

Research Article

Identification and validation of a genomic mutation signature as a predictor for immunotherapy in NSCLC

 Zemin Wang, You Ge,  Han Li, Gaoqiang Fei, Shuai Wang and Pingmin Wei

Department of Epidemiology and Health Statistics, School of Public Health, Southeast University, Nanjing, Jiangsu, China

Correspondence: Pingmin Wei (mpw1963@126.com)



Currently, the benefits of immune checkpoint inhibitor (ICI) therapy prediction via emerging biomarkers have been identified, and the association between genomic mutation signatures (GMS) and immunotherapy benefits has been widely recognized as well. However, the evidence about non-small cell lung cancer (NSCLC) remains limited. We analyzed 310 immunotherapy patients with NSCLC from the Memorial Sloan Kettering Cancer Center (MSKCC) cohort. Lasso Cox regression was used to construct a GMS, and the prognostic value of GMS could be able to verify in the Rizvi cohort ($N=240$) and Hellmann cohort ($N=75$). We further conducted immunotherapy-related characteristics analysis in The Cancer Genome Atlas (TCGA) cohort ($N=1052$). A total of seven genes (*ZFHX3*, *NTRK3*, *EPHA7*, *MGA*, *STK11*, *EPHA5*, *TP53*) were identified for GMS model construction. Compared with GMS-high patients, patients with GMS-low had longer overall survival (OS; $P<0.001$) in the MSKCC cohort and progression-free survival (PFS; $P<0.001$) in the validation cohort. Multivariate Cox analysis revealed that GMS was an independent predictive factor for NSCLC patients in both the MSKCC and validation cohort. Meanwhile, we found that GMS-low patients reflected enhanced antitumor immunity in TCGA cohort. The results indicated that GMS had not only potential predictive value for the benefit of immunotherapy but also may serve as a potential biomarker to guide clinical ICI treatment decisions for NSCLC.

Introduction

Global cancer data for 2020 has shown that lung cancer remained the malignant tumor with the highest mortality rate (18%) worldwide as well as the incidence rate (11.4%) ranked only second to female breast cancer (11.7%) [1]. Approximately 80–85% of those cases are currently classified as non-small cell lung cancer (NSCLC), of which 5-year survival rates were less than 15–20% [2]. Two of the main dominant histological phenotypes of NSCLC include lung adenocarcinoma (LUAD; 50%) and lung squamous cell carcinoma (LUSC; 40%) [3]. Recently, immune checkpoint inhibitors (ICIs) have been applied to NSCLC treatment in order to transform the therapeutic landscape for the condition [4]. In fact, substantial advances in clinical treatment have not provided the equivalent benefit of ICIs among the majority of patients, and the results of the recent clinical trial emphasized the necessity of effective selection for biomarker-based patients [5]. Therefore, it is particularly important to identify and develop potential predictive biomarkers that can be used to predict the efficacy of ICI in dominant populations.

To date, increasing numbers of biomarkers have been confirmed to predict the benefits of immunotherapy. As predictive biomarkers for the ICIs, tumor mutation burden (TMB) and programmed death ligand 1 (PD-L1) expression have been prospectively verified in the randomized controlled trials (RCTs) of NSCLC [6,7]. Nevertheless, TMB and PD-L1 are not beneficial for all NSCLC patients, and it is still required to explore novel biomarkers to maximize clinical benefits [8,9]. At present, a growing body of studies has demonstrated that genomic mutation signatures (GMS) have great potential in predicting tumor prognosis. For example, Jiao et al. established a six-gene-based signature (including genes *RNF43*, *CREBBP*,

Received: 04 May 2022
 Revised: 05 September 2022
 Accepted: 27 October 2022

Accepted Manuscript online:
 28 October 2022
 Version of Record published:
 25 November 2022

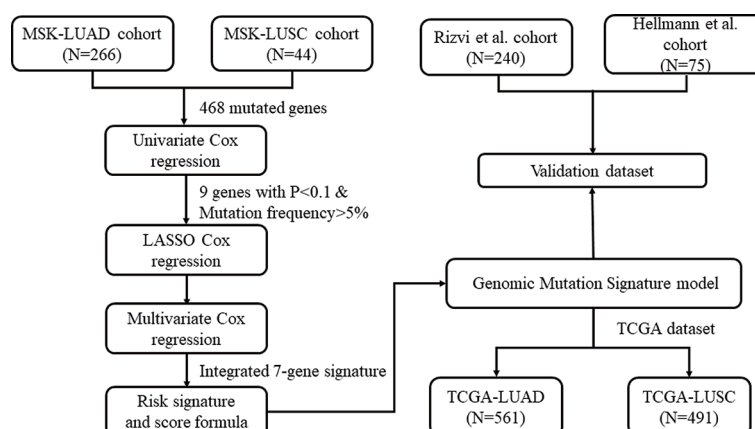


Figure 1. The flowchart of the study design

CDKN2A, *TP53*, *SPEN*, and *NOTCH3*), which was not only a powerful predictive factor of immunotherapy efficacy for gastrointestinal cancer but also may be regarded as the potential biomarker to guide clinical treatment [10]. Similarly, Bai et al. developed another eight-gene-based prognostic model (*HGF*, *KRAS*, *EGFR*, *PTPRD*, *STK11*, *KMT2C*, *SMAD4*, and *TP53*) to predict the response of nonsquamous NSCLC to PD-1 inhibitors [11]. In addition, Pan et al. also constructed a mutation classifier (*TP53*, *PIK3CA*, and *ATM*) to predict the benefits of ICI treatment in bladder cancer patients [12]. Therefore, we can deeply explore the genomic data and identify novel GMSs, so as to guide prognosis stratification and personalized treatments of patients.

In the present study, we integrated the immunotherapy cohort of NSCLC to develop and validate a novel GMS to predict immunotherapy responsiveness. Additionally, we further conducted immunotherapy-related characteristics analysis in The Cancer Genome Atlas (TCGA) cohort as well. Collectively, our results suggested the seven-gene signature could be served as a powerful predictive indicator of immunotherapy and may guide clinical ICI treatment decisions as a potential biomarker in NSCLC.

Materials and methods

Study design and samples

In the present study, a three-step approach was (discovery cohort, validation cohort, and TCGA dataset) applied for developing and validating a GMS for the predictive ability of immunotherapy among patients with NSCLC. The flow chart of the study design was illustrated in Figure 1. In order to evaluate the relationship between gene mutation and the efficacy of ICI, we obtained the clinical and genomic data of advanced cancer patients who have been treated with ICI in the Memorial Sloan Kettering Cancer Center (MSKCC) cohort (<http://www.cbioportal.org/>) [13]. A total of 310 NSCLC patients were identified as the discovery cohort, including 266 with LUAD and 44 with LUSC. Then, we adopted the clinical cohorts, which were treated by ICIs in two published cohorts during the subsequent validation phase. The cohort of 75 patients with advanced NSCLC who received combined immunotherapy was collected from the Hellmann et al.'s study [14]. The cohort of 240 patients with advanced NSCLC treated with anti-PD-1 or anti-PD-L1 therapy was obtained from Rizvi et al.'s study [15]. The mutation and clinical data of each sample in the discovery and validation cohorts were collected from cBioPortal and previous studies [13–15]. In addition, the TCGA cohort was used to explore whether GMS can be considered as a useful indicator for tumor immune microenvironment characteristics. The data of TCGA-LUAD and TCGA-LUSC were obtained from TCGA (<https://portal.gdc.cancer.gov/>). The corresponding clinical data were acquired from UCSC Xena (<http://xena.ucsc.edu/>). Patients who were involved in the three cohorts with incomplete clinical information and mutation data would be excluded.

