scientific reports

Check for updates

OPEN

Identification of a CpG-based signature coupled with gene expression as prognostic indicators for melanoma: a preliminary study

Zhen Lin & Liu Yang✉

DNA methylation is an important part of the genomic biology, which recently allowed the identification of key biomarkers for a variety of cancers, including cutaneous melanoma. Despite the current knowledge in cutaneous melanoma, there is a clear need for new efficient biomarkers in clinical application of detection. We use The Cancer Genome Atlas data as a training set and a multi-stage screening strategy to identify prognostic characteristics of melanoma based on DNA methylation. Three DNA methylation CpG sites were identified to be related to the overall survival in the skin cutaneous melanoma cohort. This signature was validated in two independent datasets from Gene Expression Omnibus. The stratified analysis by clinical stage, age, gender, and grade retained the statistical significance. The methylation signature was significantly correlated with immune cells and anti-tumor immune response. Moreover, gene expression corresponding to the candidate CpG locus was also significantly correlated with the survival rate of the patient. About 49% of the prognostic effects of methylation are mediated by affecting the expression of the corresponding genes. The prognostic characteristics of DNA methylation combined with clinical information provide a better prediction value tool for melanoma patients than the clinical information alone. However, more experiments are required to validate these findings. Overall, this signature presents a prospect of novel and wide-ranging applications for appropriate clinical adjuvant trails.

Abbreviations

TCGA	The Cancer Genome Atlas
GEO	Gene Expression Omnibus
SKCM	Skin cutaneous melanoma
TSG	Tumor suppressor genes
UCSC	University of California Santa Cruz
CIBERSORT	Cell-type identification by estimating relative subsets of known RNA transcripts
GZMA	Granzymes A
PRF1	Perforin
GO	Gene ontology
GSEA	Gene Set Enrichment Assay
TIDE	Tumor Immune Dysfunction and Exclusion
SOX9	SRY-Box Transcription Factor 9
MITF	Melanocyte Inducing Transcription Factor
TIM-3	T-Cell Immunoglobulin Mucin Receptor 3
LGALS9	Galectin 9
MDSC	Myeloid derived suppressor cell
CAF	Cancer associated fibroblast
TAM	Tumor-associated macrophage

Cutaneous melanoma is one of the deadliest neoplasms with high aggressive and metastatic characteristics[1]. The favorable prognosis of melanoma is dependent of its timely detection in early stage[2]. Once the melanoma is

Cancer Center, Department of Medical Oncology, Zhejiang Provincial People's Hospital (Affiliated People's Hospital), Hangzhou Medical College, Hangzhou, Zhejiang, China. ✉email: yangliu@hmc.edu.cn

nature portfolio

1

established and developed, it becomes a threat to the patients' life[3]. Therefore, the identification of novel effective prognostic indexes is needed to accelerate the detection of melanoma.

Aberrant DNA methylation is an epigenetic hallmark of cancer, which plays a significant role in cancer management[4]. It is increasingly recognized that epigenetic aberrations could induce cancer development and progression by inactivation of tumor suppressor genes (TSGs) at the promoter region[4]. Regarding melanoma, aberrant DNA methylation is the most widely studied dysregulated epigenetic mechanism[5]. Until now, over 100 genes contributing to melanoma pathogenesis have been identified to be aberrantly hypermethylated[5]. These genes include SOX9, MITF, TIM-3 and LGALS9[6–8], which play critical roles in tumor cell biology after hypermethylation. The methylation patterns might be recognized as a biomarker for predicting the prognosis of patients. Furthermore, it is also reported that alterations of DNA methylation status can be involved in the tumor microenvironmental immune evasion and are able to modulate the response of immunotherapy[9]. Genomic methylation alterations counteract the contribution of high mutation burden and increase immunotherapeutic resistance[10]. Moreover, several epigenetic therapies have been developed inhibiting enzymes controlling epigenetic modifications[4]. For example, 5-azacytidine and 5-aza-2-deoxycytidine were discovered as promising inhibitors of DNA methylation for treating several malignancies[11]. Meanwhile, the use of inhibitors in combination with other key enzymes will greatly improve therapy and patient recovery[11].

In the present study, we aim to construct a novel methylation biomarker signature for melanoma prognosis and further immunotherapy application. To do so, we have investigated the prevalence of CpG island methylation in TSGs silenced in cancer to design methylation profiles associated with main clinicopathological features. Moreover, we have explored the prognostic significance of these TSGs promoter methylations in melanoma outcomes. Comprehensive bioinformatics analyses were conducted to explore potential mechanism of biomarkers. This study presents an accurate predictive pattern for melanoma patients based on methylation and provides prospect of subsequent personalized medical care.

Methods and material
Data acquisition and processing
DNA methylation data (Illumina Infinium HumanMethylation450 array) from patients with skin cutaneous melanoma in the "cohort: GDC The Cancer Genome Atlas (TCGA) skin cutaneous melanoma (SKCM)" as training set were acquired from the Cancer Genomics Browser of The University of California Santa Cruz (UCSC) (https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Melanoma%20(SKCM)&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443). RNA-sequencing data, protein expression data, and clinical information were also retrieved. The clinical data were preprocessed by exclusion of patients with missing follow-up information. Additional methylation data and corresponding clinical information for independent validation set 1 and set 2 were obtained from Gene Expression Omnibus (GEO) (196 patients, GEO accession number GSE144487[12]; 47 patients, GEO accession number: GSE51547[8]). All the data of methylation beta matrix was filtered and normalized with ChAMP R package[13].

Survival model construction process
At first, 459 tumor samples were included in our study by removing the samples with unavailable DNA methylation data, gene expression or survival data. The list of common CpG probes shared between 450K and EPIC array was selected to investigate the association between gene expression and DNA methylation across tumor samples. 342,127 DNA methylation sites were remains to follow-up filtration. Then, Pearson coefficient between gene expression and corresponding methylation β level in the promoter region ("1stExon", "5'UTR", "TSS1500", "TSS200") were calculated. A high negative Pearson coefficient < − 0.6 with $P < 0.05$ was set as criteria for identification of expression related methylation probes. Subsequently, 1288 probes were identified to associated to expression and subjected to the unique Cox regression analysis. 313 probes were retained based on the criteria of $P < 0.001$. At last, the stepwise-cox regression analysis is used to construct a best fitting prognostic model. A risk score formula was established by considering the expression of optimized 3 probes measured by their estimated regression coefficients. Patients were classified into high or low risk groups with the median risk score as cutoff value. The study flowchart describing the process is shown in Fig. 1.

Determination of tumor-infiltrating immune cells and immune responses
Cell-type identification by estimating relative subsets of known RNA transcripts (CIBERSORT), an analytical algorithm, is developed to offer an estimation of the abundances of 22 immune cell population in a mixed neoplasm tissue[14]. ESTIMATE predicts tumor purity, and the presence of infiltrating stromal/immune cells in tumor tissues[15]. The cytolytic activity (CYT) score was obtained by calculating the geometric mean of granzymes A (GZMA) and perforin (PRF1) mRNA expression levels in tissue[16]. RNAseq data was analyzed by edgeR algorithm and Gene Ontology (GO), KEGG pathway enrichment and Gene Set Enrichment Assay (GSEA) that were carried by clusterProfiler package.

