

A 15-DNA repair gene prognostic signature for immunotherapy in gastric cancer

Binbin Yuan^{1*}, Chengfei Jiang^{1*}, Lingyan Chen^{1*}, Jinlong Cui², Min Chen¹, Shu Zhang¹, Yimeng Cai³, Jian-Hua Mao⁴, Xiaoping Zou^{1#}, Bo Hang^{4#}, Pin Wang^{1#}

¹Department of Gastroenterology, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing, Jiangsu 210008, China.

²Berkeley-Nanjing Research Center, Nanjing, Jiangsu, China

³Department of Molecular and Cell Biology, University of California at Berkeley, CA 94720, USA.

⁴Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

* These three authors contributed equally.

Corresponding authors:

Dr. Xiaoping Zou: 13770771661@163.com

Dr. Pin Wang: pinwang729@126.com

Department of Gastroenterology, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing, Jiangsu 210008, China.

Dr. Bo Hang: Bo_hang@lbl.gov

Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

Abbreviations

BER	base excision repair
CIN	chromosomal instability
CR	complete response
DRGS	15-DNA repair gene signature
DSB	double-strand break
GDSC	Genomics of Drug Sensitivity in Cancer
GS	genomically stable
HR	homologous recombination
HRD	homologous recombination deficiency
ICB	immune checkpoint blockade
ICI	immune checkpoint inhibitors
MMR	mismatch repair
MSI	microsatellite instability
MSI-H	microsatellite instability-high
MSI-L	microsatellite instability-low
MSS	Microsatellite Instability-Stable
mUC	metastatic urothelial carcinoma
NER	nucleotide excision repair
NHEJ	non-homologous end joining
OS	overall survival
PCA	principal component analysis
PD	progressive disease
PR	partial response
RFS	relapse-free survival
SD	stable disease
SNV	single nucleotide variants
TCGA-STAD	The Cancer Genome Atlas Stomach Adenocarcinoma
TMB	tumor mutation burden

Abstract

Gastric cancer is a heterogeneous group of diseases with only a fraction of patients responds to immunotherapy. The relationships between tumor DNA damage response, the immune system and immunotherapy have recently attracted attention. Accumulating evidence indicate that DNA repair landscape is a significant factor in driving response to immune checkpoint blockade (ICB) therapy. In this study, to explore new prognostic and predictive biomarkers for gastric cancer patients who are sensitive and responsible to immunotherapy, we developed a novel 15-DNA repair gene signature (DRGS) and its related scoring system and evaluated the efficiency of DRGS in discriminating different molecular and immune characteristics and therapeutic outcomes of gastric adenocarcinoma. The results showed that DRGS high score patients showed significantly better therapeutic outcomes compared to DRGS low score patients ($P < 0.001$). Integrated analysis of multi-omics data demonstrated that the patients with high DRGS score were characteristic of high levels of anti-tumor lymphocytes infiltration, tumor mutation burden (TMB) and PD-L1 expression, and these patients exhibited a longer overall survival and may benefit more from ICB therapy, as compared to the low-score patients. Therefore, the DRGS and its scoring system may have implications in tailoring immunotherapy in gastric cancers.

Key words: gastric cancer, immunotherapy, immune checkpoint blockade (ICB), immune checkpoint inhibitors (ICI), DNA repair gene signature (DRGS), prognostic biomarker, score system

1. Introduction

Gastric cancer is one of the most common cancer worldwide, with more than one million new cases in 2020, ranking fifth in incidence and third in mortality worldwide [1]. Up to now the prognosis of this disease remains dismal. Gastric cancer is a heterogeneous group of diseases with variable responsiveness to treatment such as chemotherapy and immunotherapy, and new biomarkers are needed to identify patients with gastric cancer for sensitivity toward such therapies.

Chemotherapy is used as standard treatment, either as preoperative or postoperative therapy, at all pathological stages of gastric cancer. Some regimens showed better 5-year survival rate (OS) or relapse-free survival (RFS) in certain patients [2,3]. However, some regimens demonstrated considerable toxicity and mortality in gastric patients.

In recent years, the approval of multiple immune checkpoint inhibitors (ICIs) and the promising results from tumor immune clinical trials have led to the development of tumor immunotherapy [2,3]. In ICI therapies, anti-PD-1 and anti-PD-L1 antibodies and anti-CTLA4 antibodies are highly effective in patients with microsatellite instability-high (MSI-H) subtype or high expression of PD-L1 [4,5]. Most recently, the effect of immunotherapy combined with chemotherapy or targeted therapy on gastric cancer has been studied. For example, in CheckMate-649, the largest global randomized phase III clinical study in gastric cancer [6], nivolumab plus chemotherapy as first-line treatment for previously untreated and advanced gastric cancer demonstrated superior overall survival (OS) and relapse-free survival (RFS) compared to chemotherapy alone, reducing the risk of death by 20%. Some conflicting results have also been obtained which may be due to different chemotherapy regimens and heterogeneity of patients. In summary, even though its impact on the outcome of gastric cancer has not been clearly defined, immunotherapy

plays an important role in the treatment of gastric cancer, and combinational therapy is becoming a new trend.

The association between genetic instability or DNA repair defects with tumor susceptibility to immunotherapy has been observed in certain types of tumors. DNA damage and genomic instability have been found to affect the anti-tumor immune response. There are several major repair pathways for the removal of different types of exogenous and endogenous DNA lesions, including direct reversal, base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR) and double-strand break (DSB) repair that includes homologous recombination (HR), non-homologous end joining (NHEJ) and the Fanconi anemia pathway [7-9]. Deficiency in such repair is associated with reduced DNA repair capacity and increased genetic instability, thus promoting cancer development, or providing an opportunity/benefit for cancer therapy, as the efficacy of certain anticancer drugs or therapies is highly influenced by cellular DNA repair capacity, for example, small-molecule inhibitors of DNA repair have been combined with conventional chemotherapy drugs [10]. As mentioned above, the noticeable effectivity of ICIs in cancers with MMR-deficient and its characteristic genetic signature MSI has recently been discovered, and a recent large-scale analysis showed that mutations of mismatch repair genes and DNA polymerases account for 13.5% of high TMB tumors [11,12]. There is also evidence of potential implications for immunotherapy in cancers with homologous recombination deficiency (HRD) and somatic changes in the NER pathway [13]. Taken together, accumulating evidence suggest that a systematic understanding of DNA repair landscape in tumor may help assess the tumor susceptibility to immunotherapy.

There are still many challenges in tumor immunotherapy, such as only small portion of tumors exhibiting an effect, the overall clinical response rate is low, and the recurrence after treatment. These prompted searching for biomarker strategies for immunotherapy, such as those based on tumor-infiltrating lymphocytes, TMB,

deficient MMR, immune gene signatures, and multiplex immunohistochemistry [14]. More recently, a role for DNA repair in selection of patients for immunotherapy has been emerged. However, no repair gene signature(s) has been developed so far. In this study, we aimed to explore the prognostic role of DNA repair genes expression in relation to predict response and better select patients for immunotherapy in gastric cancer. We identified a 15-DNA repair gene signature (DRGS) that is associated with the prognosis of gastric cancer and developed a scoring system that predicts response and better select patients for immunotherapy. In addition, the association of the DRGS score with the molecular and immune profiles were further investigated. The results suggest that our DRGS is a promising biomarker for tailoring immunotherapy.

2. Materials and Methods

2.1. Datasets used in this study

The next-generation sequencing data and clinicopathological information for gastric cancer were downloaded from The Cancer Genome Atlas Stomach Adenocarcinoma (TCGA-STAD) data collection (<https://portal.gdc.cancer.gov>). The microarray data (GSE30727) were downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE30727>). We also employed a dataset with immunotherapy information (R package “IMvigor210”) to evaluate the prediction efficiency of our scoring system [15]. Drugs sensitivity data among different gastric cancer cells in vitro were download from Genomics of Drug Sensitivity in Cancer (GDSC, <http://www.cancerrxgene.org/downloads>).

2.2. Construction of the 15-DNA repair gene signature and related scoring system

We used DEseq2 (version 1.28.1) package in R (version 4.0.5) to analyze the absolute counts for differential gene expression in TCGA-STAD and employed the online platform GEO2R to access the differentially expressed genes (adjusted p-value < 0.05 , $|\log_2FC| > 1$) between tumor and normal tissues. From the Human DNA Repair Genes website (<https://www.mdanderson.org/documents/Labs/Wood-Laboratory/human-dna-repair-genes.html>) [16] which was last modified by Drs. R.

