Am J Transl Res 2021;13(4):1952-1970
www.ajtr.org /ISSN:1943-8141/AJTR0126032

Original Article

Development and validation of a prognostic nomogram for patients with lung adenocarcinoma based on a novel 6-DNA repair-related gene signature

Minjie Chen[1], Hui Dong[2], Shengchang Deng[2], Ying Zhou[2]

[1]Queen Mary College, Nanchang University, Nanchang 330006, Jiangxi, China; [2]Departmend of Histology and Embryology, Jiangxi Medical College, Nanchang University, Nanchang 330006, Jiangxi, China

Received November 10, 2020; Accepted February 7, 2021; Epub April 15, 2021; Published April 30, 2021

Abstract: DNA repair-related genes (DRGs) have attracted much attention in the field of oncology. However, the prognostic role of DRGs and their biological function in lung adenocarcinoma (LUAD) remains rudimentary and inconclusive. In this study, 716 LUAD cases from two different cohorts were collected. Samples from The Cancer Genome Atlas (TCGA) were used as the training set, and data from Gene Expression Omnibus (GEO) datasets were used for validation. Using multivariate Cox analysis and LASSO regression, we constructed a DRG signature and used it, together with clinical indices, to develop a nomogram to predict 1-, 3-, and 5-year survival rates. We identified a six-DRG signature to estimate the survival of LUAD patients, which distinguished high-risk from low-risk patients with LUAD in both the training and validation cohorts. We also observed elevated levels of infiltrating CD4 memory activated T cells, resting NK cells, M0 and M1 macrophages, and activated mast cells in the high-risk group. Finally, a nomogram incorporating the signature and clinical parameters was superior to the American Joint Committee on Cancer (AJCC) staging system in predicting the survival of LUAD patients. The DRG prognostic signature and integrated nomogram could be a useful tool to predict prognosis in patients with LUAD.

Keywords: Lung adenocarcinoma, DNA repair genes, prognosis, nomogram, personalized prediction model

Introduction

Lung cancer is the most prevalent malignancy and the leading cause of cancer-related deaths worldwide [1]. Non-small cell lung cancer (NSCLC) accounts for approximately 85% of all lung cancers [2]. Based on the histological features of the tumor, NSCLC can be divided into three main subtypes: lung adenocarcinoma (LUAD), lung squamous cell carcinoma, and large cell carcinoma. Among these subtypes, LUAD accounts for approximately 40% of lung cancer cases [2]. Although current treatments (including surgery, chemoradiotherapy, immunotherapy, and gene-targeted therapy) prolong the survival time of patients with early-stage LUAD, nearly 10-44% of these patients still die within 5 years after the intervention [3, 4]. As such, the high rates of recurrence and distant metastasis associated with LUAD make it a major threat to patient health. Therefore, it is imperative to find new and effective therapeutics for treating LUAD. In parallel, more specific biomarkers to predict LUAD prognosis also need to be developed, which may aid personalized treatment and clinical decision-making.

DNA damage is an important factor contributing to tumorigenesis, including LUAD [5]. DNA damage takes many forms including nucleotide alterations (deletions, insertions, and substitutions), and single-and double-strand DNA breaks [6]. These processes can result in genomic instability, inactivation of tumor suppressors, or activation of oncogenes [7]. However, the outcomes of DNA damage are diverse, partly dependent on DNA repair systems in the cell [8]. It is well known that numerous genes, called DNA repair genes (DRGs), are involved in the prevention of DNA damage and they are critical for maintaining the integrity of native DNA. Recently, the use of DRGs as diagnostic or prognostic molecular biomarkers has attracted growing attention in the field of oncol-

ogy [9-12]. However, the prognostic role of DRGs and their biological function in LUAD remains rudimentary and inconclusive. In addition, there is no accurate DRG signature for the prediction of LUAD-associated survival.