Tumor immune dysfunction and exclusion (TIDE) analysis
TIDE is a computational algorithm constructed by Jiang et al.[17] to predict the immunotherapy response of immune checkpoint blockades. TIDE quantified the levels of two conventional mechanisms in the process of tumor immune evasion: T cell dysfunction and T cell exclusion. The TIDE score in patients with cutaneous melanoma from the TCGA cohort, including T cell dysfunction score and T cell exclusion score, were performed in the TIDE web (http://tide.dfci.harvard.edu) after uploading the transcriptome profiles. Then, the correlation among diverse immune parameters and signatures were evaluated.
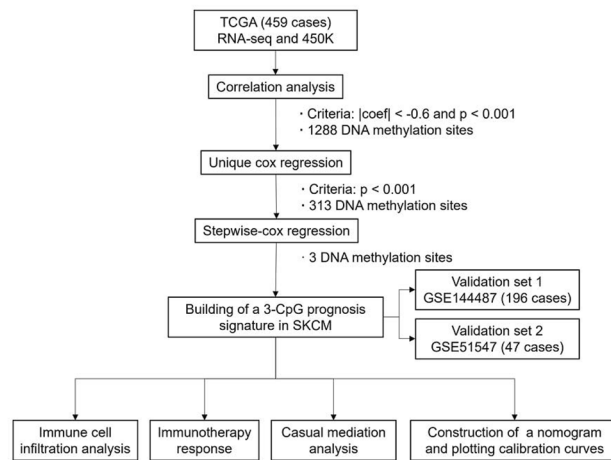
Figure 1. Flowchart showing steps involved in identification of the prognostic DNA methylation signature in melanoma.

Mediation analysis

VanderWeele's mediation analysis is widely applied to test a hypothetical causal chain[18]. In this case, the corresponding mRNA expression was regarded as mediators, and its indirect effect ($HR_{indirect}$) was determined, standing for the regulation via mRNA expression. Meanwhile, the effect of methylation status on clinical prognosis is described as direct effect (HR_{direct}). Finally, the whole effect of methylation on survival is called HR_{total}, composed of the two variables, $HR_{indirect}$ and HR_{direct}.

Nomogram construction

Nomogram was developed to predict the personalized estimation of the probability of prognosis, recurrence, or drug response[19]. In nomogram, each contributing parameter including age, gender, Tumor Node Metastasis stage (TNM stage), and risk score, was quantified by points and added up to generate a total point by individual. In the present study, we integrated the clinical characteristics and risk score in the nomogram for prediction of prognosis for each melanoma patients. The nomogram was processed via the rms package for R software. A calibration curve was applied to visualize the deviation between predicted probabilities and the actual situation, with 1000 replicates set up in the bootstrap method. The predictive accuracy of the nomogram was measured by the concordance index (C-index).

Statistical analysis

Kaplan–Meier survival curves were drawn and compared among subgroups using log-rank tests to assess survival differences. Multivariate Cox regression and subgroups stratification analysis were performed to explore whether the methylation-based risk score was independent of patients' clinical features. To evaluate the efficiency of the survival prediction among the risk score and the TNM stage, receiver operating characteristic (ROC) analysis was performed. Hazard ratios (HR) and 95% C-index were also calculated. P values were two-sided, and $P < 0.05$ was considered statistically significant. Statistical tests above were performed using R version 4.0.0 (The R Foundation).

Ethics approval and consent to participate

Consent for participation for all patients was obtained through The Cancer Genome Atlas Project.

Result

Derivation of prognostic DNA methylation sites and construction of risk formula

The study was conducted on 459 patients in TCGA cohort who are clinically and pathologically diagnosed with cutaneous melanoma. Firstly, univariate cox proportional hazard regression analysis was carried to filter the candidate mRNA-related-DNA methylation level data in the training cohort. With the strict criteria ($P < 0.001$), a total of 313 DNA methylation sites were identified as candidate markers that significantly correlated with the overall survival (OS) of patients. Subsequently, these candidate sites were used to perform multivariate cox regression analysis with both-side stepwise. Finally, 3 methylation sites were selected to contribute one hazard ratio model as the optimum prognostic model for predicting OS. The liner risk score formula was created as follows: Risk score $= (-1.29) * \beta_{cg09088834(NINL)} + (0.7904946) * \beta_{cg16393012(ARHGDIB)} + (0.7137228) * \beta_{cg09321817(HLA-DPA1)}$. The detailed information of the three methylation probes is shown in Supplementary Table 1. The risk score for each patient was calculated using this formula. Then, patients were assigned into high-risk and low risk group based on the optimized risk value within the TCGA cohort (Fig. 2A). The survival of high-risk patients was significantly shorter than of low-risk patients in a Kaplan–Meier curve (Fig. 2B, $P < 0.0001$). The univariates Cox regression analysis was operated to evaluate the prognosis value of these sites respectively (Fig. 2C). Then, we calculated
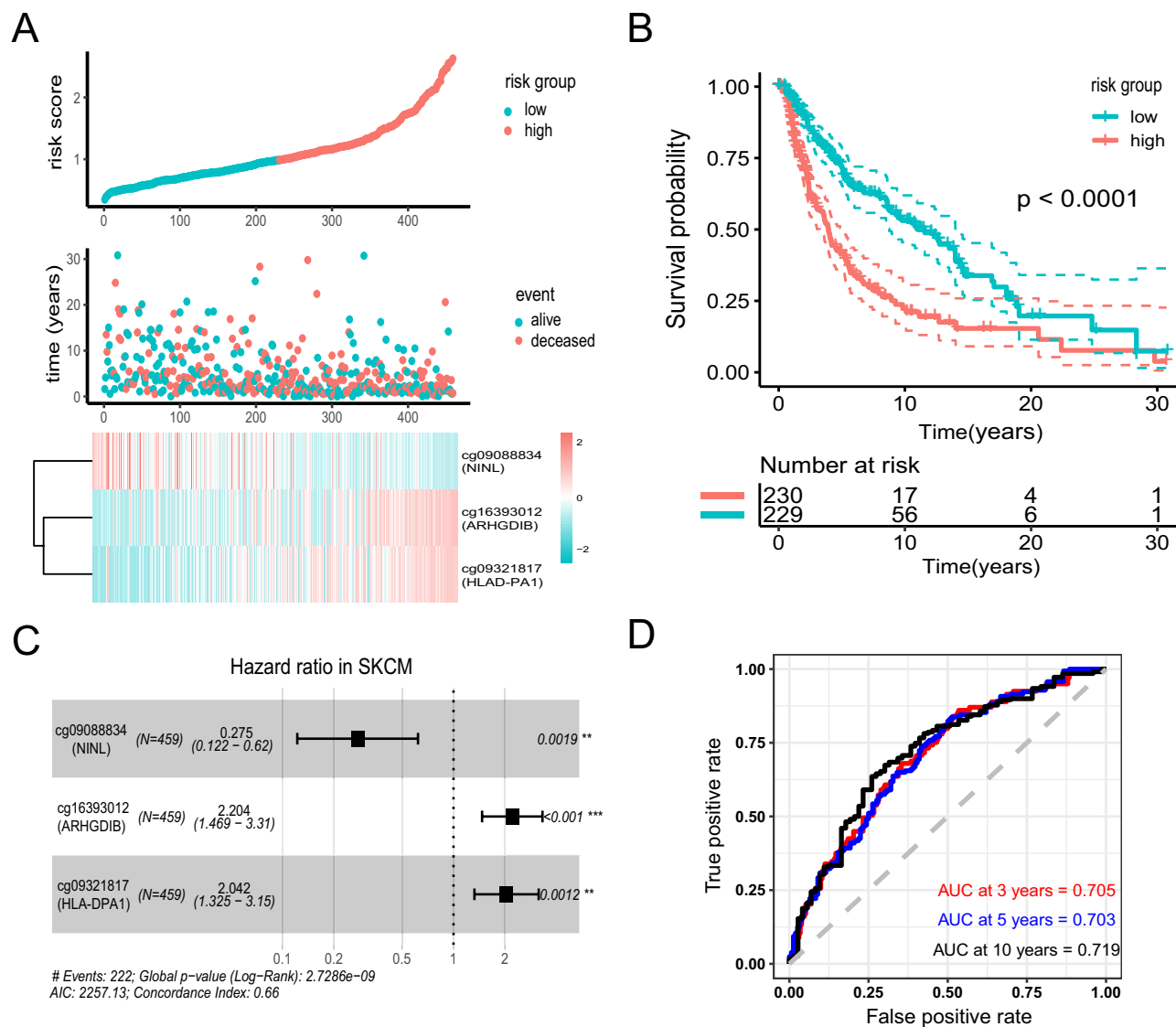
Figure 2. Construction of the 3-methylation signature in TCGA training set for determining survival outcomes. (**A**) The distribution of risk score, patients' survival status and methylation expression panel. (**B**) Kaplan–Meier curves in all patients based on risk score. (**C**) 3 methylation probes were significantly correlated with overall survival derived from the univariable cox regression analysis in melanoma patients. (**D**) ROC analysis of 3 methylation for predicting of survival at 3, 5 and 10 years in TCGA cohorts.

the time-dependent area under the ROC curves (AUC) to evaluate predictive performance. AUC values at 3-, 5- and 10-years indexes are 0.705, 0.703, and 0.719 respectively, representing a prognostic prediction (Fig. 2D).