Wood and M. Lowery on 10th June 2020, we identified 15 common genes as a DNA repair gene signature (DRGS) among the 219 DNA repair genes and the differentially expressed gene lists from TCGA-STAD and GSE30727, which are: *PRKDC*, *FANCI*, *LIG1*, *RECQL4*, *FANCA*, *BRCA2*, *FANCD2*, *UBE2T*, *POLQ*, *PARPBP*, *EXO1*, *XRCC2*, *RAD54L*, *EME1* and *FANCB*.

Before the analysis, we applied the Z-score normalization to the expression level of 15 genes from each sample. The principal component analysis (PCA) is a proven technique, which can reduce dimension, improve interpretability and minimize information loss of large datasets at the same time. We employed the PCA as a dimensionality reduction method to obtain the 1st dimension correlation coefficients of the 15 genes (Figure S1).

The formula for the DRGS scoring system was as follow:

$$DNA\ repair\ score = \sum (correlation\ coefficients\ for\ gene\ i) * (normalized\ gene\ i\ expression\ level)$$

2.3. Comparison of the immunocyte infiltration

To determine the relationship between this DRGS scoring system and immunocyte infiltration in the gastric cancer tissue, we performed the immunocyte infiltration analysis through the online platform ImmuCellAI (<http://bioinfo.life.hust.edu.cn/ImmuCellAI#!/>) using TCGA-STAD samples [17]. Then we calculated DRGS scores from the same TCGA-STAD samples and used the median score as the cut-point to divide these samples into high-score subtype and low-score subtype. Mann Whitney U test was employed to compare the multiple immunocyte infiltration results between these two subtype groups. We applied the hierarchical clustering to the immunocyte whose infiltration score were significantly different between two groups. Mann Whitney U test was used to compare the infiltration score between high-score subtype and low-score subtype. We also used ImmuCellAI to predict the patient response to ICB therapy.

2.4. Comprehensive analysis of histologic and molecular characteristic

Samples from the two groups mentioned above were also used to explore the relationship between the DRGS score and various tumor characteristics. We compared the Lauren classification, molecular subtype and MSI status using chi-square test. Tumor mutation levels were calculated through R package “Maftools” [18]. We used Mann Whitney U test to compare TMB and PD-L1 expression between the two groups. In light that the TMB and PD-L1 expression are strongly related to the ICI therapy, we performed the spearman correlation to analyze the relationship between the DRGS score and the above two immunotherapeutic biomarkers.

2.5. Relationship between the DRGS score and immunotherapy outcomes

The public dataset IMvigor210 contains a total of 348 advanced urothelial bladder cancer patients who received atezolizumab treatment. We applied DRGS to the dataset to investigate its predictive ability of immunotherapy. We performed normalization and PCA analysis with the same methodology as mentioned above, and all the 348 patients were divided into two equal groups according to their DRGS score. We then compared their response to immunotherapy and the immunotypes of tumor by the Chi-square test. Also, the Kaplan-Meier analysis of OS within the two groups was performed, and log-rank test was used to compare survival curves.

2.6. Prognostic ability of DNA repair score

We also analyzed the prognostic ability of the DRGS scoring system in TCGA-STAD samples, through Kaplan-Meier analysis of OS and the log-rank test for comparison. The independence of DRGS score system was verified by univariate and multivariate Cox regression. Independent gastric cancer datasets (GSE26901, GSE15459 and GSE26899) were served as validation sets and analyzed with the same methodology described above.

2.7. Association of the DRGS score with drug sensitivity

We downloaded the drugs sensitivity data combined with the RNA-seq data from 23 types of GC cells. We performed the Pearson correlation analysis to calculate the

correlation between the drug sensitivity level of different chemotherapeutic agents and DNA repair score. The criteria for significant correlation are p value < 0.05 and correlation > 0.4 .

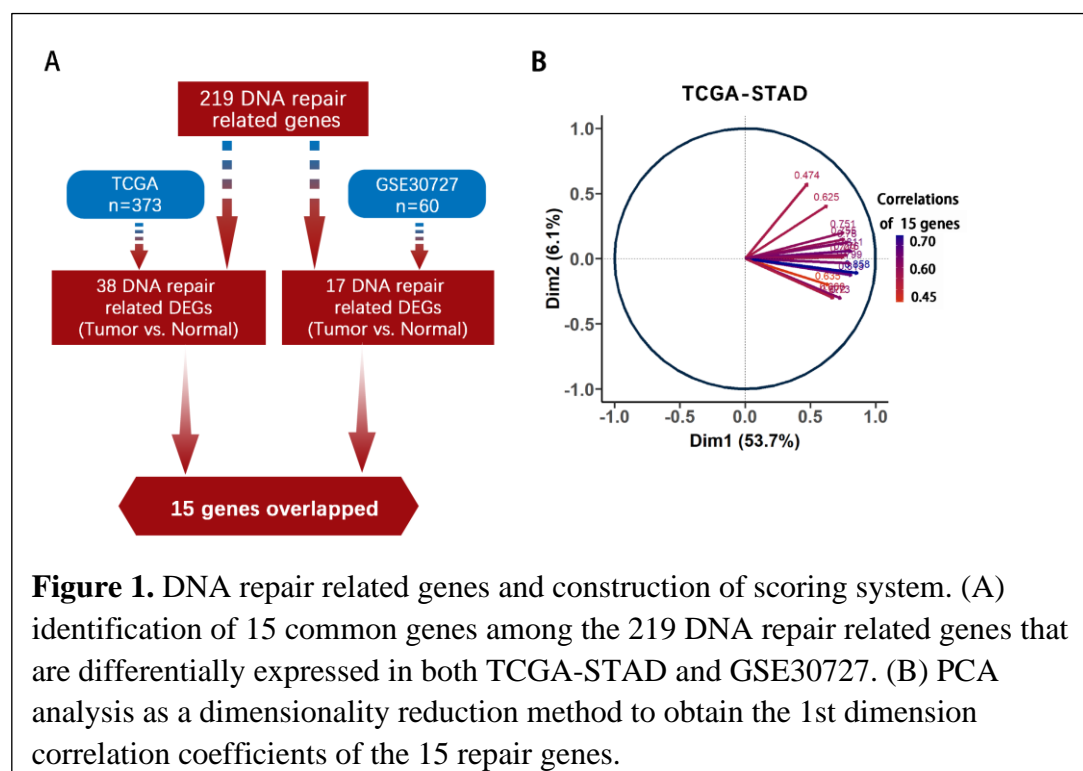
2.8. Statistical analysis

Statistical methods were described in different sections above. The significant level of these statistic tests was two-sided p -value < 0.05 . The software used for analysis is R (version 4.0.5).

3. Results

3.1. Identification of a 15-DNA repair gene signature (DRGS) and construction of the scoring system

The 219 DNA repair genes were acquired from the website Human DNA Repair Genes. Next, we used DEseq2 package of R to process the raw data from TCGA-STAD and employed the online platform GEO2R to find those differentially expressed genes (adjusted p -value < 0.05 , $|\log_2FC| > 1$) between tumor and normal tissues. Finally, we identified 15 common genes among the 219 DNA repair gene,



which were differentially expressed in both TCGA-STAD and GSE30727 (Figure 1A, Table 1). We then employed the PCA analysis as a dimensionality reduction method to obtain the first dimension correlation coefficients of 15 genes (Figure 1B). Based on these coefficients, we created a scoring system for the 15-DNA repair gene signature (DRGS) (details refer to the method). Then patients were divided into two group using the DRGS score (high and low score groups).