Clinical outcomes

The clinical outcomes of the present study mainly included progression-free survival (PFS), objective response rate (ORR), overall survival (OS), and durable clinical benefit (DCB). PFS was assessed from the date of initiation of immunotherapy to the time of progress or death due to any causes. ORR was determined based on Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 [16]. OS refers to the time from random grouping to death, which was

caused by any reasons. Complete response (CR), partial response (PR), or stable disease (SD) lasting more than 24 weeks was recognized as a durable clinical benefit (DCB); SD or progressive disease (PD) lasting less than 24 weeks was assumed as no durable benefits (NDB) [17]. Patients who did not develop symptoms and were censored before 24 weeks of follow-up should be defined as not evaluated (NE).

Construction of the GMS

First, univariate Cox regression analyses were performed on the relationship between prognosis-related gene mutations (mutation frequency > 5%) and the survival of 310 patients in two cohorts. Then, we conducted a least absolute shrinkage and selection operator (LASSO) Cox regression analysis on genes with a *P*-value less than 0.1 determined by univariate Cox regression [10]. The multivariate Cox regression analysis was used to create the optimal signature. We used the Survminer R package to generate the optimal cutoff value to divide the patients into the GMS-low group and GMS-high group. The calculation formula of risk score showed as follows:

GMS score = $(\beta_1 \times \text{mutation status of Gene}_1) + (\beta_2 \times \text{mutation status of Gene}_2) + \dots + (\beta_n \times \text{mutation status of Gene}_n)$. Gene mutation status (1 or 0) was coded by mutant and wild-type genes. β was the regression coefficient generated in the multivariate Cox regression analysis.

Immunotherapy-related characteristics analysis

As a representative of the expression in tumor-specific neoantigen, TMB may trigger the immune response. TMB was determined as the total number of nonsynonymous somatic mutations per megabase (Mb) in the genome [18]. For WES data in TCGA dataset, 38 Mb was used as the assessed exon size [19]. For samples in both of MSKCC cohort and validation cohort, TMB data were derived from the Memorial Sloan Kettering Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT).

We characterized tumor immune activation according to studies of immune-related analysis that were previously published, including cytolytic activity (CYT) score [20], inflammation signature score [21], immunologic constant of rejection (ICR) score [22], IFN- γ signaling score [23], antigen processing machinery (APM) score [24], CD8⁺T effector score [25], and the activity of 13 immune-related pathways [26]. These indicators were confirmed to correlate with the efficacy of immunotherapy. The CYT score was determined based on granzyme A (GZMA) and perforin 1 (PRF1) expression [27]. The method of single-sample gene set enrichment analysis (ssGSEA) in the GSVA R package was used to quantify the above indicators [28]. The IFN- γ signaling score was performed via the gene sets of KEYNOTE-012 [29] and POPLAR [30] in ICIs-treated clinical trials. In addition, we also applied the ssGSEA for calculating the infiltration scores of 16 immune cells to estimate the abundance of tumor-infiltrating lymphocytes [28].

Statistical analysis

Statistical analyses were performed using R v. 4.1.1, GraphPad Prism (V.8.0.2), and SPSS V.26.0 (SPSS). Cox regression analysis was performed to establish the GMS. The optimal cut-off value of GMS was conducted by the Survminer R package. The ORR and DCB in different subgroups were analyzed by the χ^2 test or Fisher's exact test. Kaplan–Meier method and log-rank test were applied to calculate PFS and OS. The Man–Whitney U test or Kruskal–Wallis test was used to compare differences between two independent subgroups. All reported *P*-values less than 0.05 were considered statistically significant.

Results

Construction of the GMS

Our study developed a predictive model named GMS based on MSKCC cohort, which included 310 lung cancer patients receiving ICI treatment (MSKCC-LUAD cohort, *N*=266; MSKCC-LUSC, *N*=44). The top 5% mutation frequency and pattern of mutations among patients with LUAD and LUSC from the MSKCC cohort were presented in Figure 2. First, a univariate Cox regression model was performed to select prognosis-related gene mutation (cases with mutation frequency > 5%). The genes with a *P*-value less than 0.1 in univariate Cox analyses were introduced into LASSO Cox regression (Figure 3A,B). After that, these candidate mutation genes were calculated by a multivariate Cox regression model to predict the OS of the MSKCC training cohort. Finally, totally of seven genes (*ZFHX3*, *NTRK3*, *EPHA7*, *MGA*, *STK11*, *EPHA5*, *TP53*) were identified to form the optimal model. Based on the mutation status of the seven genes (1 or 0) weighted by their regression coefficient, the GMS risk model was calculated for each patient (Table 1). GMS score = $(0.652 \times TP53) - (1.052 \times ZFHX3) - (1.111 \times NTRK3) - (0.974 \times EPHA7) - (0.629 \times MGA) + (0.654 \times STK11) - (0.664 \times EPHA5)$. In the calculation formula, the mutant genes were coded

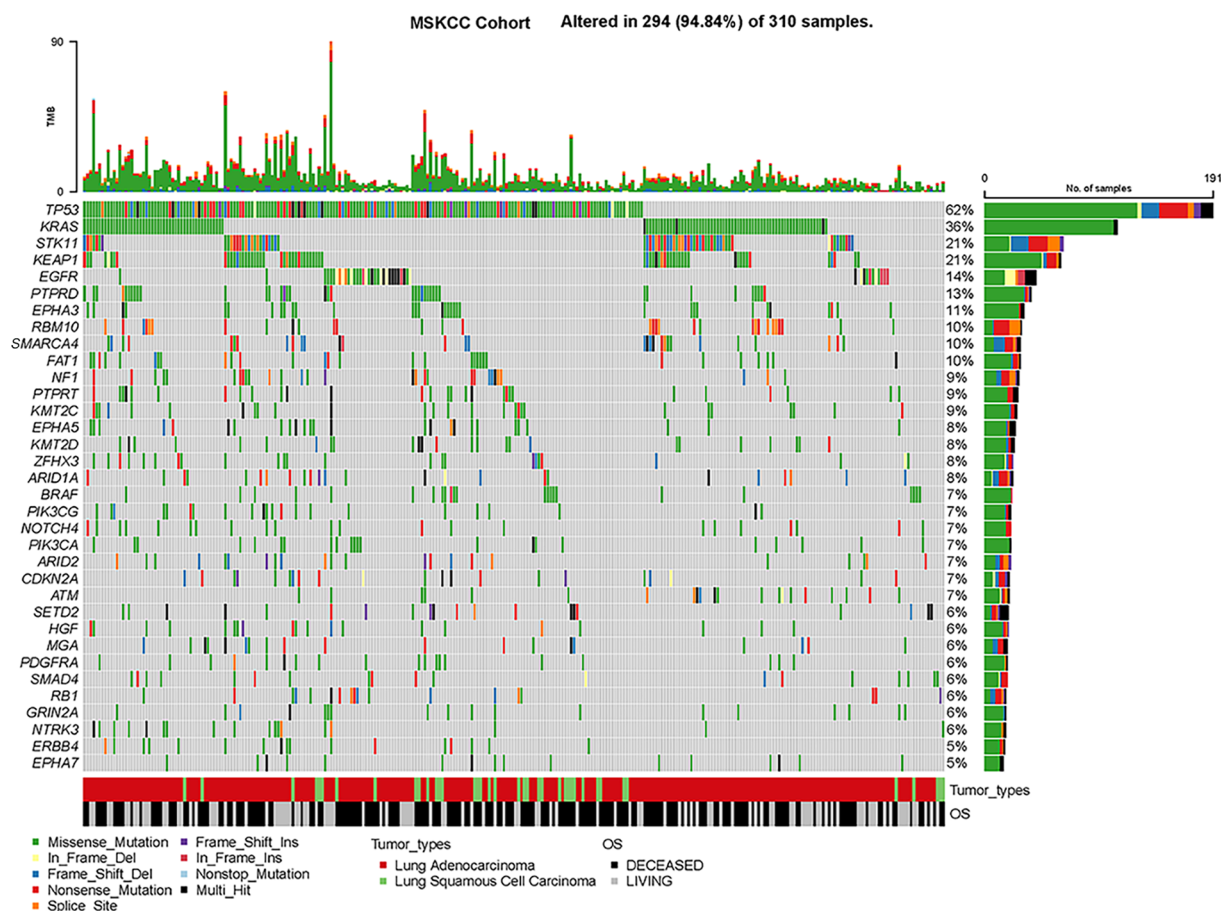


Figure 2. Oncoplot of the mutated genes

Assessment of the frequency and pattern of mutations in patients with NSCLC from the MSKCC cohort. The mutation genes of 310 patients with NSCLC in this cohort were analyzed. Genes were listed by mutation frequency > 5%.