The signature is independent from clinical and pathological characteristics

There are varieties of clinical and pathological features that are considered as predictions for patients with melanoma, such as patients' age, sex, AJCC stage, tumor thickness, the site where the sample was obtained, ulceration status, as well as mutation status[20–24]. Therefore, patients were regrouped according to different clinicopathological characteristics to assess the independence and applicability of this three-DNA methylation signature. Age, gender, tumor tissue site, pathologic stage, Breslow thickness, ulceration, BRAF and NRAS mutation status were taken into consideration. Consequently, the three-DNA methylation signature remains a good reference for different regrouped cohorts in the forest plot (Fig. 3). Meanwhile, patients with BRAF mutation shows a high-risk score comparing with counterparts without BRAF mutation. However, there is no difference about NRAS status (Supplementary Fig. 1). These results suggest that the methylation signature has a high effectiveness for risk stratification. The signature could therefore be used as an applicable independent prognostic predictor and guide clinicals to choose proper targeted drugs.

Subgroup	Number		P value	Hazard Ratio(95%CI)
All patients	459	◆	6.18E−11	2.156(1.713−2.715)
Gender				
Male	286	■	1.87E−06	1.909(1.464−2.491)
Female	173	⊢■⊣	3.14E−06	3.297(1.996−5.444)
Age of diagnosis				
<60 y	249	■	4.88E−05	1.905(1.396−2.599)
>60 y	210	⊢■⊣	8.41E−07	2.562(1.762−3.725)
Tumor tissue site				
Primary Tumor	103	■	0.049	1.670(1.010−2.330)
Regional Cutaneous or Subcutaneous Tissue	73	⊢■⊣	0.0101	2.319(1.222−4.401)
Regional Lymph Node	216	⊢■⊣	3.05E−08	2.859(1.972−4.147)
Distant Metastasis	64	■	0.0441	1.605(1.013−2.544)
Pathologic stage				
0, I and II	230	⊢■⊣	3.04E−06	2.293(1.618−3.248)
III and IV	193	⊢■⊣	9.12E−08	2.896(1.961−4.277)
Breslow thickness				
<2 mm	135	⊢■⊣	1.60E−04	2.717(1.617−4.566)
>2 mm	220	■	7.70E−04	1.762(1.267−2.451)
Ulceration				
Present	166	⊢■⊣	3.19E−05	2.508(1.626−3.867)
Absent	145	■	0.00319	1.793(1.216−2.644)
BRAF mutation				
Yes	233	⊢■⊣	1.15E−06	3.258(2.024−5.244)
No	221	■	5.68E−05	1.832(1.364−2.461)
NRAS mutation				
Yes	126	⊢■⊣	2.41E−04	2.115(1.418−3.154)
No	328	■	2.18E−07	2.163(1.616−2.896)

1 3 5

Figure 3. The prognostic value in subgroups based on clinicopathologic characteristics in TCGA cohort. The SKCM cohort was divided into subgroups based on clinicopathologic characteristics, such as gender, age of diagnosis, tumor tissue site, pathologic stage, Breslow thickness, ulceration, BRAF mutation and NRAS mutation.

Evaluation of the methylation signature for survival prediction in other independent cohorts

To further examine the prognostic values of the three-DNA methylation signature in another independent cohorts, Kaplan–Meier and ROC analyses were carried out in two other independent cohorts (GSE144487 N = 198; GSE51547, N = 47). Similarly, patients with high or low risk were grouped based on the median risk score of the training cohort (Fig. 4A, D). Patients in the low-risk group had a significantly longer overall survival in GSE144487, suggesting the three-DNA methylation performed well in prediction ($P < 0.0001$, Fig. 4B). 3-, 5-, and 10-year AUC were 0.678, 0.697 and 0.713, respectively (Fig. 4C). However, due to the limited sample size in the GSE51547 cohort, the log-Rank test was not significant but the AUC value in 3- and 5-year were 0.705 and 0.833 (Fig. 4E, F), confirming that the three-DNA methylation signature can also predict the survival of melanoma patients in other independent cohorts.

Association between three-DNA methylation signature with immune cells and immune response

Immune cells calculated by CIBERSORT algorithm were compared in the two groups in the SKCM cohort. Diverse anti-tumor immune cells, including plasma B cells, CD8+ T cells, activated memory CD4+ T cells, activated natural killer (NK) cells, M1-like macrophages and activated dendritic cells (DCs) were found highly enriched in the low-risk population. Meanwhile, resting memory CD4+ T cells, resting NK cells, M0 macrophages, M2-like macrophages and resting mast cells were elevated in the high-risk population (Fig. 5A). Considering the role of TILs as an important marker, we analyze the prognostic value with this parameter. This
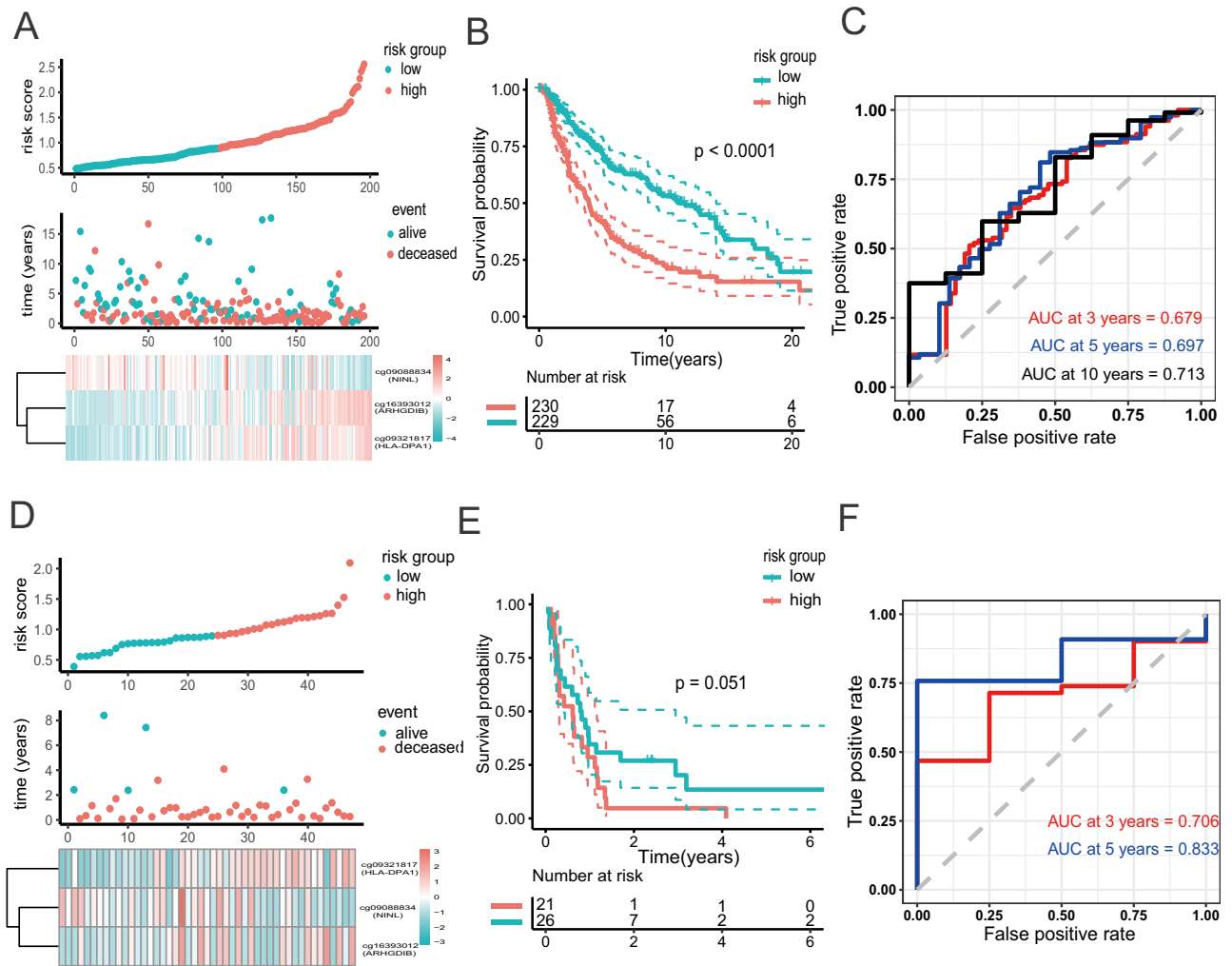
Figure 4. Validation of 3-methymation-signature by revealing the prognostic significance in the validation sets. (**A–C**). The distribution of risk score, patients' survival status and methylation expression panel; Kaplan–Meier curves and ROC analysis for survival prediction in GSE144487. (**D–F**). The distribution of risk score, patients' survival status and methylation expression panel; Kaplan–Meier curves and ROC analysis for survival prediction in GSE51547.