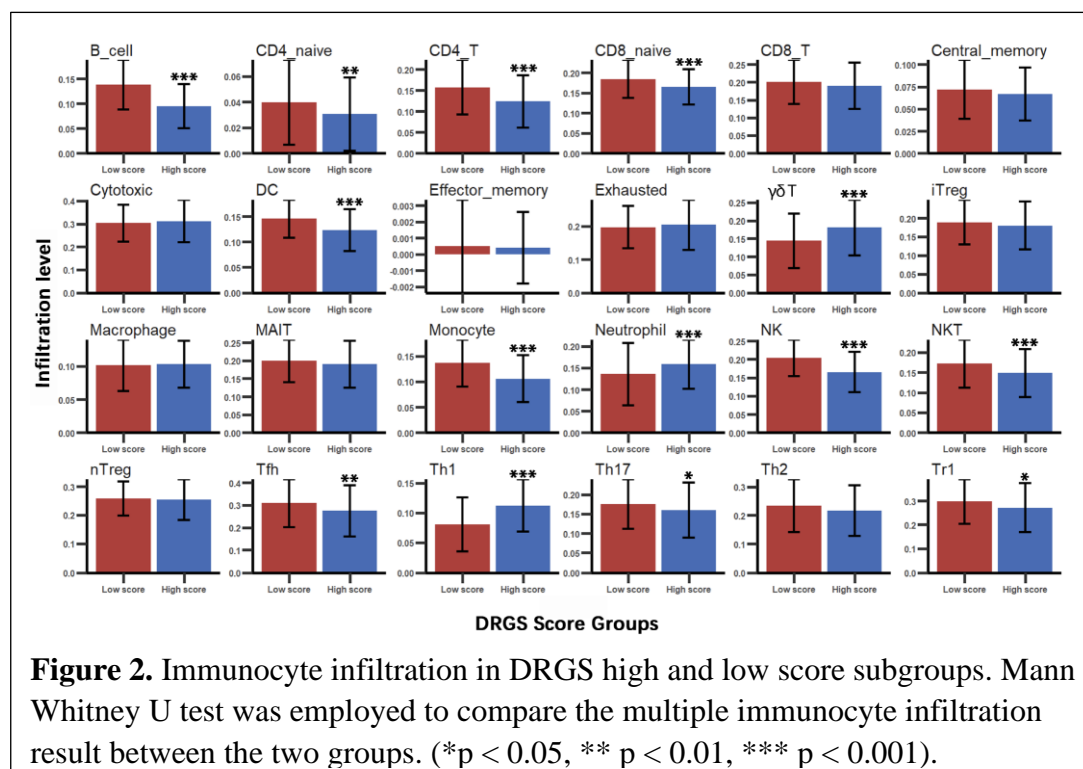
Table 1. The 15 DNA repair genes in the DRGS and their respective values

Gene symbol	logFC		adj.P.Val	
	TCGA-STAD	GSE30727	TCGA-STAD	GSE30727
<i>PRKDC</i>	1.081	1.042	1.71E-16	9.93E-04
<i>FANCI</i>	1.825	1.012	1.40E-31	3.08E-05
<i>LIG1</i>	1.012	1.009	4.80E-13	1.70E-02
<i>RECQL4</i>	2.112	1.007	5.64E-28	2.28E-03
<i>FANCA</i>	1.647	1.412	5.54E-23	1.19E-05
<i>BRCA2</i>	1.740	1.091	2.90E-22	3.30E-03
<i>FANCD2</i>	1.582	1.196	3.12E-25	1.04E-03
<i>UBE2T</i>	1.618	1.036	2.96E-21	1.01E-03
<i>POLQ</i>	1.906	1.182	5.96E-22	8.10E-04
<i>PARPBP</i>	1.386	1.255	4.54E-18	2.17E-03
<i>EXO1</i>	2.152	1.136	3.02E-32	6.97E-03
<i>XRCC2</i>	2.177	1.016	5.83E-37	1.33E-03
<i>RAD54L</i>	1.735	1.290	4.55E-23	4.36E-04
<i>EME1</i>	2.160	1.030	5.61E-32	1.05E-03
<i>FANCB</i>	1.789	1.057	9.37E-25	9.79E-04

3.2. Immune characteristics of different DRGS score groups

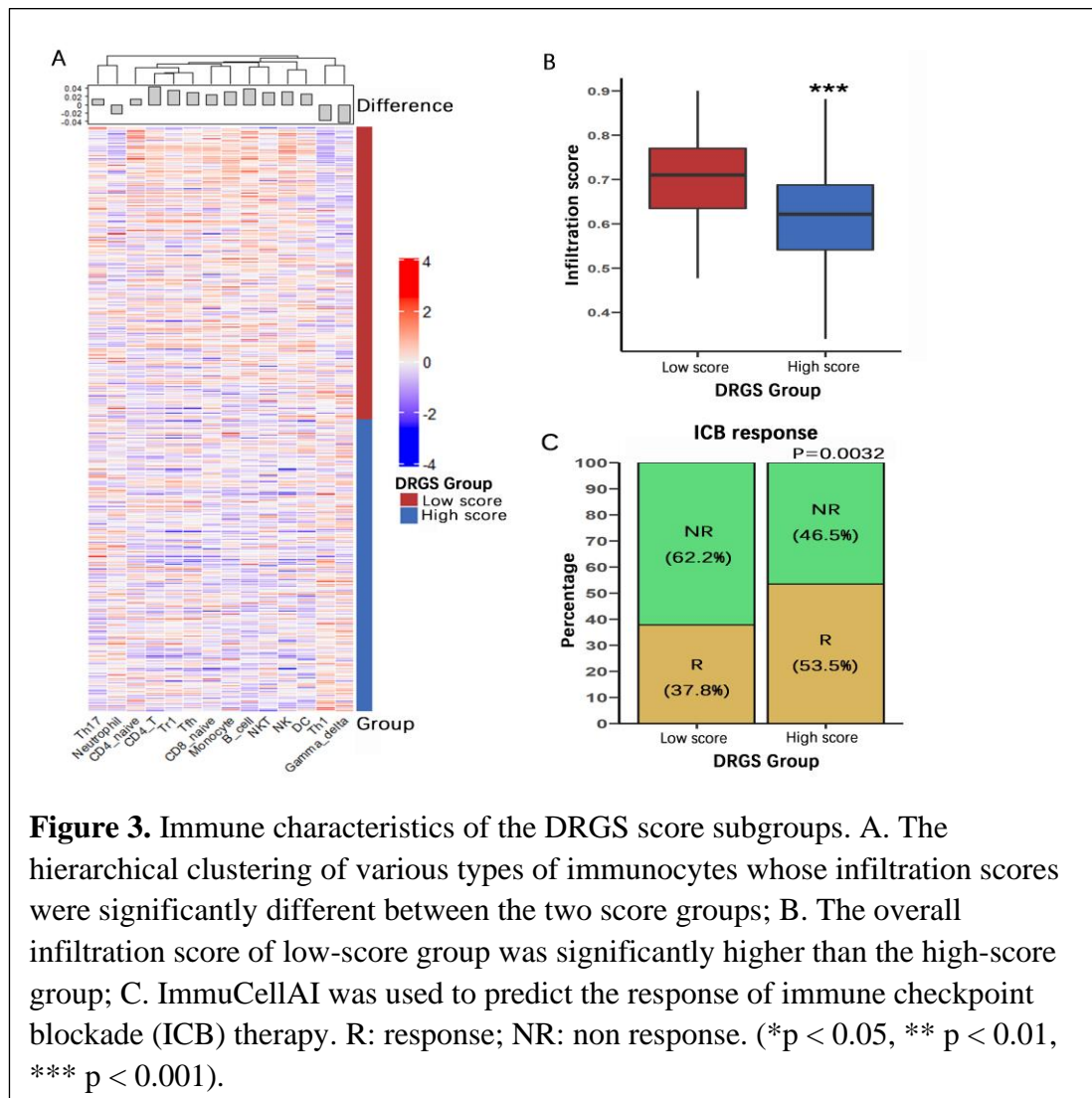
We identified immune characteristics of 375 TCGA-STAD samples by ImmuCellAI to estimate the specific fractions of 24 types of immune cells in each gastric cancer sample. To analyze the composition of immune cells in different score subgroups, we used the Mann Whitney U test to compare the distribution of immune cells in different score groups. The result showed significant differences in the infiltration rates of 14 immune cells between the two groups. We found that anti-tumor cells such as $\gamma\delta$ T cells, neutrophil cells and Th1 cells were more abundant in

the DRGS high score group ($P < 0.001$), while immunosuppressive or tumor enhancement cells such as B cell ($P < 0.001$), CD4 T cells ($P < 0.001$), monocyte ($P < 0.001$), Th17 cells ($P = 0.031$), Tfh cells ($P = 0.002$) and Tr1 ($P = 0.012$) cells were more abundant in the DRGS low score group (Figure 2A, Figure 3A). In our study, infiltration score of samples in DRGS low score group was significantly higher than those in high score group ($P < 0.001$, Mann-Whitney U test, Figure 3B). Furthermore, we used ImmuCellAI to predict the response of ICB therapy and found that there were 53.5% of patients in DRGS high score group and fewer patients (37.8%) in DRGS low score group, in response to ICB therapy ($P = 0.003$, chi-square test, Figure 3C).