Table 1 Multivariable Cox regression analysis of candidate mutation genes in the MSKCC cohort

Variable	B	HR	95% CI	P-value
ZFHX3	-1.052	0.349	0.152 to 0.802	0.013
NTRK3	-1.111	0.329	0.132 to 0.822	0.017
EPHA7	-0.974	0.377	0.138 to 1.030	0.057
MGA	-0.629	0.533	0.256 to 1.110	0.092
STK11	0.654	1.924	1.345 to 2.752	<0.001
EPHA5	-0.664	0.515	0.276 to 0.962	0.037
TP53	0.652	1.919	1.391 to 2.647	<0.001

Abbreviations: B, regression coefficient; CI, confidence interval; HR, hazard ratio.

as 1 and the wild-type genes were coded as 0. The patients were divided into the GMS-high group and GMS-low group by optimal cutoff value 1, which was calculated by the Survminer R package. Compared with the GMS-high group, patients in the GMS-low group ($P < 0.001$) had longer OS (Figure 3C). In order to appraise the sensitivity and specificity of predictions that identified GMS, receiver operating characteristic (ROC) curves were plotted as well as area under the curve (AUC) values were calculated. The ROC curve results illustrated that GMS performed better predictor ability (AUC = 0.667) compared with TMB (AUC = 0.479) in the MSKCC cohort (Figure 3D). In addition, GMS was considered as an independent prognostic factor in the MSKCC cohort. Table 2 provided the clinical characteristics between GMS-low and GMS-high patients of the MSKCC cohort. After multivariate adjustment of clinicopathological factors, GMS, TMB, and drug-type remained the powerful and independent prognostic factors

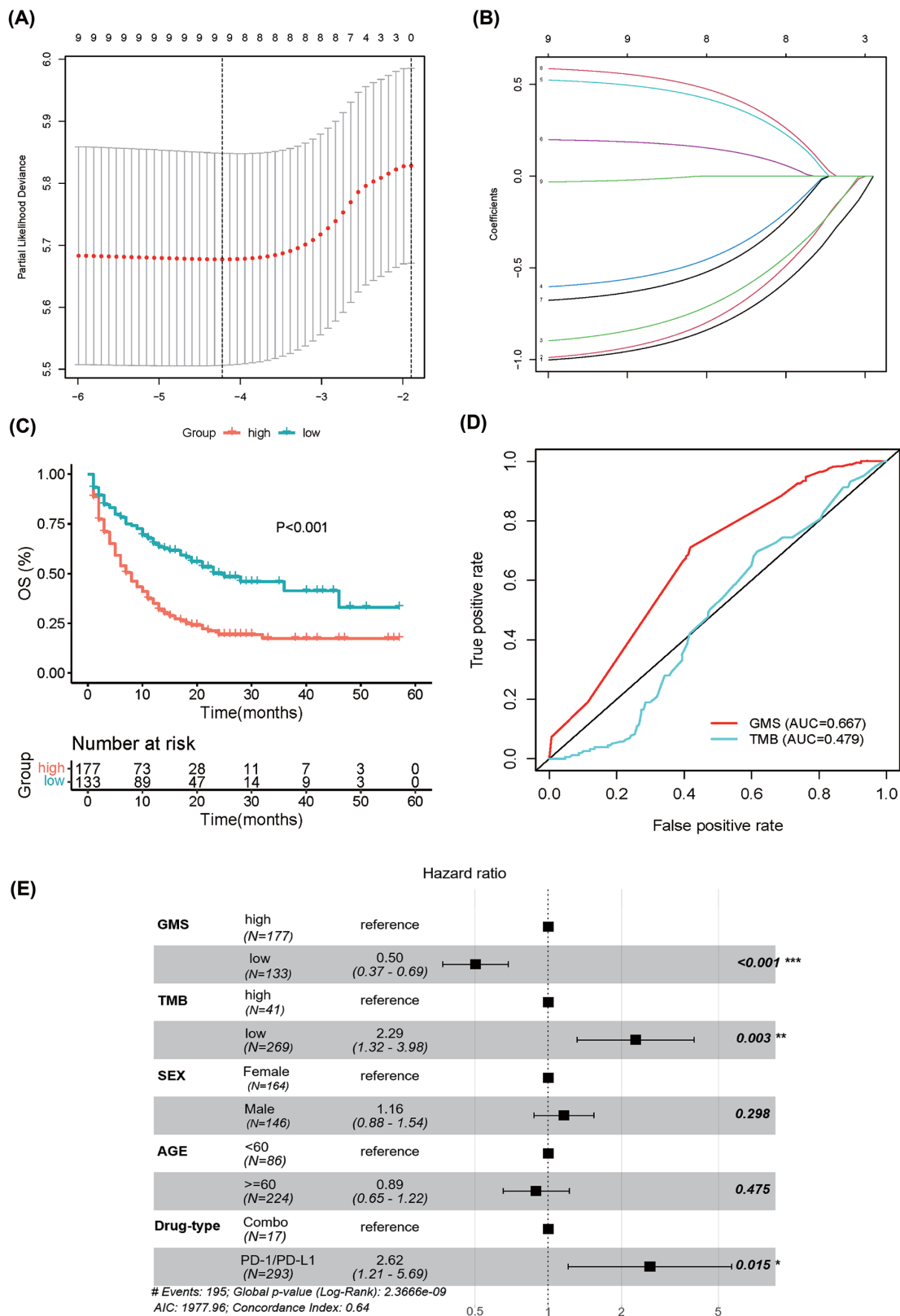


Figure 3. Construction of the GMS
(A) The LASSO regression was performed for variable selection. (B) LASSO coefficient profiles of the nine candidate genes. (C) Kaplan–Meier curves for the OS of patients in the GMS-high group and GMS-low group in the MSKCC cohort. (D) The ROC curve measuring the predictive value of GMS and TMB in the MSKCC cohort. (E) Multivariate Cox analysis for OS in the MSKCC cohort.

Table 2 Clinical characteristics between GMS-low and GMS-high patients of the MSKCC cohort

Characteristics	Classification	GMS-low	GMS-high	χ^2	P-value
Age	≥ 60	94	130	0.291	0.590
	< 60	39	47		
Sex	Male	59	87	0.700	0.403
	Female	74	90		
Drug type	PD-1/PD-L1	122	171	3.490	0.062
	Combo	11	6		
Cancer type	LUAD	116	150	0.381	0.573
	LUSC	17	27		

Abbreviations: LUAD, lung adenocarcinoma, LUSC, lung squamous cell carcinoma.