signature may serve as a reliable independent prognostic predictor even when considering immune cells as covariates (Supplementary Fig. 2). Immune score calculated by ESTIMATE algorithm was higher in the risk-low group (Fig. 5B). The cytolytic activity (CYT) score is a new index of cancer immunity calculated from the mRNA expression levels of GZMA and PRF1, representing immune cytotoxicity[16]. CYT score was higher in risk-low group (Fig. 5C). Moreover, GSEA analysis GSEA analysis was performed by comparing the risk-high and risk-low group, uncovering that T cell receptor (TCR), CD8[+] TCR downstream and B cell receptor pathways were significantly downregulated in the high-risk group (NES = − 0.24, p = 2.0e−6; NES = − 0.24, p = 2.0e−6; NES = − 0.24, p = 2.0e−6; Fig. 5D–F). Gene Ontology (GO) enrichment analysis showed that immune response related pathways, including T cell receptor signaling pathway, CD8 TCR downstream pathway, and T cell receptor signaling pathway, were also highly enriched in the low-risk group (Fig. 5G). Finally, in Reverse Phase Protein Array (RPPA) of SKCM cohorts, the protein levels of LCK and CD20 were higher in low-risk group, suggesting a boost of the anti-tumor immune response (Fig. 5H, I).

Relationship between the methylation-based signature and immunotherapy response

Next, in order to predict the response of immunotherapy, we introduced the TIDE analysis into our study[17]. As expected, patients in the high-risk group displayed a significantly lower TIDE score, as well as T cell dysfunction (Fig. 6A). However, the high-risk population is characterized by higher exclusion scores (Fig. 6A). It is also widely accepted that microsatellite instability could be a predictor for immunotherapy efficiency[25]. MSI signature, a prediction model based on ridge regression, was higher in the low-risk group as compared to the high-risk group (Fig. 6B). Merk18 expression is considered as a T-cell inflamed signature for pan-cancer predictors of clinical benefit from anti-programmed cell death Protein 1 (PD-1) treatment[26]. Whereas IFN-γ-related mRNA
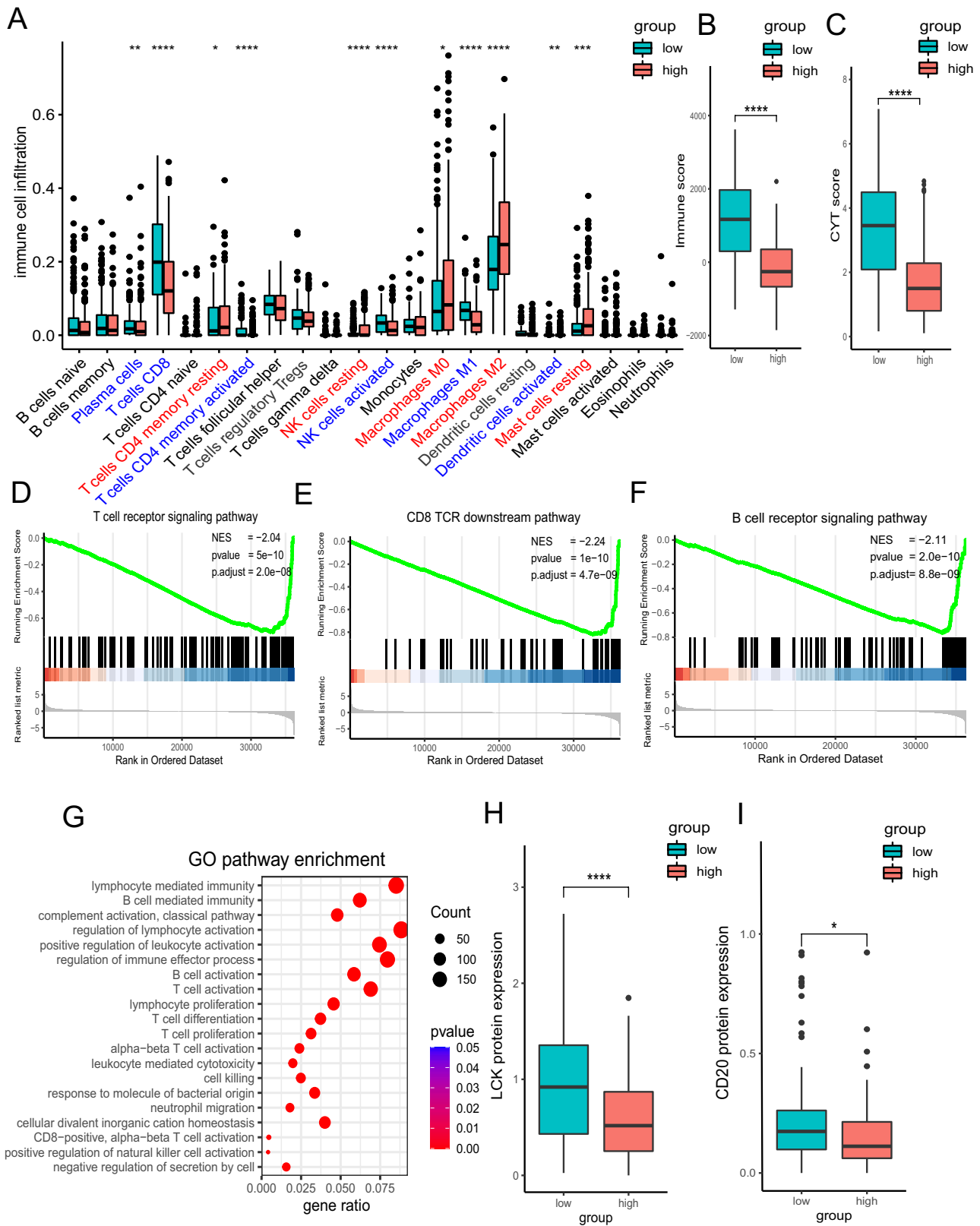
Figure 5. Analysis of composition of tumor infiltering leukocytes and immune response. (**A**) The comparison of 22 immune cell subsets between high score and low score groups based on CIBERSORT. (**B**) Estimated immune score comparison based on ESTIMATE. (**C**) CYT score comparison. (**D**–**F**). GSEA analysis of T cell receptor signaling pathway, CD8 TCR downstream pathway, and T cell receptor signaling pathway. (**G**) GO enrichment analysis. (**H**, **I**) Protein levels of LCK and CD20.