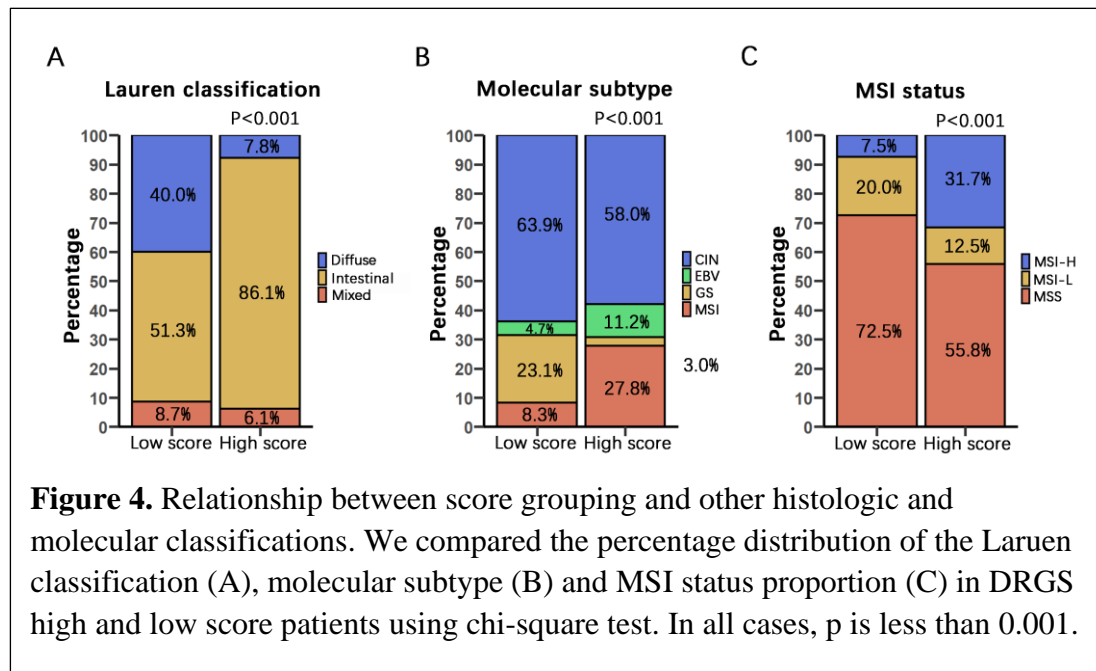
3.3. Relationship between DRGS score groups and other histologic and molecular classifications

To examine the relationship between score grouping and other histologic and molecular classification, we analyzed clinical data of 372 samples of gastric cancer patients from the TCGA-STAD database. According to the Lauren classification, gastric cancers are divided into 3 main subtypes, intestinal, diffuse and mixed. In our



study, DRGS high score group was characterized by high proportion (86.1%) of intestinal subtype samples, while diffuse and mixed subtypes samples were only 7.8% and 6.1%, respectively, while DRGS low score group comprised 51.3% intestinal, 40.0% diffuse and 8.7% mixed samples. There were more intestinal and fewer diffuse samples in the DRGS high score group ($P < 0.001$, chi-square test, Figure 4A).

TCGA proposed 4 subtypes of gastric cancer: (1) EBV-positive, (2) microsatellite instability (MSI), (3) genomically stable (GS), and (4) chromosomal instability (CIN) [19]. In our study, DRGS high score group had a higher percentage of EBV (11.2%) and MSI (27.8%) subtype samples. However, DRGS low score group had a higher percentage of GS (23.1%) and CIN (63.9%) subtype samples ($P < 0.001$, chi-square test, Figure 4B).



On the basis of the frequency of mutations within microsatellite markers, gastric cancer can be classified as MSS, MSI-H and MSI-L [20,21]. In our study, the percentage of MSI-H subtype patients (31.7%) was higher in DRGS high score group. In contrast, there were more MSS subtype patients (72.5%) in DRGS low score group ($P < 0.001$, chi-square test, Figure 4C).

3.4. Prognostic ability of the DRGS score

To investigate whether the prognostic impact of the DRGS scoring system is independent of clinical factors that could be associated with clinical outcomes, we first performed univariate Cox regression analysis on all available clinical parameters in the TCGA-STAD dataset ($n = 372$). Clinical parameters that demonstrated significant prognostic impact ($P < 0.05$ by Wald test) were selected for multivariate Cox regression analysis along with the DRGS. The result indicated that DRGS is an independent prognostic factor ($P = 0.038$, Table 2). Taking the median score as the cut-off value, patients with DRGS high score had a better OS than these with DRGS low score ($p = 0.007$, log-rank test, Figure 5A). Then, the prognostic value of scoring system was validated using the GSE26901 ($n = 109$), GSE15459 ($n = 192$) and GSE26899 ($n = 93$) GC dataset. As showed in Figure 5, the patients of GSE26901 in

the DRGS high score group had a significantly better prognosis than those in the DRGS low score group ($p = 0.011$, log-rank test, Figure 5B), consistent with the result of the TCGA dataset. However, there were no significant OS difference in GSE15459 ($P = 0.77$, Figure 5C) or GSE26899 ($P = 0.25$, Figure 5D) between the two groups.

Table 2. Univariate and multivariate Cox regression confirm the independence of the DRGS scoring system

Variables	HR	95% CI for HR		P values
		Lower	Upper	
Clinical factors				
Age	1.041	1.020	1.062	<0.001
Stage				0.190
Stage II vs. Stage I	0.916	0.319	2.629	0.871
Stage III vs. Stage I	0.830	0.210	3.280	0.790
Stage IV vs. Stage I	1.778	0.431	7.330	0.426
T grade				0.260
T2 vs. T1	4.571	0.588	35.547	0.146
T3 vs. T1	6.761	0.795	57.478	0.080
T4 vs. T1	5.603	0.636	49.379	0.121
N grade				0.079
N1 vs. N0	1.236	0.583	2.621	0.580
N2 vs. N0	1.823	0.734	4.523	0.196
N3 vs. N0	2.517	1.016	6.236	0.046
M grade	1.788	0.761	4.206	0.183
DRGS signature	0.972	0.946	0.998	0.038

Note: HR: hazard risk; CI: confidence interval.

3.5. Molecular characteristics of different DRGS score groups

To identify the relationship between TMB and the DRGS score, we analyzed the somatic mutation data of 365 stomach adenocarcinoma samples from TCGA-STAD. The result indicated that the DRGS score was positively associated with TMB ($\rho = 0.50$, $P < 0.001$, Figure 6A). TMB of samples in DRGS high score group was significantly higher than that in DRGS low score group ($P < 0.001$, Mann-Whitney U