(GMS: HR 0.50, 0.37–0.69, $P < 0.001$; TMB: HR 2.29, 1.32–3.98, $P = 0.003$; Drug-type: HR 2.62, 1.21–5.69, $P = 0.015$) for OS (Figure 3E).

Validation of the predictive value of GMS

In order to further validate the prognostic ability of the GMS classifier, we integrated two previously published independent cohorts of NSCLC patients received ICIs treatment (Rizvi et al. cohort with 240 patients and Hellman et al. cohort with 75 patients). The patients in the validation cohort were divided into the GMS-high group and GMS-low group base on the optimal cutoff value. Compared with patients in GMS-high group, PFS was detected among patients of GMS-low group ($P < 0.001$) (Figure 4A). The proportion of objective response (CR/PR) in GMS-low patients was over double than that in GMS-high patients (45% vs. 20%, $P = 0.002$) (Figure 4B). The rate of DCB in GMS-low patients was significantly higher than that in GMS-high patients (65.0% vs. 29.1%, $P < 0.001$) (Figure 4C). Furthermore, ROC analysis of the validation cohort indicated that GMS (AUC = 0.619) exhibited better predictive value compared with TMB (AUC = 0.336) and PD-L1 (AUC = 0.350) (Figure 4D). The clinical distribution characteristics between GMS-low and GMS-high patients in the Rizvi and Hellman cohorts were listed in Table 3. After multivariate Cox regression excluded other confounding factors, GMS and Smoke (GMS: HR 0.40, 0.21–0.78; $P = 0.007$; Smoke: HR 1.74, 1.07–2.82, $P = 0.026$) were independent predictors of PFS (Figure 4E). Overall, GMS can be used as a powerful predictor of the outcome of immunotherapy for NSCLC.

Comparison of the immune activation characteristics of GMS in TCGA cohort

According to the above observations, we assumed that GMS would be an indicator of tumor immune microenvironment characteristics for NSCLC patients. For subsequent analysis, we combined TCGA-LUAD ($N = 561$) and TCGA-LUSC ($N = 491$) cohorts. The TCGA cohort was divided into GMS-high and GMS-low based on the GMS risk model. As for OS analysis, Kaplan–Meier survival curves showed no significant differences were identified between GMS-low patients and GMS-high patients ($P = 0.20$) in TCGA cohort (Figure 5A). Tumors with GMS-low showed remarkably more nonsynonymous mutations than those with GMS-high ($P < 0.001$) (Figure 5B). The CYT score calculated by the ssGSEA method of the expression of GZMA and PRF1 was also higher among GMS-low patients ($P = 0.03$) (Figure 5C). Using the ssGSEA methodology, we further evaluated the enrichment scores of 16 immune cells and the activity of 13 immune-related pathways between the GMS-high and GMS-low groups in TCGA cohort. In TCGA cohort, the immune infiltration level of the GMS-high group was commonly lower than that of the GMS-low group, especially B cells ($P = 0.042$), CD8⁺ T cells ($P = 0.010$), and NK cells ($P < 0.001$). Moreover, induced dendritic cells (iDCs) have higher expression in the GMS-high group ($P = 0.037$) (Figure 5D). In addition, four immune pathways showed lower activity in the GMS-high group compared with the GMS-low group in TCGA cohort, including CYT ($P = 0.030$), Inflammation-promoting ($P = 0.008$), MHC class I ($P = 0.016$), and T-cell coinhibition ($P = 0.020$) (Figure 5E). As for tumor immune activation, unlike GMS-high group we noticed that GMS-low group showed higher IFN- γ signaling score (KEYNOTE012, $P = 0.041$; POPLAR, $P = 0.021$), ICR score ($P = 0.036$), inflammation signature score ($P = 0.020$), APM score ($P = 0.015$), and CD8⁺T-effector score ($P = 0.013$) (Figure 5F).

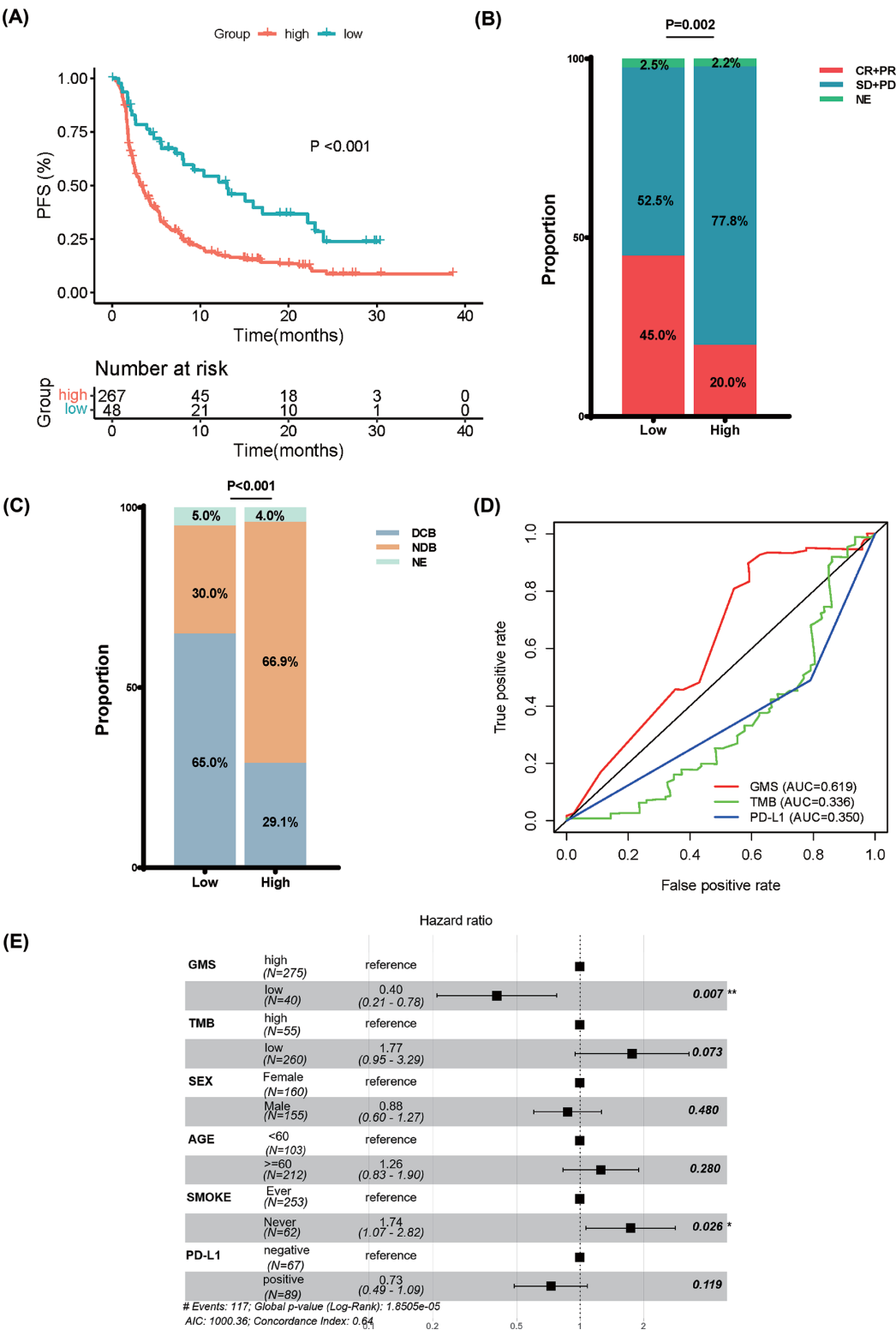


Figure 4. Validation of the predictive value of GMS

(A) Kaplan-Meier curves for the PFS of patients in the GMS-high group and GMS-low group in the Rizvi and Hellman cohorts. (B) The proportion of ORR for patients with GMS-low and GMS-high groups in the Rizvi and Hellman cohorts. (C) The proportion of DCB for patients with GMS-low and GMS-high groups in the Rizvi and Hellman cohorts. (D) ROC curves measuring the predictive value of GMS, TMB, and PD-L1 in the Rizvi and Hellman cohorts. (E) Multivariate Cox analysis for PFS in the Rizvi and Hellman cohorts.