Figure 6. Distribution of immunotherapy response markers in high- and low-risk groups. (**A**) The distribution of TIDE score, T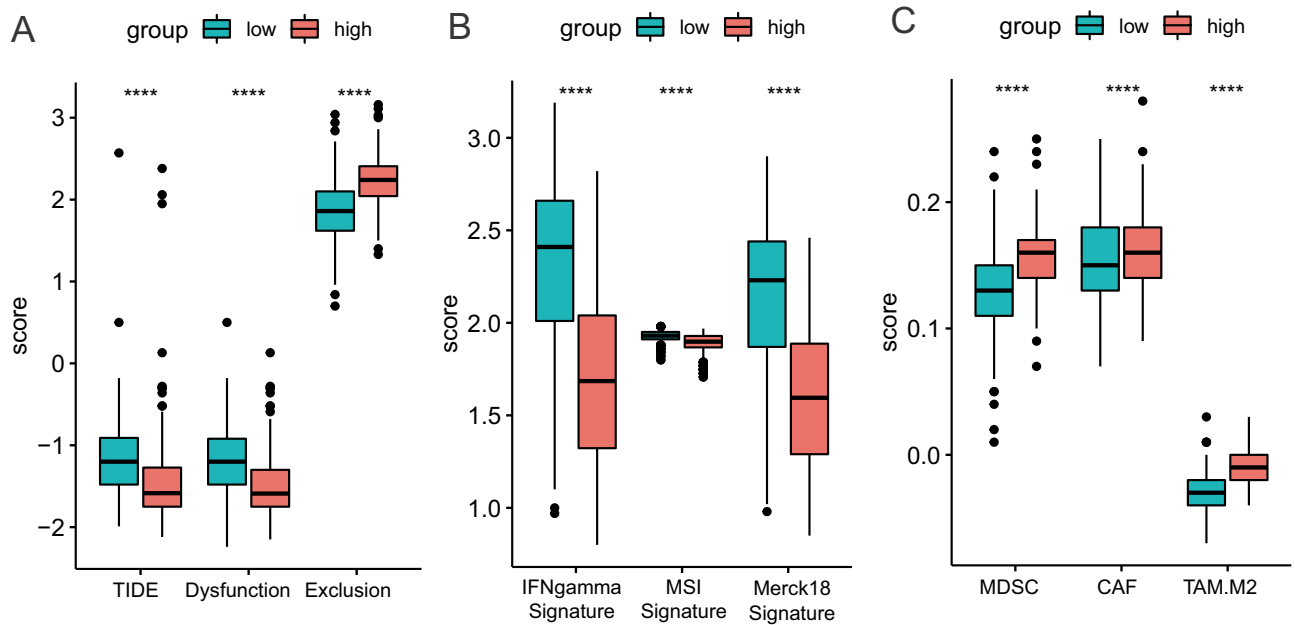 cell dysfunction score, and T cell exclusion score in high- and low-risk groups. (**B**) The distribution of MSI signature, Merk18 signature, and IFN signature in high- and low-risk groups. (**C**) The distribution of MDSC, CAF and TAM in high- and low-risk groups. *, **, *** and **** represent $P < 0.05$, $P < 0.01$, $P < 0.001$ and $P < 0.0001$, respectively.

profile predicts clinical response to PD-1 blockade[27]. Both Merk18 and IFN-γ-signature were elevated in the low-risk group (Fig. 6B). Finally, we investigated myeloid-derived suppressor cells (MDSC), cancer associated fibroblasts (CAF) and tumor-associated macrophages (TAM) in the tumor suppressive immune microenvironment. All of these are well-known biomarkers to predict the response of immunotherapy. We found higher MDSC, CAF and TAM levels in high-risk patients (Fig. 6C). These results suggest that the CpG-based signature could highlight a population with low-risk patients who may be more responsive to immunotherapy, especially immune checkpoint blockage (ICB).

The underlying mediation of mRNA expression in the effect of methylation on overall survival

To illustrate the underlying mediation pathway between methylation, and overall survival, the VanderWeele's mediation analysis was performed. In this overall mediation model, the mediator was risk score calculated by linear combination including three genes' mRNA expression (Fig. 7A). There is a clear relation between the methylation status and the corresponding mRNA expression level for the three CpG sites (Supplementary Fig. 3). Furthermore, these 3 genes were also associated with patients' outcomes (Supplementary Fig. 4). Subsequently, prognostic equation based on the mRNA expression was also performed as described above: $\text{score}_{\text{expression}} = (0.10) * \text{NINL} + (-0.05) * \text{ARHGDIB} + (-0.14) * \text{HLADPA1}$. Interestingly, this score was significantly associated with the prognosis. In the well-constructed mediation model, $\text{HR}_{\text{indirect}}$ is 2.15 (95% CI: 1.71–2.72; $P < 0.0001$). The methylation signature was highly mediated via affecting the respective mRNA expression to predict the survival outcomes of patients (proportion mediated, 49%; Fig. 7B). To avoid bias in the findings, sensitivity analysis was performed by excluding each gene expression from $\text{score}_{\text{expression}}$. We were able to find that the consequence of the mediation role of mRNA expression remained statistically significant (Fig. 7B).

Construction of a predictive nomogram

To develop a clinically applicable method predicting an individual's outcome probability, we used a nomogram to build a predictive model, taking into consideration clinicopathologic covariates. Considering the basis of the multivariate analysis of overall survival, we generated a nomogram to predict the 3-, 5-, and 10-year OS in the training cohort (Fig. 8A). The predictors included age, gender and TNM stage. The calibration plots for the 3-, 5-, and 10-year OS rate were successfully predicted (Fig. 8B–D). Combination of clinical information, methylation data (AUC = 0.773, 0.725, and 0.77 for 3-, 5- and 10-year) showed a superior prediction ability in comparison to the model using clinical data only (AUC = 0.691, 0.613, and 0.673 for 3-, 5- and 10-year; Fig. 8E–G).