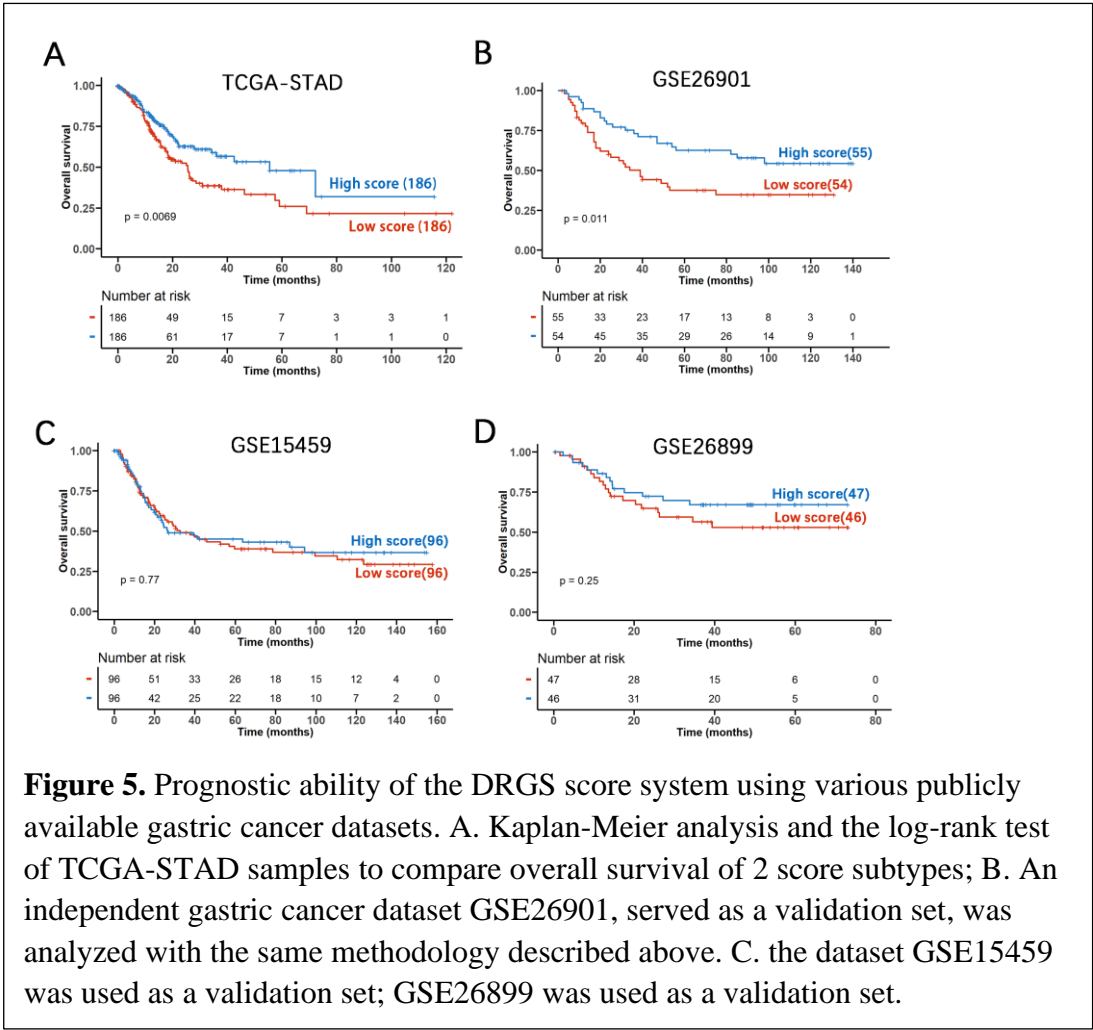
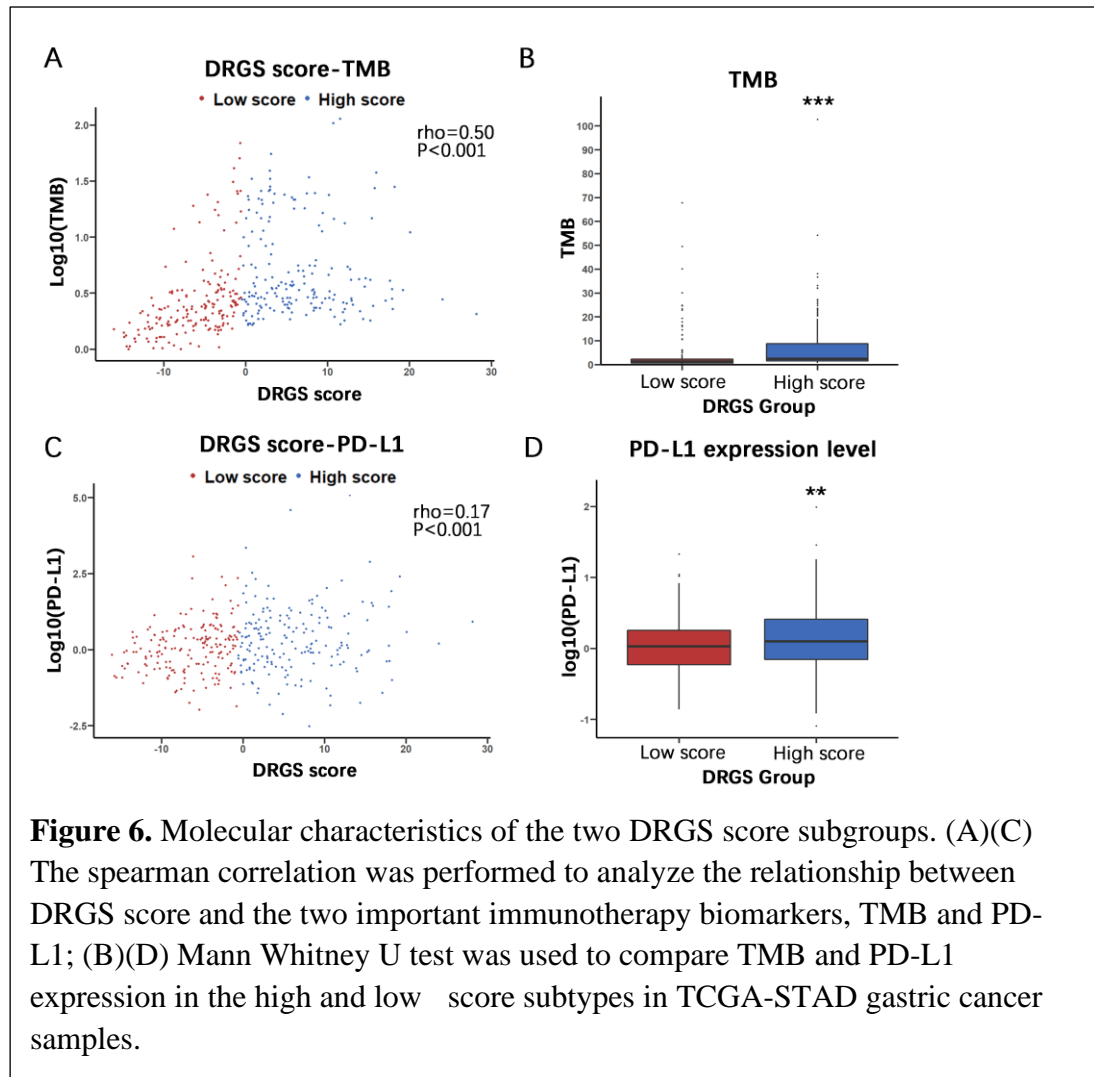


Figure 5. Prognostic ability of the DRGS score system using various publicly available gastric cancer datasets. A. Kaplan-Meier analysis and the log-rank test of TCGA-STAD samples to compare overall survival of 2 score subtypes; B. An independent gastric cancer dataset GSE26901, served as a validation set, was analyzed with the same methodology described above. C. the dataset GSE15459 was used as a validation set; GSE26899 was used as a validation set.

test, Figure 6B). In summary, all mutation events were divided into different categories, among which missense mutations accounted for the largest proportion in the classification of variation (Figure S2A, B), SNPs occurred much more frequently than insertion or deletion, and C>T was the most general of single nucleotide variants (SNV) in all STAD samples. Notably, median variant per sample of DRGS high score group (121) was higher than that of DRGS low score group (59.5). Furthermore, mutation events of the top 10 mutated genes in each sample were shown in the waterfall plot. Among DRGS high score group, the mutation frequency of TTN was the highest (59%), followed by TP53 (54%) and MUC16 (39%, Figure S2B). Among DRGS low score group, the highest mutation frequency gene was TP53 (34%), followed by TTN (33%) and MUC16 (21%, Figure S2A).