Therefore, the development of new biomarkers, based on DRG signatures, may optimize the selection of patients at the highest risk of mortality and provide novel insights into gene-targeting therapy. In this study, we identified a DRG-based prognostic signature to estimate the survival of LUAD. Moreover, we also integrated this signature with clinical features of LUAD to construct a personalized prediction model.

Methods

Data acquisition

Gene expression data and the corresponding clinical data of patients with LUAD were obtained from The Cancer Genome Atlas database (TCGA) (https://portal.gdc.cancer.gov/) and Gene Expression Omnibus database (GEO) (https://www.ncbi.nlm.nih.gov/geo/). After intersecting these samples and deleting those with a follow-up time of less than 1 day, 490 patients with LUAD were obtained from TCGA database as a training set, with patient data from the GEO database (GSE31210, n = 226) used as a validation set. A total of 716 patients were enrolled for analysis.

A list of DRGs was obtained from the UALCAN database (http://ualcan.path.uab.edu/index.html) and previously published studies [13, 14]. In addition, protein expression data of the DRGs (assessed by immunohistochemistry staining) in normal and tumor lung tissues were collected from the Human Protein Atlas (HPA) database (www.proteinatlas.org). Mutation data for LUAD in TCGA database were acquired from cBioPortal Cancer Genomics (https://www.cbioportal.org).

Development and validation of DRG signature

To narrow down the DRGs associated with prediction of survival, we selected 47 overlapping DRGs that were related to overall survival (OS) from the GSE31210 and TCGA datasets, and performed univariate Cox analysis on them, with a *P*-value of less than 0.05. Then, we used least absolute shrinkage and selection operator (LASSO) Cox regression to further reduce the number of DRGs in the training set. Finally, identified DRGs were entered into a multivariate Cox regression analysis to establish a signature for predicting the survival of patients with LUAD in the training set.

The risk score of each patient was calculated using the following formula:

$$\text{Risk score} = \sum_{i=0}^{N} (\beta_i \times Exp_i)$$

Where N is the number of prognostic DRGs, Expi is the corresponding expression data of the identified DRGs, and βi is the regression coefficient derived from the LASSO Cox regression model coefficients. The median risk score was set at a cut-off to categorize patients with LUAD as low- or high-risk cohorts.

Gene set enrichment analysis

We identified molecular pathways that were differentially activated between the low- and high-risk groups using Gene Set Enrichment Analysis (GSEA) software (V4.1.0, http://software.broadinstitute.org/gsea/index.jsp). Two sets of genes, from the gene ontology (GO) (c5.all.v7.1.symbols) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (c2.cp.kegg.v7.1.symbols) databases, were evaluated in this study. For every analysis, gene set permutations were performed 1000 times. Genes with a false discovery rate (FDR) < 0.05 with a normalized enrichment score (NES) of |NES| > 1 were considered to be significantly enriched.

Estimation of immune cell type fractions

CIBERSORT (https://cibersort.stanford.edu/) is a computational program for characterizing immune cell composition based on their gene expression profiles [15]. We utilized the CIBERSORT method to estimate the relative fractions of 22 types of tumor-infiltrating immune cells based on the gene expression data between the low- and high-risk groups.

Construction and validation of the nomogram

To construct a scoring system capable of evaluating the 1-, 3-, and 5-year survival prospects of the patients, we established a nomogram scoring system based on the DRG signature and

other clinical features, using the "rms", "Hmisc", "lattice", "Formula", and "foreign" R packages. To assess the predictive power of the prognostic nomogram, we performed Kaplan-Meier, C-index, area under the receiver operating characteristic (ROC) curve (AUC), and calibration plot analyses. In addition, the clinical applicability of the nomogram was evaluated using decision curve analysis (DCA).