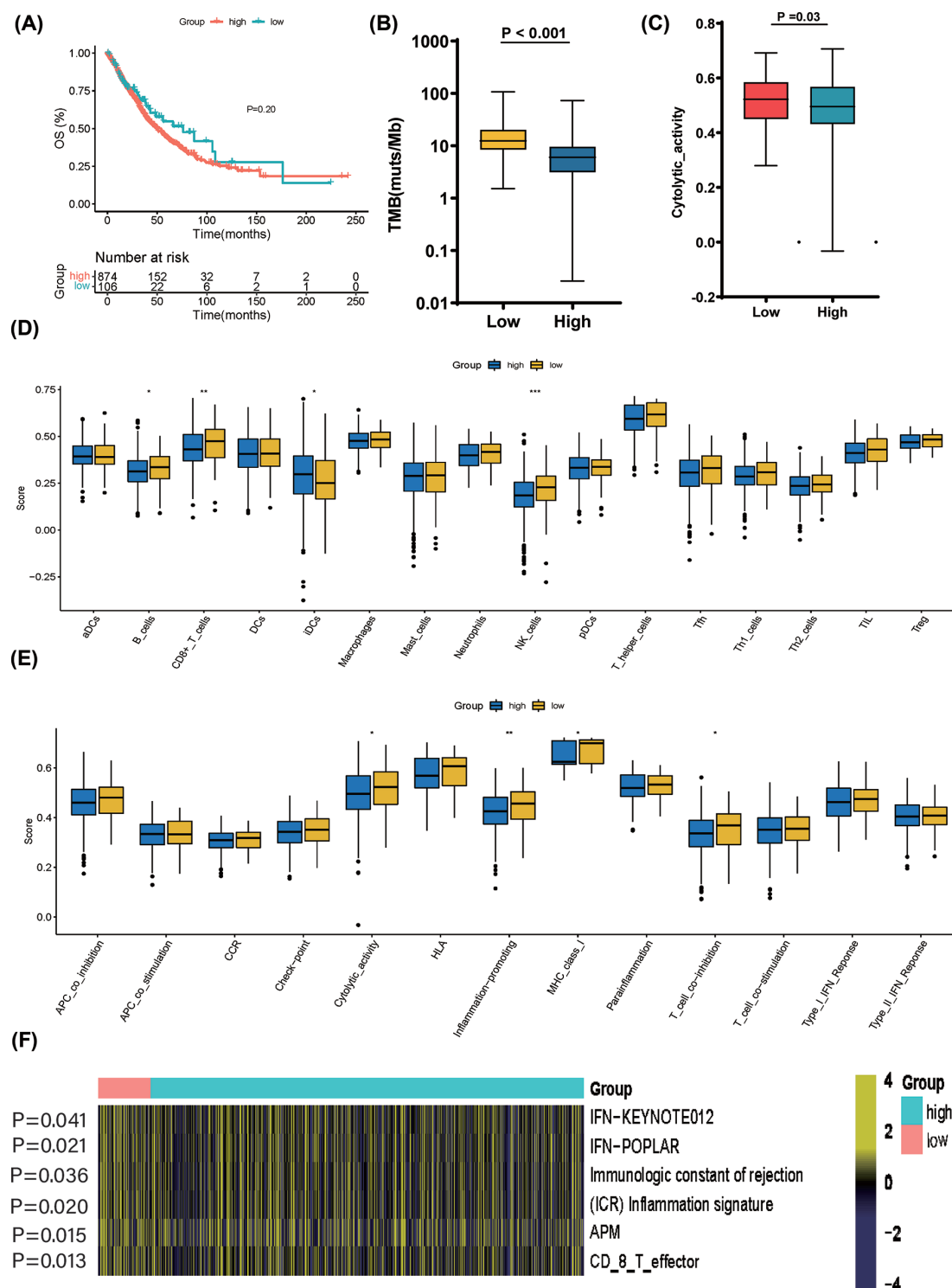


Figure 5. Comparison of the activation characteristics of GMS in TCGA cohort

(A) Kaplan-Meier curves for the OS of patients in the GMS-high group and GMS-low group in the TCGA cohort. (B) Comparison of nonsynonymous mutations between the GMS-low and GMS-high groups in the TCGA cohort. (C) Comparison of cytolytic score between the GMS-low and GMS-high groups in the TCGA cohort. (D) Comparison of the enrichment scores of 16 types of immune cells between the GMS-low group (yellow box) and GMS-high group (blue box) in the TCGA cohort. (E) Comparison of the enrichment scores of 13 immune-related pathways between the GMS-low group (yellow box) and GMS-high group (blue box) in the TCGA cohort. P -values were showed as: *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$. (F) Heatmap of immune-related signatures between the GMS-low and GMS-high groups in the TCGA cohort.

Table 3 Clinical distribution characteristics between GMS-low and GMS-high patients in the Rizvi and Hellman cohorts

Characteristics	Classification	GMS-low	GMS-high	χ^2	P-value
Age	≥ 60	26	186	0.110	0.740
	< 60	14	89		
Sex	Male	18	137	0.324	0.569
	Female	22	138		
Smoke	Ever	34	219	0.635	0.425
	Never	6	56		
PD-(L)1	Positive	16	73	3.268	0.195
	Negative	8	59		
	NE	16	143		
Best overall response	CR/PR	18	55	12.434	0.002
	SD/PD	21	214		
	NE	1	6		
Durable clinical benefit	DCB	26	80	21.111	<0.001
	NDB	12	184		
	NE	2	11		

Abbreviations: CR, complete response, DCB, durable clinical benefit, NDB, no durable benefit, NE, not evaluable, PD, progressive disease, PD-(L)1, programmed cell death-1 or programmed death-ligand 1, PR, partial response, SD, stable disease.

Discussion

Predictive biomarkers may provide a cost-effective method to identify potential responds to immunotherapy and offer an accurate guide for patients receiving ICI therapy. In this multicohort analysis, we systematically gathered and integrated clinical and genomic data, so that could estimate the connection between genomic signatures and clinical response among patients from NSCLC who had been treated with ICI. We developed and validated a prognostic model based on seven gene mutations in order to predict the survival benefits of patients receiving ICI treatment. According to the present study, GMS divided patients into two different subgroups, with the significant OS and PFS advantage in the GMS-low group. Moreover, the patients with NSCLC in GMS-low group had a better prognosis in the ICI treatment cohort than those in the GMS-high group, and they were independent prognostic factors.

In particular, we utilized the TCGA cohort to characterize tumor immune activation. We found that the GMS-low group can be considered as better immunogenicity, as indicated by higher TMB and an immune-inflammatory phenotype, such as increased CD8⁺ T-cell infiltration. When we used the ssGSEA method to calculate the overall immune cell infiltration level of cancer, GMS-low had significantly higher immune scores than GMS-high, which also confirmed that the GMS-low had stronger immune activity. In addition, the tumor microenvironment (TME) is closely associated with ICI efficacy among NSCLC patients. As the central effector cells in the TME, several studies have suggested that highly infiltrated CD8⁺ T cells were associated with the beneficial prognosis for most tumors, including NSCLC [31]. Peripheral immune cells also play an important role in the antitumor immune response. Increased peripheral CD8⁺ CD28⁺ T cells correlated with favorable survival and better treatment response for patients with NSCLC [32–34]. Studies have indicated that high levels of tumor-infiltrated and peripheral Tregs were related to unfavorable prognosis in NSCLC compared with normal levels [35–37]. Thus, higher TMB, B cells, CD8⁺ T cells, and NK cells, among others, may be the reason that ICI was more effective in patients with GMS-low than in those with GMS-high.