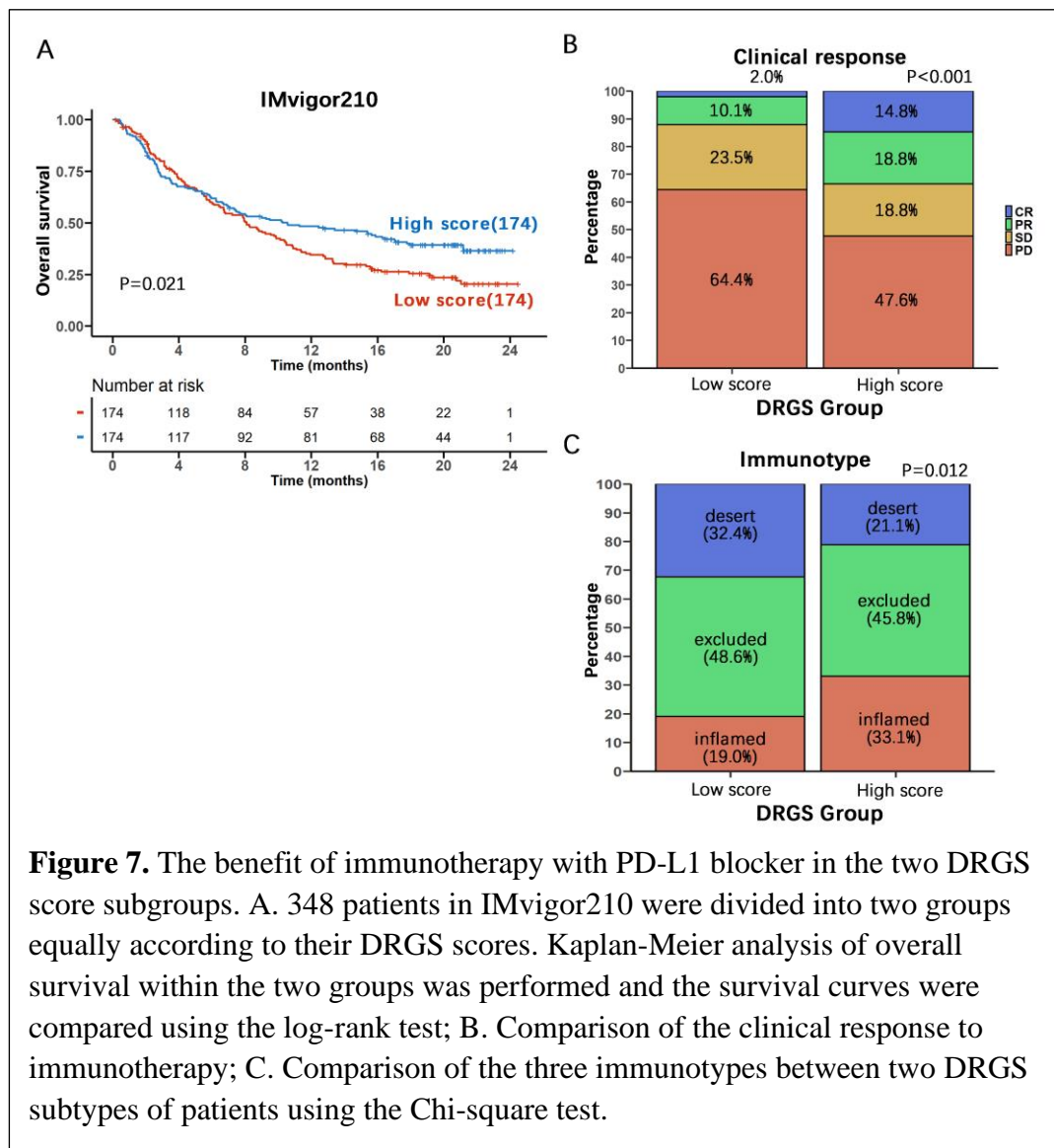


PD-L1 expression is another important biomarker of susceptibility to PD-1 blockade. We found that PD-L1 expression in DRGS high score group was significantly higher than in DRGS low score group ($P = 0.002$, Figure 6D). The DRGS score is positively correlated to PD-L1 expression ($\rho = 0.17$, $P < 0.001$, Figure 6C).

3.6. The benefit of immunotherapy with PD-L1 blocker in different DRGS score groups

In view of the association between our score system and the immune microenvironment of the tumor, as described above, we tested the ability of the DRGS score system to predict the response of patients to immunotherapy. This analysis was based on IMvigor210 cohort, a large phase 2 trial investigating the

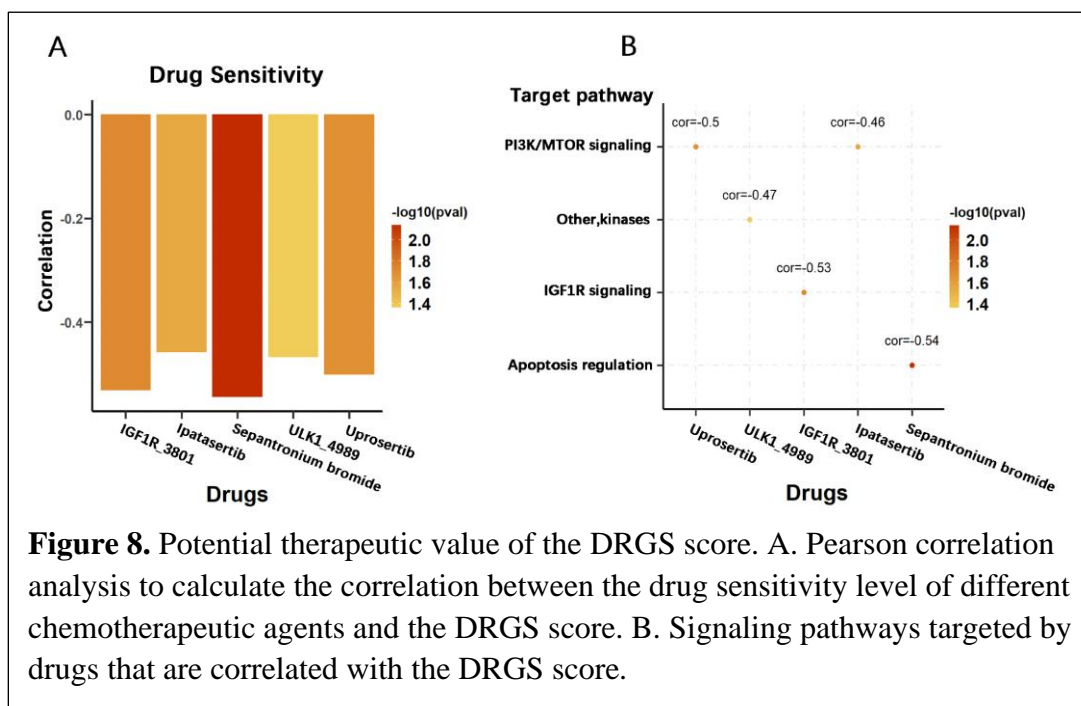
clinical activity of PD-L1 blockade with atezolizumab in metastatic urothelial carcinoma (mUC). We found that DRGS high score subtype patients exhibited prominent prolonged overall survival ($P = 0.021$, Figure 7A). Also, the 348 patients in the IMvigor210 cohort exhibited different degrees of response to anti-PD-L1 blocker. DRGS high score patients comprised 14.8% complete response (CR), 18.8% partial response (PR), 18.8% stable disease (SD) and 47.6% progressive disease (PD). DRGS low score patients comprised 2.0% CR, 10.1% PR, 23.5% SD and 64.4% PD. The chi-squared test performed between DRGS high and low score groups also showed significantly better therapeutic outcomes in DRGS high score patients ($P < 0.001$, Figure 7B).



We also analyzed the three immune subtypes of IMvigor210 (immune inflamed, immune excluded and immune desert) in the two subgroup patients. DRGS high score patients comprised 33.1% inflamed, 45.8% excluded and 21.1% desert, while DRGS low score subtype patients comprised 19.0% inflamed, 48.6% excluded and 32.4% desert. We found that the DRGS high score patients had a significantly higher percentage of “immune inflamed” tumors and a lower percentage of “immune desert” and “immune excluded” tumors ($P = 0.012$, chi-square test, Figure 7C).

3.7. Potential therapeutic value of the DRGS score

To explore the effect of the DRGS score system on drug response, we evaluated the association between the DRGS score and the response to drugs in gastric cancer cell lines. We identified 5 significantly correlated pairs between the score and the drug in the Genomics of Drug Sensitivity in Cancer (GDSC) database (<https://www.cancerrxgene.org>), and all the 5 pairs exhibited drug resistance correlated with the DRGS score, including insulin-like growth factor receptor IGF1R_3801 ($R_s = -0.53$, $p = 0.019$), AKT inhibitor ipatasertib ($R_s = -0.46$, $p = 0.028$), surviving inhibitor sepantronium bromide ($R_s = -0.54$, $p = 0.007$), ULK1 protein



kinase ULK1_4989 ($R_s = -0.47$, $p = 0.043$) and AKT inhibitor uprosertib ($R_s = -0.50$, $p = 0.021$)(Figure 8A). In addition, we found that drugs were mostly targeting PI3K/MTOR, IGF1R and Apoptosis regulation signaling pathways (Figure 8B).

4. Discussion

In the emerging immunotherapy strategies such as the ICB therapy in various types of cancer [22-24], the interaction of tumor DNA damage and repair landscape with the immune system and related therapy has recently been revealed as a complex biological process [2,13,25]. Nevertheless, many prognostic and predictive biomarkers are being evaluated clinically to identify criteria for establishing customized therapeutic approaches. This has been verified in cancers with MMR deficiency which leads to microsatellite instability (MSI) phenotype, in which ICIs were found to be higher effective [12].

In this study, taking into the known association of DNA damage and repair with immunotherapy response, we identified a 15-DNA repair gene signature (DRGS) using the gene expression data in TCGA-STAD database and developed the score system with the median score as the cut-point to divide gastric cancer patients into high and low score subtypes. We assessed the capacity of the DRGS score to predict the patient response to ICB therapy using the IMvig210 cohort. The results demonstrated that DRGS high score patients showed significantly better therapeutic outcomes compared to DRGS low score patients ($P < 0.001$), suggesting that the DRGS is a promising predictor of patient response to immune therapies.

Of the 15 genes in DRGS, six genes are associated with tolerance and repair of DNA crosslinks: five genes from the FA pathway *FANCI*, *FANCA*, *FANCD2*, *FANCB*, *BRCA2* (*FANCD1*) and *XRCC2*. This may indicate a damage response to DNA crosslinks formed in the gastric cancer, although no information is available on a possible higher level of such lesions in this tumor. The FA pathway has also been associated with cancer susceptibility and with sensitivity and resistance to

chemotherapeutic agents. Also, four genes in the DRGS are in the HR pathway for repair of DNA double-strand breaks, i.e., *XRCC2*, *EME1*, *RAD54L* and *BRCA2*. This pathway is among the most well-studied of DNA repair defects in tumor immunotherapy response [25,26]. Accumulating evidence suggests more immunogenic in HR-deficient (HRD) tumors, and more susceptible to checkpoint inhibitors [27]. It should also be noted that in our DRGS high score group, the high expression level of the repair genes does not mean an increased cellular repair capacity, rather, it could well be a damage response following the increased levels of DNA damage inside tumor cells. Therefore, a better understanding of the DNA damage in the tumor will help to know the mechanism of DRGS as a biomarker. In addition, change in gene expression levels and protein levels do not correlate well in most of the cases, mainly due to the regulation control at many different levels.

