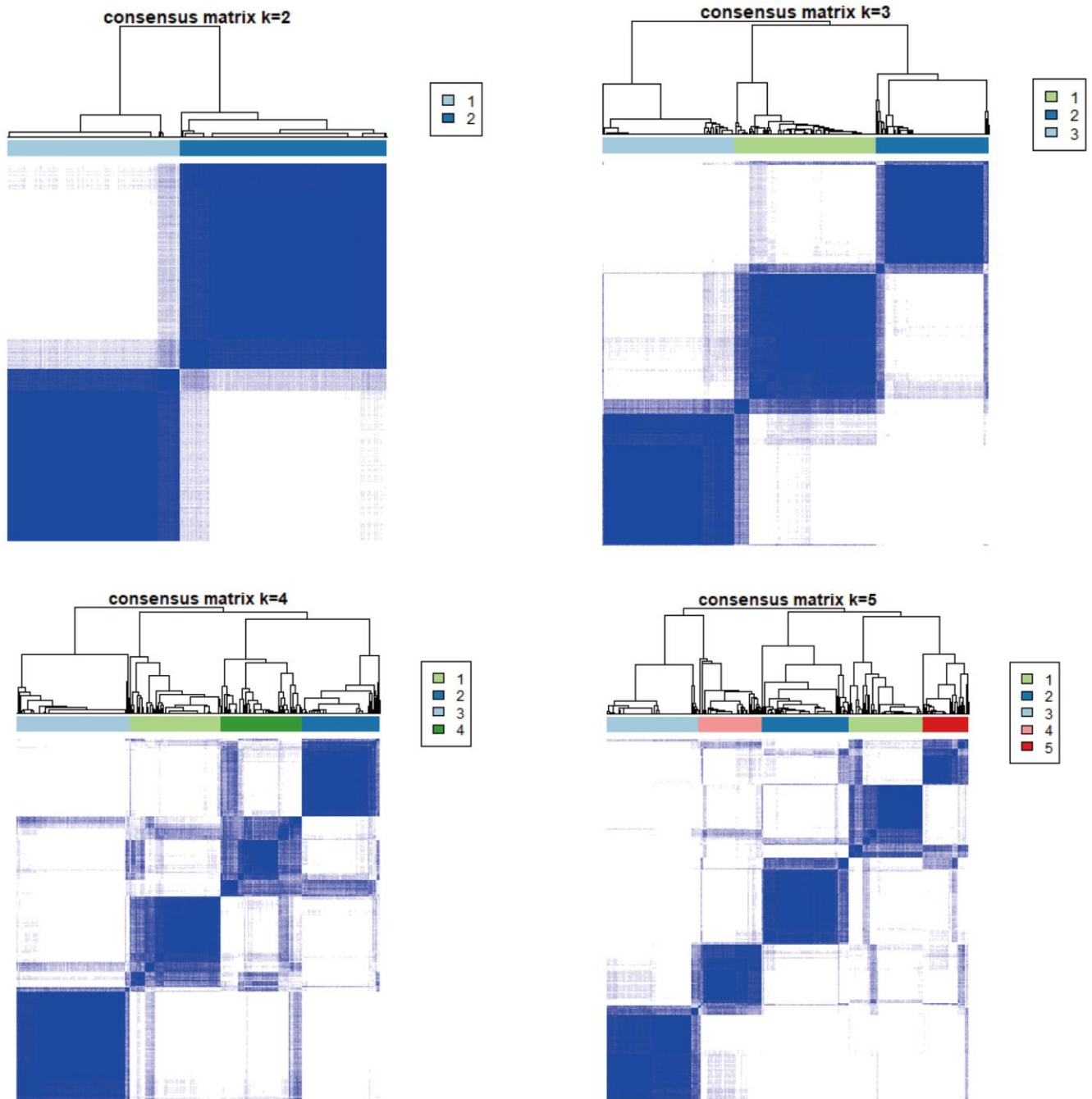
Supplementary Figures



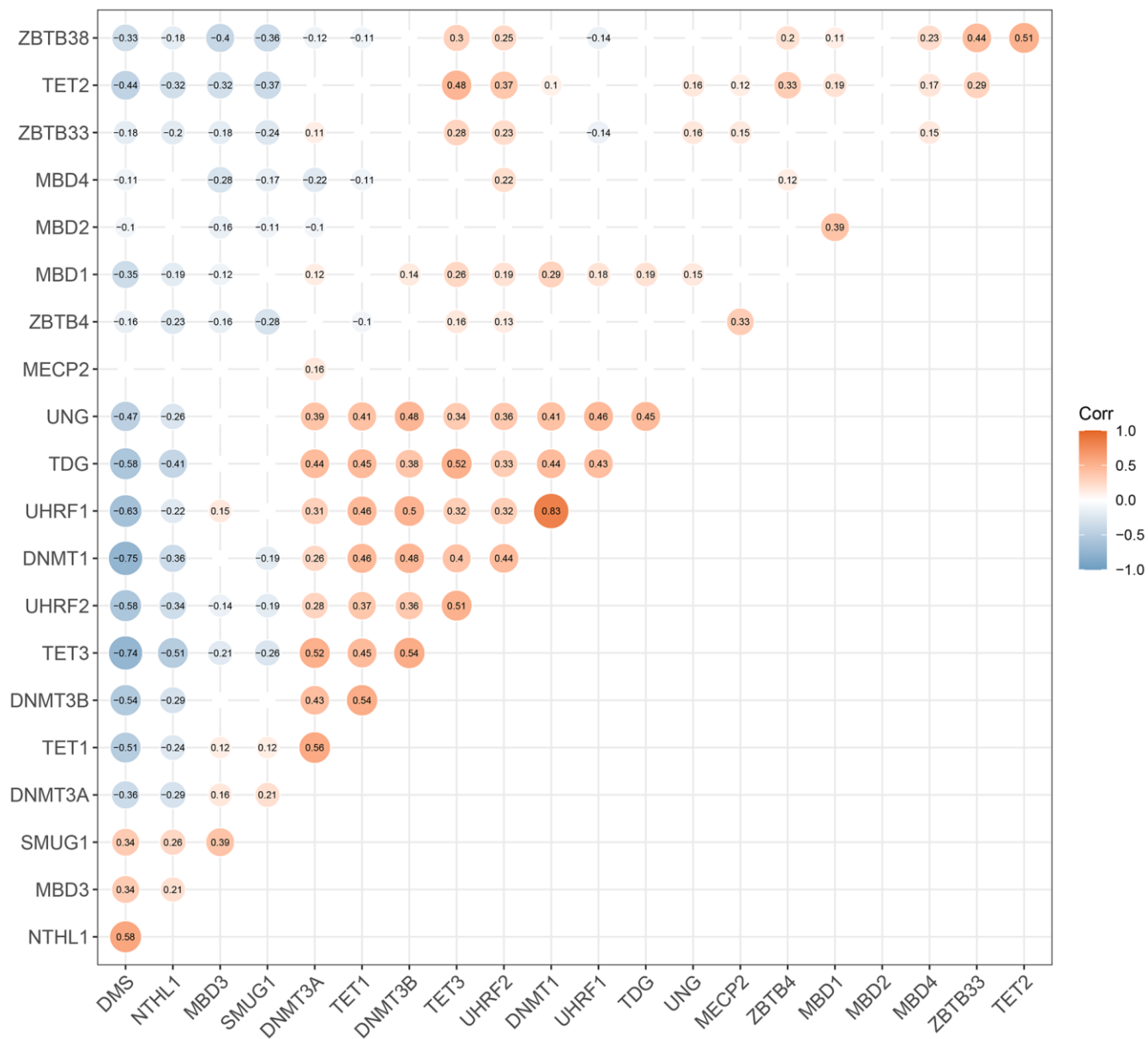
Supplementary Figure 1. Difference in writer gene expression between low and high eraser genes. (A) TET1. (B) TET2. (C) TET3. Data analyses were based on the TCGA-LIHC cohort.



Supplementary Figure 2. Consensus matrices of 20 DNA methylation regulators for k = 2, k=3, k=4 and k=5. Data analyses were based on the TCGA-LIHC cohort.



Supplementary Figure 3. Consensus matrices of 1037 overlapping differentially expressed genes for k = 2, k=3, k=4 and k=5. Data analyses were based on the TCGA-LIHC cohort.



Supplementary Figure 4. Correlation between DMS score and expression of 20 DNA methylation regulators. The numbers in the circles represented the correlation coefficients.

Supplementary Table

Please browse Full Text version to see the data of Supplementary Table 1.

Supplementary Table 1. Prognostic analysis for these 1037 overlapping differentially expressed genes using the univariate Cox regression model.