At present, some biomarkers have been found to predict treatment outcomes. MSI analysis could be an appropriate biomarker, while the low incidence of MSI-H in lung cancer might limit its clinical application in this population [38]. Emerging studies indicated that gene mutations including ARID1A, TP53, PBRM1, KEAP1, STK11, NOTCH, and JAK may have different effects for ICI treatment [32–36]. Nevertheless, a single gene mutation may not be adequate to transform the ICI treatment landscape and probably is not a sufficient comprehensive biomarker. For example, TP53 mutations were associated with improved PD-L1 expression and promoted CD8⁺ T-cell infiltration, but the single

mutation failed to distinguish sensitive patients with LUAD-receiving immunotherapy. In contrast, the TP53/KRAS comutation combination exhibited a stronger improvement in PD-L1 expression and enhanced tumor immunogenicity compared with the KRAS or TP53 single mutation [39], highlighting the importance of a model that would combine the different genes. We have consequently included most of the possible determinant genes to identify a novel stable GMS to guide patient stratification and personalized treatment.

Currently, PD-L1 and TMB were confirmed to be the primary biomarkers for predicting clinical efficacy of immune checkpoint inhibitors in NSCLC. The ROC curves indicated that the GMS classifier performed better than both TMB and PD-L1. In addition, as PD-L1 and TMB are continuous variables, there is no clear threshold to define whether a response occurs or not. And both biomarkers vary considerably between detection platforms and methods [40,41]. In the present study, seven genes were constructed to form a signature based on mutation data. Nevertheless, risk-scoring formulas and thresholds for mutation-based gene sets can be verified with various methods of tumor analysis, for instance, DNA-sequencing and single nucleotide polymorphism microarray analyses. In this way, mutation-based gene sets are independent from different technical sources, even if various platforms are applied in different centers [27]. Therefore, it is worthwhile to consider a prospective trial to incorporate GMS as a biomarker.

Previous studies have demonstrated the immunological function and prognostic value of seven genes in GMS classifiers for immune therapy. A number of specific genetic variations, for instance, TP53 and STK11, have been proved to create an influence on the infiltration and function of immune cells and the clinical outcomes of ICI therapy [42,43]. As previously reported, the mutation in the DDR pathway can increase the immunogenicity of tumors by accumulating incorrect DNA damage response and promote ICI efficacy [44,45]. There is a possibility that this mechanism contributes to the increased effectiveness of ICIs in patients with ZFX3 mutations [46]. Moreover, the research indicated that NTRK3 has been proven to be a prognostic biomarker related to TMB and can contribute to the development of bladder cancer immunotherapy [47]. EPHA5 mutation may be used as a biomarker to predict the immune response of patients with LUAD by providing insight into genome-wide mutational burden [48]. Besides, EPHA7 mutations have been demonstrated to be predictive biomarkers for immune checkpoint blockades in several cancer types [49]. MAG has also been proved to be an individual indicator affecting the prognosis of liver cancer, which can be used to guide the effective prognosis and treatment with liver cancer patients [50]. In our study, we integrated genomic mutations associated with ICIs prognosis in NSCLC into a risk model. Therefore, GMS offers a crucial promise in predicting immunotherapy efficacy, and nearly all genes play a key role in regulating the tumor immune microenvironment. The GMS-low can be regarded as an immune inflammatory phenotype, with the activity of immune pathway, higher TMB, and effective immune cell infiltration. Our findings suggested that the GMS may offer critical important insights into the immunological characteristics of NSCLC.

There are several limitations of the present study. First of all, the present study is a retrospective research based on multiple public databases, and the problem of inadequate and limited available data is inherent. The limited size of patients involved in the study may limit application of the conclusions. For example, the sample size for squamous lung cancer was restricted due to the availability of samples in the MSKCC cohort; in the validation cohort, PD-L1 could not be evaluable in some patients due to missing data. Second, our analysis only discussed two of the most important NSCLC subtypes, LUSC, and LUAD, and remaining subtypes were not involved. Finally, our results need to be further verified in future prospective trials and validated through laboratory or clinical trials.

Conclusion

Ultimately, we explored and validated a novel seven-gene GMS model, which was correlated with better ICI treatment in patients with NSCLC. Therefore, this signature may be used as one of the predictive biomarkers for ICIs. Furthermore, the signature provides a cost-effective approach and a research framework for the construction and evaluation of predictive biomarkers based on immunotherapy in other tumor types. However, the GMS should be validated in future prospective trials and the mechanisms explored in further molecular studies.

Data Availability

The data that support the findings of the present study are openly available in cBioPortal (<http://www.cbioportal.org/>), TCGA (<https://portal.gdc.cancer.gov/>), and UCSC Xena (<http://xena.ucsc.edu/>).

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

This work was supported by the (Humanities and Social Sciences of Ministry of Education Planning Fund of China) [grant number 16YJA84001].

CRedit Author Contribution

Zemin Wang: Data curation, Formal Analysis, Validation, Investigation, Visualization, Methodology, Writing—original draft. **You Ge:** Data curation, Formal Analysis. **Han Li:** Software, Methodology. **Gaoqiang Fei:** Data curation. **Shuai Wang:** Data curation. **Pingmin Wei:** Supervision.

Acknowledgements

The authors thank MSCKK, TCGA, and UCSC Xena for providing platforms and contributors for uploading meaningful datasets. The authors appreciate the authors of the ICIs-treated cohorts for making the genomic, transcriptomic, and clinical data available. The authors thank Yangyang Liu and Meng Zhao for their contributions to the manuscript.

Abbreviations

APM, antigen processing machinery; AUC, area under the curve; CR, complete response; CYT, cytolytic activity; DCB, durable clinical benefit; GMS, genomic mutation signature; GZMA, granzyme A; ICI, immune checkpoint inhibitor; ICR, immunologic constant of rejection; iDC, induced dendritic cell; LASSO, least absolute shrinkage and selection operator; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; Mb, megabase; MSKCC, Memorial Sloan Kettering Cancer Center; NDB, no durable benefit; NE, not evaluated; NK, natural killer; NSCLC, non-small cell lung cancer; ORR, objective response rate; OS, overall survival; PD-L1, programmed death ligand 1; PFS, progression-free survival; PR, partial response; RCT, randomized controlled trials; SD, stable disease; ssGSEA, single-sample gene set enrichment analysis; TMB, tumor mutation burden; TME, tumor microenvironment; WES, whole exome sequencing.