It is known that the tumor immune microenvironment has great implications in gastric cancer progression and susceptibility to immunotherapy [28]. Given that a multifaceted condition may affect gastric cancer primed for response to immunotherapy, we analyzed several important factors or potential immunotherapy response biomarkers in relation to the two DRGS score groups using TCGA-STAD, including tumor-infiltrating lymphocytes, immune subtypes, TMB, PD-L1 expression and other histologic and molecular classifications (detailed see results). In TCGA-STAD samples analyzed by ImmuCellAI, we found that anti-tumor cells including $\gamma\delta$ T cells, neutrophils and Th1 cells were more abundant in the DRGS high score samples, while several types of immunosuppressive or tumor enhancement cells were more abundant in the low-score patients. Next, we analyzed the somatic mutation and PD-L1 expression data in TCGA-STAD and found that both were positively associated with the score, suggesting that DRGS high score patients are more likely to respond to immunotherapy. We also analyzed the three immune subtypes of IMvigor210 and found that the DRGS high score patients were associated with a significantly higher percentage of “immune inflamed” subtype, which has been shown to be infiltrated by a number of subtypes of immune cells [29]. Finally, DRGS high

score group had a higher percentage of MSI-H phenotype based on TCGA subtype classifications. MSI-H is characterized by better survival [30] and a high number of microsatellite mutations [21]. Recent clinical trials have confirmed that MSI-positive gastric tumors are sensitive to ICB therapy [31]. In summary, our high-score patients are characterized by “immune inflamed”, “intestinal”, and “MSI-H” phenotypes, which indicate an abundant immune cell infiltration, high expression of PD-L1, and high number of microsatellite mutations. All of which are associated with higher sensitivity and better response to immunotherapy.

There were also limitations in this work. In the study we used the data from a different type of tumor (IMvigor210 cohort on locally advanced or metastatic urothelial bladder cancer) to evaluate the ability of the DRGS score to predict PD-L1 blockade-based immunotherapy response, as a lack of such dataset in gastric cancer. Future studies will use data from the gastric cancer patients, including those from our own hospital.

5. Conclusion

The DRGS is a promising predictor of patient response to immune-directed therapies. Further studies are needed to evaluate the clinical utility of DRGS for tailoring immunotherapy.

Funding

This work was supported by the National Natural Science Foundation of China, Grant Number: 81802388 (to Pin Wang); Natural Science Foundation from the Department of Science & Technology of Jiangsu Province, Grant Number: BK20180120 (to Pin Wang).

Author Contributions

Conceptualization, Bo Hang, Pin Wang, Xiaoping Zou and Jian-Hua Mao; Data curation, Binbin Yuan, Chengfei Jiang, Lingyan Chen, Min Chen and Shu Zhang; Formal analysis, Binbin Yuan, Chengfei Jiang, Lingyan Chen, Jinlong Cui, Yimeng Cai, Jian-Hua Mao, Bo Hang and Pin Wang; Funding acquisition, Pin Wang; Investigation, Binbin Yuan, Chengfei Jiang, Jinlong Cui, Min Chen and Shu Zhang; Methodology, Jinlong Cui and Jian-Hua Mao; Project administration, Lingyan Chen; Supervision, Pin Wang, Xiaoping Zou and Bo Hang; Validation, Binbin Yuan, Yimeng Cai and Jian-Hua Mao; Writing – original draft, Binbin Yuan, Bo Hang; Writing – review & editing, Pin Wang, Jinlong Cui, Min Chen, Shu Zhang, Yimeng Cai, Jian-Hua Mao and Xiaoping Zou.

Conflicts of Interest

The authors declare no conflict of interest.

Reference

1. Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F., Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021, 71 (3), 209-249.
2. Joshi, S. S.; Badgwell, B. D., Current treatment and recent progress in gastric cancer. *CA Cancer J Clin* 2021, 71 (3), 264-279.
3. Harada, K.; Baba, H.; Ajani, J. A., Recent trend in gastric cancer treatment in the USA. *J Cancer Metastasis Treat* 2018, 4, 18.
4. Chao, J.; Fuchs, C. S.; Shitara, K.; Tabernero, J.; Wainberg, Z. A., Pembrolizumab (pembro) in microsatellite instability-high (MSI-H) advanced

- gastric/gastroesophageal junction (G/GEJ) cancer by line of therapy. *Journal of Clinical Oncology* 2020, 38 (4_suppl), 430-430.
5. Fuchs, C. S.; Özgüroğlu, M.; Bang, Y. J.; Di Bartolomeo, M.; Mandal, M.; Ryu, M. H.; Fornaro, L.; Olesinski, T.; Caglevic, C.; Chung, H. C., et al., Pembrolizumab versus paclitaxel for previously treated PD-L1-positive advanced gastric or gastroesophageal junction cancer: 2-year update of the randomized phase 3 KEYNOTE-061 trial. *Gastric Cancer* 2021.
 6. Moehler, M.; Shitara, K.; Garrido, M.; Salman, P.; Shen, L.; Wyrwicz, L.; Yamaguchi, K.; Skoczylas, T.; Bragagnoli, A. C.; Liu, T., LBA6_PR Nivolumab (nivo) plus chemotherapy (chemo) versus chemo as first-line (1L) treatment for advanced gastric cancer/gastroesophageal junction cancer (GC/GEJC)/esophageal adenocarcinoma (EAC): First results of the CheckMate 649 study - ScienceDirect. *Annals of Oncology* 31.
 7. Friedberg EC, Walker GC, Siede W, Wood RD, Schultz RA, et al. (2006) DNA repair and mutagenesis. Washington: ASM Press.
 8. Hang, B., Repair of exocyclic DNA adducts: rings of complexity. *Bioessays* 2004, 26 (11), 1195-208.
 9. Rodríguez, A.; D'Andrea, A., Fanconi anemia pathway. *Curr Biol* 2017, 27 (18), R986-r988.
 10. Helleday, T.; Petermann, E.; Lundin, C.; Hodgson, B.; Sharma, R. A., DNA repair pathways as targets for cancer therapy. *Nat Rev Cancer* 2008, 8 (3), 193-204.
 11. Chalmers, Z. R.; Connelly, C. F.; Fabrizio, D.; Gay, L.; Ali, S. M.; Ennis, R.; Schrock, A.; Campbell, B.; Shlien, A.; Chmielecki, J., et al., Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med* 2017, 9 (1), 34.
 12. Le, D. T.; Durham, J. N.; Smith, K. N.; Wang, H.; Bartlett, B. R.; Aulakh, L. K.; Lu, S.; Kemberling, H.; Wilt, C.; Lubner, B. S., et al., Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* 2017, 357 (6349), 409-413.
 13. Bever, K. M.; Le, D. T., DNA repair defects and implications for immunotherapy. *J Clin Invest* 2018, 128 (10), 4236-4242.
 14. Gibney, G. T.; Weiner, L. M.; Atkins, M. B., Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *Lancet Oncol* 2016, 17 (12), e542-e551.
 15. Balar, A. V.; Galsky, M. D.; Rosenberg, J. E.; Powles, T.; Petrylak, D. P.; Bellmunt, J.; Loriot, Y.; Necchi, A.; Hoffman-Censits, J.; Perez-Gracia, J. L., et al., Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. *Lancet* 2017, 389 (10064), 67-76.
 16. Wood, R.; Mitchell, M.; Lindahl, T., Human DNA repair genes, 2005. *Mutation research* 2005, 577, 275-83.
 17. Miao, Y. R.; Zhang, Q.; Lei, Q.; Luo, M.; Xie, G. Y.; Wang, H.; Guo, A. Y., ImmuCellAI: A Unique Method for Comprehensive T-Cell Subsets