Discussion

In recent years, the importance of DNA methylation in the biology of cutaneous melanoma has been increasingly appreciated[5]. For instance, it is well described that tumor suppressor gene promoters display the focal DNA hypermethylation in many cancers, including melanoma[5]. Furthermore, methylation features could be used to classify distinct subgroups with differing survival outcomes and biologic behavior within melanoma patient cohorts[28]. Previous studies discovered that DNA methylation regulates the expression of key genes such as
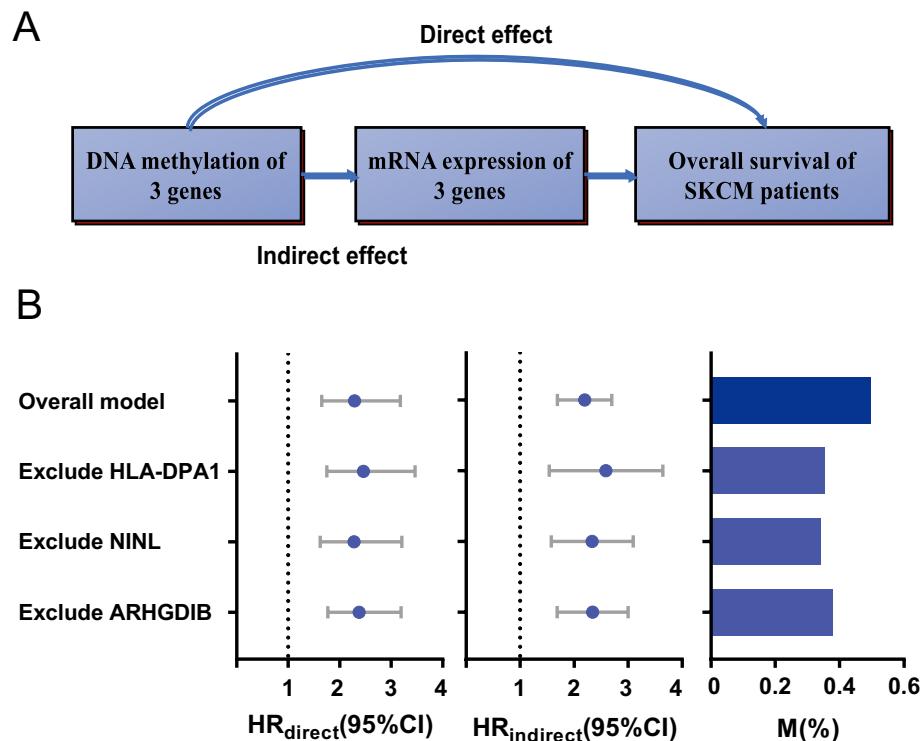
Figure 7. The mediation analysis for methylation prognostic signature through mRNA expression. (**A**) Diagram of mediation model. (**B**) The 3 CpG sites were treated as "exposure"; mediator was the linear combination of the corresponding 3 mRNA expression level (score$_{exp}$) in overall model. Prognostic effect in hazard ratio (HR) were described as direct effect (HR$_{direct}$) and indirect effect (HR$_{indirect}$), corresponding 95% confidence interval (95% CI), and the proportion of effect mediated (M%). Further, sensitivity analyses were performed by excluding each gene from score$_{exp}$, respectively.

tyrosinase (TYR), tyrosinase-related protein 1 (TYRP1), dopachrome tautomerase (DCT) and microphthalmia-associated transcription factor (MITF), as well as paracrine factors such as stem cell factor (SCF) and endothe-lin-1 (ET-1) in melanogenesis, a very important process in both cutaneous and ocular melanoma biology[29–32]. Previous study have revealed that a "Yin and Yang" role for melanin and active melanogenesis in melanoma development, progression, and therapy[33]. However, these studies usually concentrated on single gene methyla-tion, which were unable to achieve a good performance of prediction. Moreover, their application was limited to specific clinical characteristics due to small patient numbers. Consequently, compared with individual DNA methylation, a combination of DNA methylation as biomarkers could achieve a better sensitivity and specificity as predictive pattern[34].

This signature is confirmed by its high reproducibility and utility in various clinical groups with potential clinical applicability. Cutaneous melanoma has a high heterogeneity in terms of genome, clinical features, and histopathological characteristics[35]. Several parameters, including age, gender, stage of disease, Breslow thickness, and ulceration status, have significant influence on melanoma patient prognosis, relapse, and therapy responses[23]. For instance, the methylation signature could present reliable independence of the clinical factors mentioned above in regard to utilization. BRAF and NRAS are mutant genes in skin cutaneous melanoma associated with different tumor immune microenvironments[24]. Thus, we analyzed the performance of the signature among patients with different BRAF and NRAS mutation status. We could demonstrate that our signature is independ-ent of these mutations and could be applicable as a prognostic biomarker. Meanwhile, considering that an ideal prognostic marker could efficiently stratify in other independent cohorts, we employed two other GEO datasets (GSE14487 and GSE51547) to validate the practicality of our three-DNA methylation signature. The 3-DNA methylation signature performed well in distinguishing low-risk and high-risk groups in these two independ-ent cohorts as well, suggesting that this signature may be of high clinical value with wide-ranging applications.

In the process of anti-tumor response, immune cells infiltrate into tumor tissue. The tumor cells work as antigens and activate the host immunity system[36]. The grade of tumor-infiltrating lymphocytes is an independent protective factor, which displays an excellent prognostic marker in melanoma[37]. Meanwhile, immune checkpoints targeted therapy has facilitated great progress in treating multiple cancers. It is well known that immune inhibi-tors including PD-1, programmed death-ligand 1 (PD-L1), and cytotoxic T-lymphocyte-associated Protein 4 (CTLA-4) enable patients to produce an effective antitumor response, especially in melanoma[38]. However, there is a limitation that only one-third of patients show clinical benefit from immunotherapy[39]. Given the present situation, methylation status in DNA promoters play a critical role in cell lineage specification. Therefore, we hypothesize that DNA methylation signature may serve as a specific molecular predictor for the evaluation of
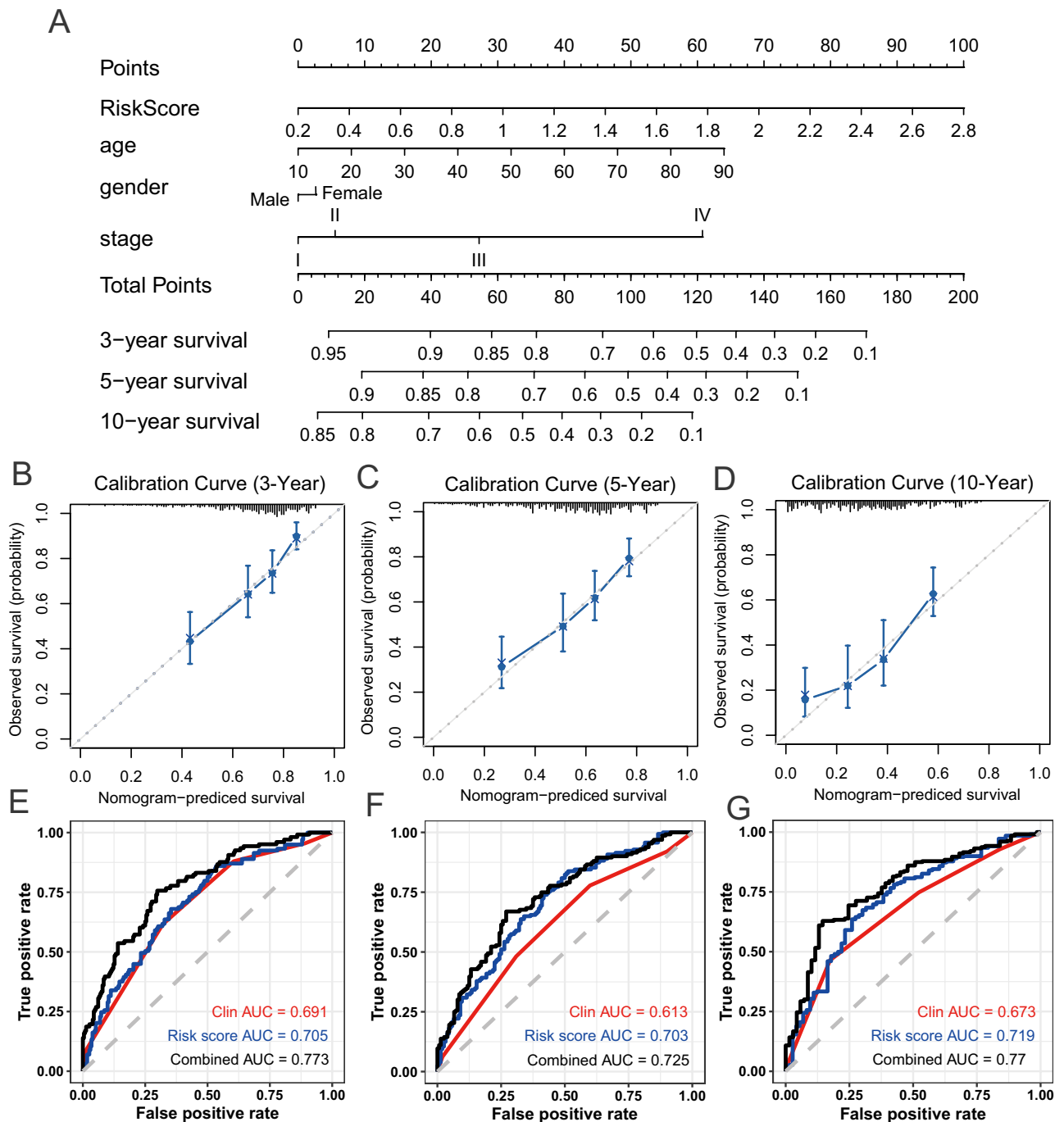
Figure 8. Nomogram and Calibration plot. (**A**) Nomogram to predict the 3-, 5- and 10-year OS. (**B–D**) Calibration curve for OS nomogram model in 3-, 5- and 10-year. The dot line represents the ideal nomogram, and the blue line represents the observed nomogram. (**E–G**) ROC analysis of the sensitivity and specificity of OS prediction by the DNA methylation signature risk score, clinical information, and combination of the two factors.