Statistical analysis

The expression profiles of the mRNAs are shown as raw data, and each mRNA was log2 normalized for further analysis. Kaplan-Meier analysis with a log-rank test was performed between the low- and high-risk groups. A ROC survival analysis was conducted to compare the predictive accuracy of the DRG signature with respect to patients with LUAD. All statistical analyses were performed using R version 3.6.2 (https://www.rproject.org/). *P*-values of less than 0.05 were considered statistically significant.

Results

Development of a DRG signature for LUAD

To screen prognostic DRGs and to construct a prognostic signature, the mRNA levels of each DRG were subjected to univariable Cox regression analysis. Eventually, 93 DRGs from TCGA and 169 DRGs from the GSE31210 databases were found to be significantly associated with OS ($P < 0.05$, Tables S1 and S2). Eventually, 47 overlapping prognostic DRGs was selected for subsequent studies (**Figure 1A**).

To further refine the list of DRGs capable of predicting LUAD survival, the 47 previously selected DRGs were included in the LASSO regression analysis conducted on the training set data, and 9 DRGs were identified (**Figure 1B** and **1C**). These genes were then subjected to multivariate Cox regression analysis to establish a risk signature, and 6 DRGs, including BTG anti-proliferation factor 2 (*BTG2*), damage specific DNA binding protein 1 (*DDB1*), DNA methyltransferase 1 associated protein 1 (*DMAP1*), PTTG1 regulator of sister chromatid separation, securing (*PTTG1*), SMAD family member 3 (*SMAD3*), and single-strand-selective monofunctional uracil-DNA glycosylase 1 (*SMUG1*), were finally selected as predictors of OS in patients with LUAD (**Figure 1D**). Risk scores were calculated using the following formula derived from the regression coefficients and DRG expression levels (see Materials and Methods, section 2.2): Risk score = (0.02374 × DDB1 mRNA level) + (-0.00421 × BTG2 mRNA level) + (-0.07563 × DMAP1 mRNA level) + (0.02302 × PTTG1 mRNA level) + (0.03270 × SMAD3 mRNA level) + (0.04118 × SMUG1 mRNA level).

DRG expression and alterations in LUAD

To further investigate the expression of DRGs included in the gene signature, we analyzed the mRNA levels of these genes in normal and tumor tissues in the training set cohort. The mRNA levels of *DDB1*, *PTTG1*, *DMA*P1, and SMUG1 were significantly increased in LUAD, whereas BTG2 mRNA levels were significantly decreased (**Figure 2A**). However, there was no significant difference in SMAD3 mRNA levels between the LUAD and control samples. We also analyzed the protein levels of the DRGs using the HPA database, which holds immunohistochemistry results. Our analyses show that the protein levels of the DRGs matched their mRNA expression levels (**Figure 2B**).

To investigate whether genetic alterations of the 6 DRGs play significant roles in LUAD, we used the cBioPortal online tool to search for any alterations in these genes. Among the datasets analyzed, the frequency of gene alterations, including amplifications, deep deletions, missense mutations, and truncating mutations, ranged from 0.9% to 4%. Of these alterations, amplifications and missense mutations were the most commonly observed changes (**Figure 2C**).

Functional enrichment analysis

Next, we explored the potential functional mechanisms leading to a differential prognosis for LUAD with this gene signature. To this end, we conducted separate GSEA analyses for each DRG, based upon stratification of their respective risk scores, using both GO term and KEGG pathway enrichment analyses. The top 5 upregulated GO terms were ADP metabolic process, cadherin binding, cell cycle G2M phase transition, mitotic nuclear division, and nucleotide phosphorylation (**Figure 3A**). The results from the KEGG pathway analysis revealed that