- Abundance Prediction and its Application in Cancer Immunotherapy. *Adv Sci (Weinh)* 2020, 7 (7), 1902880.
18. Mayakonda, A.; Lin, D. C.; Assenov, Y.; Plass, C.; Koeffler, H. P., Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* 2018, 28 (11), 1747-1756.
 19. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014, 513 (7517), 202-9.
 20. Rhyu, M. G.; Park, W. S.; Meltzer, S. J., Microsatellite instability occurs frequently in human gastric carcinoma. *Oncogene* 1994, 9 (1), 29-32.
 21. Yoon, K.; Lee, S.; Han, T. S.; Moon, S. Y.; Yun, S. M.; Kong, S. H.; Jho, S.; Choe, J.; Yu, J.; Lee, H. J., et al., Comprehensive genome- and transcriptome-wide analyses of mutations associated with microsatellite instability in Korean gastric cancers. *Genome Res* 2013, 23 (7), 1109-17.
 22. Gentles, A. J.; Newman, A. M.; Liu, C. L.; Bratman, S. V.; Feng, W.; Kim, D.; Nair, V. S.; Xu, Y.; Khuong, A.; Hoang, C. D., et al., The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med* 2015, 21 (8), 938-945.
 23. Fridman, W. H.; Pagès, F.; Sautès-Fridman, C.; Galon, J., The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer* 2012, 12 (4), 298-306.
 24. Zhang, C.; Li, D.; Yu, R.; Li, C.; Song, Y.; Chen, X.; Fan, Y.; Liu, Y.; Qu, X., Immune Landscape of Gastric Carcinoma Tumor Microenvironment Identifies a Peritoneal Relapse Relevant Immune Signature. *Front Immunol* 2021, 12, 651033.
 25. Mouw, K. W.; Goldberg, M. S.; Konstantinopoulos, P. A.; D'Andrea, A. D., DNA Damage and Repair Biomarkers of Immunotherapy Response. *Cancer Discov* 2017, 7 (7), 675-693.
 26. Strickland, K. C.; Howitt, B. E.; Shukla, S. A.; Rodig, S.; Ritterhouse, L. L.; Liu, J. F.; Garber, J. E.; Chowdhury, D.; Wu, C. J.; D'Andrea, A. D., et al., Association and prognostic significance of BRCA1/2-mutation status with neoantigen load, number of tumor-infiltrating lymphocytes and expression of PD-1/PD-L1 in high grade serous ovarian cancer. *Oncotarget* 2016, 7 (12), 13587-98.
 27. van Wilpe, S.; Tolmeijer, S. H.; Koornstra, R. H. T.; de Vries, I. J. M.; Gerritsen, W. R.; Ligtenberg, M.; Mehra, N., Homologous Recombination Repair Deficiency and Implications for Tumor Immunogenicity. *Cancers (Basel)* 2021, 13 (9).
 28. Lazăr, D. C.; Avram, M. F.; Romoșan, I.; Cornianu, M.; Tăban, S.; Goldiș, A., Prognostic significance of tumor immune microenvironment and immunotherapy: Novel insights and future perspectives in gastric cancer. *World J Gastroenterol* 2018, 24 (32), 3583-3616.
 29. Chen, D. S.; Mellman, I., Elements of cancer immunity and the cancer-immune set point. *Nature* 2017, 541 (7637), 321-330.

30. Yamamoto, H.; Perez-Piteira, J.; Yoshida, T.; Terada, M.; Itoh, F.; Imai, K.; Perucho, M., Gastric cancers of the microsatellite mutator phenotype display characteristic genetic and clinical features. *Gastroenterology* 1999, 116 (6), 1348-57.
31. Le, D. T.; Uram, J. N.; Wang, H.; Bartlett, B. R.; Kemberling, H.; Eyring, A. D.; Skora, A. D.; Lubner, B. S.; Azad, N. S.; Laheru, D., et al., PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* 2015, 372 (26), 2509-20.

Supplementary Materials

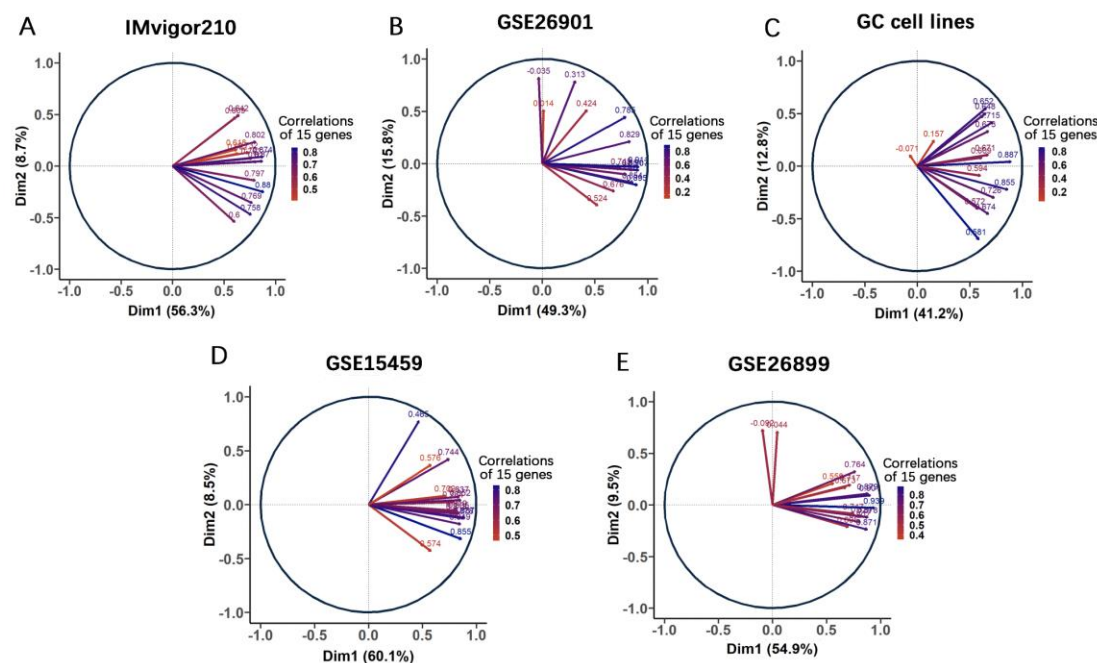


Figure S1. PCA analysis was used as a dimensionality reduction method to obtain the 1st dimension correlation coefficients of 15 genes in IMvigor210, GSE26901, GSE15459, GSE26899 and GC cell lines (from A to E).

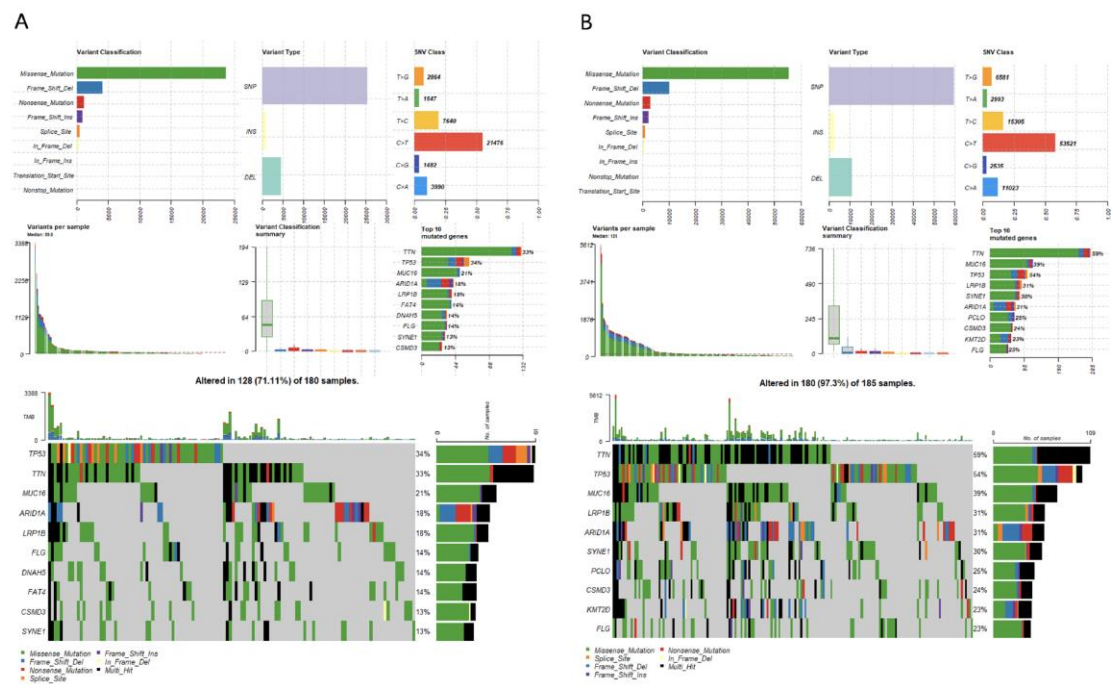


Figure S2. Overview of TCGA-STAD mutations. Statistics and visual display of mutation type, SNV classification, TMB and mutation classification in DRGS low score subtype (A) versus high score subtype (B) samples. Waterfall plot of top 10 mutation genes, which classifies the mutations of each gene in each sample, and display the results in the waterfall chart.