immune activity[40]. To preliminarily assess the predictive ability of the methylation-based signature, six different well-validated immunotherapy biomarkers were used. The TIDE score was created to serve as an accurate biomarker for the immune checkpoint blockade response[17]. Microsatellite instability has been reported as a predictive factor for immunotherapy in malignant melanoma[25]. Merk18 expression represents a T cell inflamed signature that predicts therapeutic efficacy in patients treated with pembrolizumab across 20 cancers[26]. IFN-γ signature predicts clinical response to PD-1 blockade[27]. MDSCs, CAFs and TAMs are well known classic biomarkers of the response to anti-PD-1/PD-L1 therapies[41–43]. We found that high-risk patients had significantly lower levels of TIDE score, MSI, Merk18, IFNγ-signature, MDSCs, CAFs and TAMs. In the current study, these novel findings represent that our methylation-based signature, although developed for accurate prognosis, also

significantly correlates with immune cell infiltration and anti-tumor immune response. However, further studies are needed to verify the ability of this signature to predict immunotherapy response.

In addition to DNA methylation, mRNA expression levels of three genes also affected prognosis significantly. NINL is identified as an oncogenic protein which causes spontaneous tumorigenesis in transgenic mice[44]. ARHGDIB is a metastasis suppressor gene affecting the migration of T cells in varieties of cancers[45,46]. HLA-DPA1 is mostly expressed in antigen presenting cells and plays a central role in the immune system by presenting peptides derived from extracellular proteins[47]. Mediation analysis showed that around 49% of the methylation prognostic effect is mediated through affecting corresponding gene expression. Meanwhile, the mediational role of transcriptional regulation is confirmed by sensitivity analysis.

At last, we built a nomogram including methylation score and clinical pathology to predict individual prognosis. Prognostic signature integrating DNA methylation and clinical information provides a better prognostic prediction value for melanoma patients than clinical information only. Our nomogram provides a simple and accurate prognostic tool for OS indication for patients with melanoma.

The study's limitations should be noted. Firstly, the biologic mechanisms of the candidate markers NINL, ARHGDIB, and HLADPA1 are still unknown. Secondly, although this study was validated using two independent cohorts, clinical investigations in different populations are needed to validate the prognostic value of this signature. Last but not least, wet-lab experiments are required to further prove the correlation with the immune cells and the anti-tumor immune response.

Conclusion

In this study, we constructed a powerful DNA methylation signature with a high predictive efficiency for prognosis in melanoma. The identified signature is independent from clinical and pathological characteristics and had a good performance in multiple cohorts. In the future, we hope that this signature could be used to refine the current prognostic model and to optimize the treatment strategy for patients with melanoma.

Data availability

Clinical information, high-throughput sequencing-counts, and DNA methylation data were retrieved from the UCSC xena (https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Melanoma%20(SKCM)&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443), which is a publicly available database.

References

1. Tracey, E. H. & Vij, A. Updates in melanoma. *Dermatol. Clin.* **37**(1), 73–82. https://doi.org/10.1016/j.det.2018.08.003 (2019).
2. Piris, A., Lobo, A. C. & Duncan, L. M. Melanoma staging: Where are we now?. *Dermatol. Clin.* **30**(4), 581–592. https://doi.org/10.1016/j.det.2012.06.001 (2012).
3. Schadendorf, D. *et al.* Melanoma. *Lancet* **392**(10151), 971–984. https://doi.org/10.1016/S0140-6736(18)31559-9 (2018).
4. Egger, G., Liang, G., Aparicio, A. & Jones, P. A. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **429**(6990), 457–463. https://doi.org/10.1038/nature02625 (2004).
5. Micevic, G., Theodosakis, N. & Bosenberg, M. Aberrant Dna methylation in melanoma: Biomarker and therapeutic opportunities. *Clin. Epigenetics* **9**, 34. https://doi.org/10.1186/s13148-017-0332-8 (2017).
6. Holderried, T. *et al.* Molecular and immune correlates of TIM-3 (HAVCR2) and galectin 9 (LGALS9) mRNA expression and DNA methylation in melanoma. *Clin. Epigenetics* **11**(1), 161. https://doi.org/10.1186/s13148-019-0752-8 (2019).
7. Cheng, P. F. *et al.* Methylation-dependent Sox9 expression mediates invasion in human melanoma cells and is a negative prognostic factor in advanced melanoma. *Genome Biol.* **16**, 42. https://doi.org/10.1186/s13059-015-0594-4 (2015).
8. Lauss, M. *et al.* Genome-wide DNA methylation analysis in melanoma reveals the importance of CpG methylation in MITF regulation. *J. Investig. Dermatol.* **135**(7), 1820–1828. https://doi.org/10.1038/jid.2015.61 (2015).
9. Jung, H. *et al.* DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load. *Nat. Commun.* **10**(1), 4278. https://doi.org/10.1038/s41467-019-12159-9 (2019).
10. Vizoso, M. *et al.* Epigenetic activation of a cryptic TBC1D16 transcript enhances melanoma progression by targeting EGFR. *Nat. Med.* **21**(7), 741–750. https://doi.org/10.1038/nm.3863 (2015).
11. Zhang, J., Yang, C., Wu, C., Cui, W., Wang, L. DNA methyltransferases in cancer: biology, paradox, aberrations, and targeted therapy. *Cancers.* https://doi.org/10.3390/cancers12082123 (2020).
12. Mitra, S. *et al.* Analysis of DNA methylation patterns in the tumor immune microenvironment of metastatic melanoma. *Mol. Oncol.* https://doi.org/10.1002/1878-0261.12663 (2020).
13. Tian, Y. *et al.* Champ: Updated methylation analysis pipeline for illumina beadchips. *Bioinformatics* **33**(24), 3982–3984. https://doi.org/10.1093/bioinformatics/btx513 (2017).
14. Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M. & Alizadeh, A. A. Profiling tumor infiltrating immune cells with cibersort. *Methods Mol. Biol.* **1711**, 243–259. https://doi.org/10.1007/978-1-4939-7493-1_12 (2018).
15. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612. https://doi.org/10.1038/ncomms3612 (2013).
16. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**(1–2), 48–61. https://doi.org/10.1016/j.cell.2014.12.033 (2015).
17. Jiang, P. *et al.* Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**(10), 1550–1558. https://doi.org/10.1038/s41591-018-0136-1 (2018).
18. VanderWeele, T. J. Explanation in causal inference: Developments in mediation and interaction. *Int. J. Epidemiol.* **45**(6), 1904–1908. https://doi.org/10.1093/ije/dyw277 (2016).
19. Iasonos, A., Schrag, D., Raj, G. V. & Panageas, K. S. How to build and interpret a nomogram for cancer prognosis. *J. Clin. Oncol.* **26**(8), 1364–1370. https://doi.org/10.1200/JCO.2007.12.9791 (2008).
20. Weiss, S. A., Hanniford, D., Hernando, E. & Osman, I. Revisiting determinants of prognosis in cutaneous melanoma. *Cancer-Am. Cancer Soc.* **121**(23), 4108–4123. https://doi.org/10.1002/cncr.29634 (2015).
21. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2021. *CA-Cancer J. Clin.* **71**(1), 7–33. https://doi.org/10.3322/caac.21654 (2021).