References

- Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A. et al. (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249, <https://doi.org/10.3322/caac.21660>
- Fitzmaurice, C., Allen, C., Barber, R.M., Barregard, L., Bhutta, Z.A., Brenner, H. et al. (2017) Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the Global Burden of Disease Study. *JAMA Oncol.* **3**, 524–548, <https://doi.org/10.1001/jamaoncol.2016.5688>
- Chen, Z., Fillmore, C.M., Hammerman, P.S., Kim, C.F. and Wong, K.K. (2014) Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat. Rev. Cancer* **14**, 535–546, <https://doi.org/10.1038/nrc3775>
- Doroshov, D.B., Sanmamed, M.F., Hastings, K., Politi, K., Rimm, D.L., Chen, L. et al. (2019) Immunotherapy in non-small cell lung cancer: facts and hopes. *Clin. Cancer Res.* **25**, 4592–4602, <https://doi.org/10.1158/1078-0432.CCR-18-1538>
- Havel, J.J., Chowell, D. and Chan, T.A. (2019) The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* **19**, 133–150, <https://doi.org/10.1038/s41568-019-0116-x>
- Sun, Y., Duan, J., Fang, W., Wang, Z., Du, X., Wang, X. et al. (2021) Identification and validation of tissue or ctDNA PTPRD phosphatase domain deleterious mutations as prognostic and predictive biomarkers for immune checkpoint inhibitors in non-squamous NSCLC. *BMC Med.* **19**, 239, <https://doi.org/10.1186/s12916-021-02075-5>
- Mok, T.S.K., Wu, Y.-L., Kudaba, I., Kowalski, D.M., Cho, B.C., Turna, H.Z. et al. (2019) Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *Lancet North Am. Ed.* **393**, 1819–1830, [https://doi.org/10.1016/S0140-6736\(18\)32409-7](https://doi.org/10.1016/S0140-6736(18)32409-7)
- Boeri, M., Milione, M., Proto, C., Signorelli, D., Lo Russo, G., Galeone, C. et al. (2019) Circulating miRNAs and PD-L1 tumor expression are associated with survival in advanced NSCLC patients treated with immunotherapy: a prospective study. *Clin. Cancer Res.* **25**, 2166–2173, <https://doi.org/10.1158/1078-0432.CCR-18-1981>
- Cristescu, R., Mogg, R., Ayers, M., Albright, A., Murphy, E., Yearley, J. et al. (2018) Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science* **362**, eaar3593, <https://doi.org/10.1126/science.aar3593>
- Jiao, X., Wei, X., Li, S., Liu, C., Chen, H., Gong, J. et al. (2021) A genomic mutation signature predicts the clinical outcomes of immunotherapy and characterizes immunophenotypes in gastrointestinal cancer. *NPJ Precis Oncol.* **5**, 36, <https://doi.org/10.1038/s41698-021-00172-5>
- Bai, X., Wu, D.H., Ma, S.C., Wang, J., Tang, X.R., Kang, S. et al. (2020) Development and validation of a genomic mutation signature to predict response to PD-1 inhibitors in non-squamous NSCLC: a multicohort study. *J. Immunother. Cancer* **8**, e000381, <https://doi.org/10.1136/jitc-2019-000381>
- Pan, Y.H., Zhang, J.X., Chen, X., Liu, F., Cao, J.Z., Chen, Y. et al. (2021) Predictive value of the TP53/PIK3CA/ATM mutation classifier for patients with bladder cancer responding to immune checkpoint inhibitor therapy. *Front. Immunol.* **12**, 643282, <https://doi.org/10.3389/fimmu.2021.643282>
- Samstein, R.M., Lee, C.H., Shoushtari, A.N., Hellmann, M.D., Shen, R., Janjigian, Y.Y. et al. (2019) Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* **51**, 202–206, <https://doi.org/10.1038/s41588-018-0312-8>
- Hellmann, M.D., Nathanson, T., Rizvi, H., Creelan, B.C., Sanchez-Vega, F., Ahuja, A. et al. (2018) Genomic features of response to combination immunotherapy in patients with advanced non-small-cell lung cancer. *Cancer Cell.* **33**, 843.e4–52.e4, <https://doi.org/10.1016/j.ccell.2018.03.018>

- 15 Rizvi, H., Sanchez-Vega, F., La, K., Chatila, W., Jonsson, P., Halpenny, D. et al. (2018) Molecular determinants of response to anti-programmed cell death (PD)-1 and anti-programmed death-ligand 1 (PD-L1) blockade in patients with non-small-cell lung cancer profiled with targeted next-generation sequencing. *J. Clin. Oncol.* **36**, 633–641, <https://doi.org/10.1200/JCO.2017.75.3384>
- 16 Hugo, W., Zaretsky, J.M., Sun, L., Song, C., Moreno, B.H., Hu-Lieskovan, S. et al. (2016) Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* **165**, 35–44, <https://doi.org/10.1016/j.cell.2016.02.065>
- 17 Rizvi, N.A., Hellmann, M.D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J.J. et al. (2015) Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128, <https://doi.org/10.1126/science.aaa1348>
- 18 Yarchoan, M., Hopkins, A. and Jaffee, E.M. (2017) Tumor mutational burden and response rate to PD-1 inhibition. *N. Engl. J. Med.* **377**, 2500–2501, <https://doi.org/10.1056/NEJMc1713444>
- 19 Chalmers, Z.R., Connelly, C.F., Fabrizio, D., Gay, L., Ali, S.M., Ennis, R. et al. (2017) Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* **9**, 34, <https://doi.org/10.1186/s13073-017-0424-2>
- 20 Narayanan, S., Kawaguchi, T., Yan, L., Peng, X., Qi, Q. and Takabe, K. (2018) Cytolytic activity score to assess anticancer immunity in colorectal cancer. *Ann. Surg. Oncol.* **25**, 2323–2331, <https://doi.org/10.1245/s10434-018-6506-6>
- 21 Thompson, J.C., Hwang, W.T., Davis, C., Deshpande, C., Jeffries, S., Rajpurohit, Y. et al. (2020) Gene signatures of tumor inflammation and epithelial-to-mesenchymal transition (EMT) predict responses to immune checkpoint blockade in lung cancer with high accuracy. *Lung Cancer* **139**, 1–8, <https://doi.org/10.1016/j.lungcan.2019.10.012>
- 22 Roelands, J., Hendrickx, W., Zoppoli, G., Mall, R., Saad, M., Halliwill, K. et al. (2020) Oncogenic states dictate the prognostic and predictive connotations of intratumoral immune response. *J. Immunother. Cancer* **8**, e000617, <https://doi.org/10.1136/jitc-2020-000617>
- 23 Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D.R. et al. (2017) IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* **127**, 2930–2940, <https://doi.org/10.1172/JCI91190>
- 24 Maggs, L., Sadagopan, A., Moghaddam, A.S. and Ferrone, S. (2021) HLA class I antigen processing machinery defects in antitumor immunity and immunotherapy. *Trends Cancer* **7**, 1089–1101, <https://doi.org/10.1016/j.trecan.2021.07.006>
- 25 Chen, X., Xu, R., He, D., Zhang, Y., Chen, H., Zhu, Y. et al. (2021) CD8(+) T effector and immune checkpoint signatures predict prognosis and responsiveness to immunotherapy in bladder cancer. *Oncogene* **40**, 6223–6234, <https://doi.org/10.1038/s41388-021-02019-6>
- 26 Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G. and Hacohen, N. (2015) Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61, <https://doi.org/10.1016/j.cell.2014.12.033>
- 27 Long, J., Wang, D., Wang, A., Chen, P., Lin, Y., Bian, J. et al. (2022) A mutation-based gene set predicts survival benefit after immunotherapy across multiple cancers and reveals the immune response landscape. *Genome Medicine* **14**, 20, <https://doi.org/10.1186/s13073-022-01024-y>
- 28 Hänzelmann, S., Castelo, R. and Guinney, J. (2013) GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7, <https://doi.org/10.1186/1471-2105-14-7>
- 29 Muro, K., Chung, H.C., Shankaran, V., Geva, R., Catenacci, D., Gupta, S. et al. (2016) Pembrolizumab for patients with PD-L1-positive advanced gastric cancer (KEYNOTE-012): a multicentre, open-label, phase 1b trial. *Lancet Oncol.* **17**, 717–726, [https://doi.org/10.1016/S1470-2045\(16\)00175-3](https://doi.org/10.1016/S1470-2045(16)00175-3)
- 30 Fehrenbacher, L., Spira, A., Ballinger, M., Kowanzet, M., Vansteenkiste, J., Mazieres, J. et al. (2016) Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. *Lancet* **387**, 1837–1846, [https://doi.org/10.1016/S0140-6736\(16\)00587-0](https://doi.org/10.1016/S0140-6736(16)00587-0)
- 31 Donnem, T., Hald, S.M., Paulsen, E.E., Richardsen, E., Al-Saad, S., Kilvaer, T.K. et al. (2015) Stromal CD8+ T-cell density—a promising supplement to TNM staging in non-small cell lung cancer. *Clin. Cancer Res.* **21**, 2635–2643, <https://doi.org/10.1158/1078-0432.CCR-14-1905>
- 32 Liu, C., Hu, Q., Hu, K., Su, H., Shi, F., Kong, L. et al. (2019) Increased CD8+CD28+ T cells independently predict better early response to stereotactic ablative radiotherapy in patients with lung metastases from non-small cell lung cancer. *J. Transl. Med.* **17**, 120, <https://doi.org/10.1186/s12967-019-1872-9>
- 33 Liu, C., Hu, Q., Xu, B., Hu, X., Su, H., Li, Q. et al. (2019) Peripheral memory and naïve T cells in non-small cell lung cancer patients with lung metastases undergoing stereotactic body radiotherapy: predictors of early tumor response. *Cancer Cell Int.* **19**, 121, <https://doi.org/10.1186/s12935-019-0839-5>
- 34 Liu, C., Jing, W., An, N., Li, A., Yan, W., Zhu, H. et al. (2019) Prognostic significance of peripheral CD8+CD28+ and CD8+CD28- T cells in advanced non-small cell lung cancer patients treated with chemo(radio)therapy. *J. Transl. Med.* **17**, 344, <https://doi.org/10.1186/s12967-019-2097-7>
- 35 Liu, C., Sun, B., Hu, X., Zhang, Y., Wang, Q., Yue, J. et al. (2019) Stereotactic ablative radiation therapy for pulmonary recurrence-based oligometastatic non-small cell lung cancer: survival and prognostic value of regulatory T Cells. *Int. J. Radiat. Oncol. Biol. Phys.* **105**, 1055–1064
- 36 Peng, H., Wu, X., Zhong, R., Yu, T., Cai, X., Liu, J. et al. (2021) Profiling tumor immune microenvironment of non-small cell lung cancer using multiplex immunofluorescence. *Front. Immunol.* **12**, 750046, <https://doi.org/10.3389/fimmu.2021.750046>
- 37 Chuckran, C.A., Cillo, A.R., Moskovitz, J., Overacre-Delgoffe, A., Somasundaram, A.S., Shan, F. et al. (2021) Prevalence of intratumoral regulatory T cells expressing neuropilin-1 is associated with poorer outcomes in patients with cancer. *Sci. Transl. Med.* **13**, eabf8495, <https://doi.org/10.1126/scitranslmed.abf8495>
- 38 Bonneville, R., Krook, M.A., Kautto, E.A., Miya, J., Wing, M.R., Chen, H.Z. et al. (2017) Landscape of microsatellite instability across 39 cancer types. *JCO Precision Oncol.* **2017**, 1–15, <https://doi.org/10.1200/PO.17.00073>
- 39 Dong, Z.Y., Zhong, W.Z., Zhang, X.C., Su, J., Xie, Z., Liu, S.Y. et al. (2017) Potential predictive value of TP53 and KRAS mutation status for response to PD-1 blockade immunotherapy in lung adenocarcinoma. *Clin. Cancer Res.* **23**, 3012–3024, <https://doi.org/10.1158/1078-0432.CCR-16-2554>
- 40 Tsao, M.S., Kerr, K.M., Kockx, M., Beasley, M.B., Borczuk, A.C., Botling, J. et al. (2018) PD-L1 immunohistochemistry comparability study in real-life clinical samples: results of blueprint phase 2 project. *J. Thoracic Oncol.* **13**, 1302–1311, <https://doi.org/10.1016/j.jtho.2018.05.013>
- 41 Addeo, A., Banna, G.L. and Weiss, G.J. (2019) Tumor mutation burden—from hopes to doubts. *JAMA Oncol.* **5**, 934–935, <https://doi.org/10.1001/jamaoncol.2019.0626>

- 42 Wellenstein, M.D., Coffelt, S.B., Duits, D.E.M., van Miltenburg, M.H., Slagter, M., de Rink, I. et al. (2019) Loss of p53 triggers WNT-dependent systemic inflammation to drive breast cancer metastasis. *Nature* **572**, 538–542, <https://doi.org/10.1038/s41586-019-1450-6>
- 43 Papillon-Cavanagh, S., Doshi, P., Dobrin, R., Szustakowski, J. and Walsh, A.M. (2020) STK11 and KEAP1 mutations as prognostic biomarkers in an observational real-world lung adenocarcinoma cohort. *ESMO Open* **5**, e000706, <https://doi.org/10.1136/esmoopen-2020-000706>
- 44 Wang, Z., Zhao, J., Wang, G., Zhang, F., Zhang, Z., Zhang, F. et al. (2018) Comutations in DNA damage response pathways serve as potential biomarkers for immune checkpoint blockade. *Cancer Res.* **78**, 6486–6496, <https://doi.org/10.1158/0008-5472.CAN-18-1814>
- 45 Teo, M.Y., Seier, K., Ostrovnya, I., Regazzi, A.M., Kania, B.E., Moran, M.M. et al. (2018) Alterations in DNA damage response and repair genes as potential marker of clinical benefit from PD-1/PD-L1 blockade in advanced urothelial cancers. *J. Clin. Oncol.* **36**, 1685–1694, <https://doi.org/10.1200/JCO.2017.75.7740>
- 46 Zhang, J., Zhou, N., Lin, A., Luo, P., Chen, X., Deng, H. et al. (2021) ZFH3 mutation as a protective biomarker for immune checkpoint blockade in non-small cell lung cancer. *Cancer Immunol. Immunother.* **70**, 137–151, <https://doi.org/10.1007/s00262-020-02668-8>
- 47 Zhang, Z., Yu, Y., Zhang, P., Ma, G., Zhang, M., Liang, Y. et al. (2021) Identification of NTRK3 as a potential prognostic biomarker associated with tumor mutation burden and immune infiltration in bladder cancer. *BMC Cancer* **21**, 458, <https://doi.org/10.1186/s12885-021-08229-1>
- 48 Huang, W., Lin, A., Luo, P., Liu, Y., Xu, W., Zhu, W. et al. (2021) EPHA5 mutation predicts the durable clinical benefit of immune checkpoint inhibitors in patients with lung adenocarcinoma. *Cancer Gene Ther.* **28**, 864–874, <https://doi.org/10.1038/s41417-020-0207-6>
- 49 Zhang, Z., Wu, H.X., Lin, W.H., Wang, Z.X., Yang, L.P., Zeng, Z.L. et al. (2021) EPHA7 mutation as a predictive biomarker for immune checkpoint inhibitors in multiple cancers. *BMC Med.* **19**, 26, <https://doi.org/10.1186/s12916-020-01899-x>
- 50 Liu, M., Liu, X., Liu, S., Xiao, F., Guo, E., Qin, X. et al. (2020) Big data-based identification of multi-gene prognostic signatures in liver cancer. *Front. Oncol.* **10**, 847, <https://doi.org/10.3389/fonc.2020.00847>