22. Wouters, J. *et al.* Comprehensive DNA methylation study identifies novel progression-related and prognostic markers for cutaneous melanoma. *BMC Med.* **15**(1), 101. https://doi.org/10.1186/s12916-017-0851-3 (2017).
23. Messeguer, F. *et al.* Risk factors for the development of locoregional cutaneous metastases as the sole form of recurrence in patients with melanoma. *Actas Dermo-Sifilogr.* **104**(1), 53–60. https://doi.org/10.1016/j.ad.2012.05.003 (2013).
24. Akbani, R., *et al.* Genomic classification of cutaneous melanoma. *Cell* **161**(7), 1681–1696. https://doi.org/10.1016/j.cell.2015.05.044 (2015).
25. Kubecek, O., Trojanova, P., Molnarova, V. & Kopecky, J. Microsatellite instability as a predictive factor for immunotherapy in malignant melanoma. *Med. Hypotheses* **93**, 74–76. https://doi.org/10.1016/j.mehy.2016.05.023 (2016).
26. Ott, P. A. *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.* **37**(4), 318–327. https://doi.org/10.1200/JCO.2018.78.2276 (2019).
27. Ayers, M. *et al.* IFN-Gamma-related mRNA profile predicts clinical response to Pd-1 blockade. *J. Clin. Investig.* **127**(8), 2930–2940. https://doi.org/10.1172/JCI91190 (2017).
28. Lauss, M. *et al.* DNA methylation subgroups in melanoma are associated with proliferative and immunological processes. *BMC Med. Genom.* **8**, 73. https://doi.org/10.1186/s12920-015-0147-4 (2015).
29. Brozyna, A. A., Jozwicki, W., Roszkowski, K., Filipiak, J., Slominski, A. T. Melanin content in melanoma metastases affects the outcome of radiotherapy. *Oncotarget* **7**(14):17844–17853. https://doi.org/10.18632/oncotarget.7528 (2016).
30. Shields, C. L. *et al.* Choroidal nevus transformation into melanoma: analysis of 2514 consecutive cases. *Arch. Ophthalmol.* **127**(8), 981–987. https://doi.org/10.1001/archophthalmol.2009.151 (2009).
31. Shields, C. L. *et al.* Prognosis of uveal melanoma based on race in 8100 patients: The 2015 Doyne lecture. *Eye* **29**(8), 1027–1035. https://doi.org/10.1038/eye.2015.51 (2015).
32. Brozyna, A. A., Jozwicki, W., Carlson, J. A. & Slominski, A. T. Melanogenesis affects overall and disease-free survival in patients with stage III and IV melanoma. *Hum. Pathol.* **44**(10), 2071–2074. https://doi.org/10.1016/j.humpath.2013.02.022 (2013).
33. Slominski, R. M. *et al.* Melanoma, melanin, and melanogenesis: The yin and yang relationship. *Front. Oncol.* **12**, 842496. https://doi.org/10.3389/fonc.2022.842496 (2022).
34. Dai, W. *et al.* Systematic CpG islands methylation profiling of genes in the wnt pathway in epithelial ovarian cancer identifies biomarkers of progression-free survival. *Clin. Cancer Res.* **17**(12), 4052–4062. https://doi.org/10.1158/1078-0432.CCR-10-3021 (2011).
35. Lin, Z. *et al.* Intratumor heterogeneity correlates with reduced immune activity and worse survival in melanoma patients. *Front. Oncol.* **10**, 596493. https://doi.org/10.3389/fonc.2020.596493 (2020).
36. Angell, H. & Galon, J. From the immune contexture to the immunoscore: The role of prognostic and predictive immune markers in cancer. *Curr. Opin. Immunol.* **25**(2), 261–267. https://doi.org/10.1016/j.coi.2013.03.004 (2013).
37. Azimi, F. *et al.* Tumor-infiltrating lymphocyte grade is an independent predictor of sentinel lymph node status and survival in patients with cutaneous melanoma. *J Clin. Oncol.* **30**(21), 2678–2683. https://doi.org/10.1200/JCO.2011.37.8539 (2012).
38. Riaz, N. *et al.* Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell* **171**(4), 934–949. https://doi.org/10.1016/j.cell.2017.09.028 (2017).
39. Sharma, P., Hu-Lieskovan, S., Wargo, J. A. & Ribas, A. Primary, adaptive, and acquired resistance to cancer immunotherapy. *Cell* **168**(4), 707–723. https://doi.org/10.1016/j.cell.2017.01.017 (2017).
40. Jeschke, J. *et al.* DNA methylation-based immune response signature improves patient diagnosis in multiple cancers. *J. Clin. Investig.* **127**(8), 3090–3102. https://doi.org/10.1172/JCI91095 (2017).
41. Gabrilovich, D. I. Myeloid-derived suppressor cells. *Cancer Immunol. Res.* **5**(1), 3–8. https://doi.org/10.1158/2326-6066.CIR-16-0297 (2017).
42. Belgiovine, C., D'Incalci, M., Allavena, P. & Frapolli, R. Tumor-associated macrophages and anti-tumor therapies: Complex links. *Cell Mol. Life Sci.* **73**(13), 2411–2424. https://doi.org/10.1007/s00018-016-2166-5 (2016).
43. Mhaidly, R. & Mechta-Grigoriou, F. Role of cancer-associated fibroblast subpopulations in immune infiltration, as a new means of treatment in cancer. *Immunol. Rev.* **302**(1), 259–272. https://doi.org/10.1111/imr.12978 (2021).
44. Shao, S. *et al.* Centrosomal Nlp is an oncogenic protein that is gene-amplified in human tumors and causes spontaneous tumorigenesis in transgenic mice. *J Clin. Investig.* **120**(2), 498–507. https://doi.org/10.1172/JCI39447 (2010).
45. Harding, M. A. & Theodorescu, D. Rhogdi2: A new metastasis suppressor gene: Discovery and clinical translation. *Urol. Oncol.-Semin. Orig. Investig.* **25**(5), 401–406. https://doi.org/10.1016/j.urolonc.2007.05.006 (2007).
46. Liu, W. *et al.* Rhogdi2 positively regulates the Rho GTPases activation in response to the beta2 outside-in signaling in T cells adhesion and migration on ICAM-1. *J. Leukocyte Biol.* **106**(2), 431–446. https://doi.org/10.1002/JLB.2A0718-272RR (2019).
47. Young JA, Lindsay J, Bodmer JG, Trowsdale J. Epitope Recognition by a Dp Alpha Chain-Specific Monoclonal Antibody (Dp111) is Influenced by the Interaction Between the Dp Alpha Chain and its Polymorphic Dp Beta Chain Partner. Hum Immunol. 1988; 23(1):37–44. https://doi.org/10.1016/0198-8859(88)90016-x.

Acknowledgements

Author contributions

Z.L. and L.Y. created the idea for the paper. Z.L. performed the collection and assembly of data, conducted the analysis, drafted the manuscript, and prepared the figures. L.Y. revised the manuscript. All authors read and approved the final manuscript.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-50614-2.

Correspondence and requests for materials should be addressed to L.Y.

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