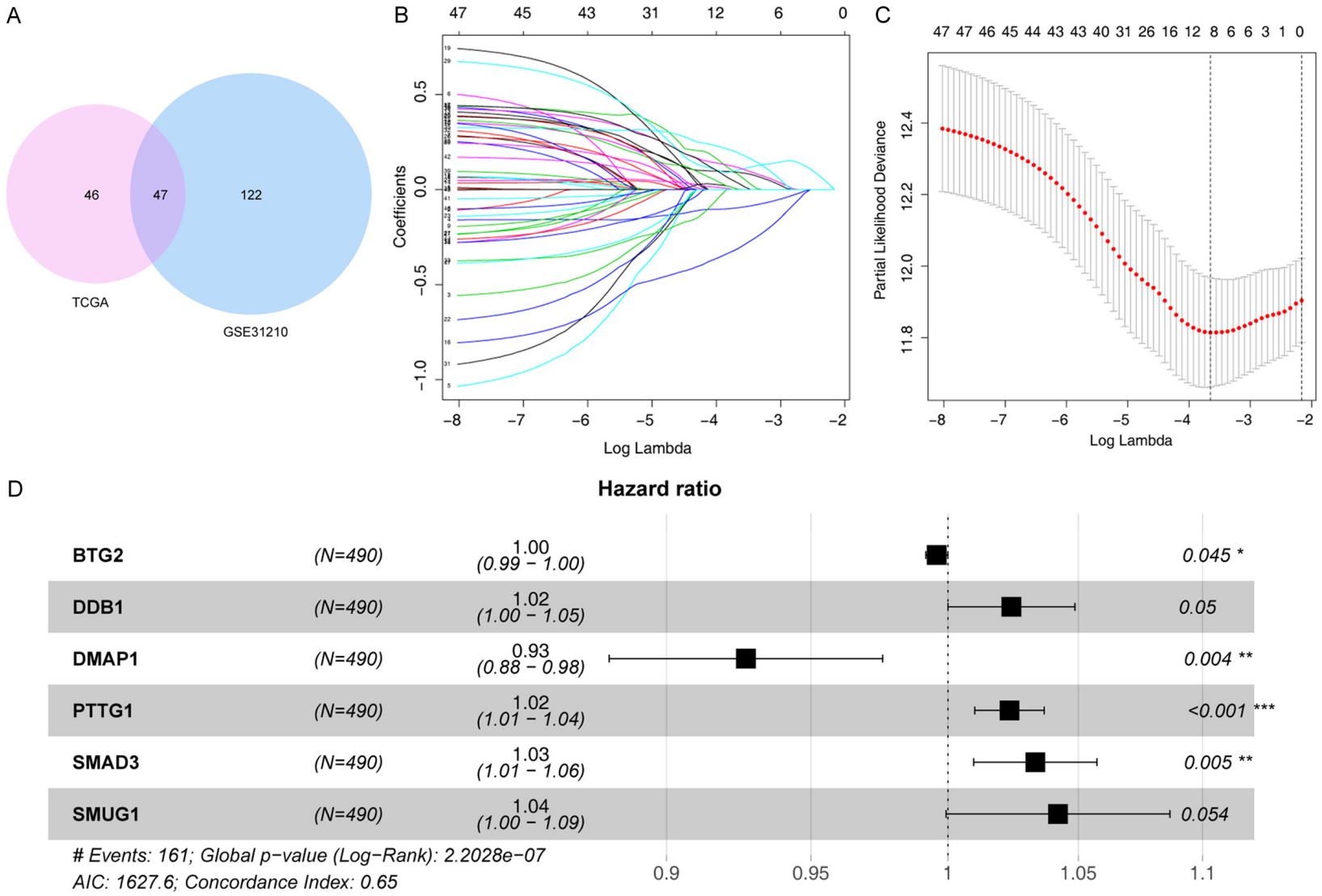


Figure 1. Construction of the DRG signature. A. Screen of 47 DRGs associated with OS based on the overlapping genes from TCGA and GSE31210. B. Selection of the optimal parameter in the least absolute shrinkage and selection operator (LASSO) regression with tenfold cross-validation. C. LASSO coefficient profiles of the candidate prognosis-related DRGs. D. DRG signature constructed by multivariate Cox regression analysis. *P < 0.05, **P < 0.01, ***P < 0.001.

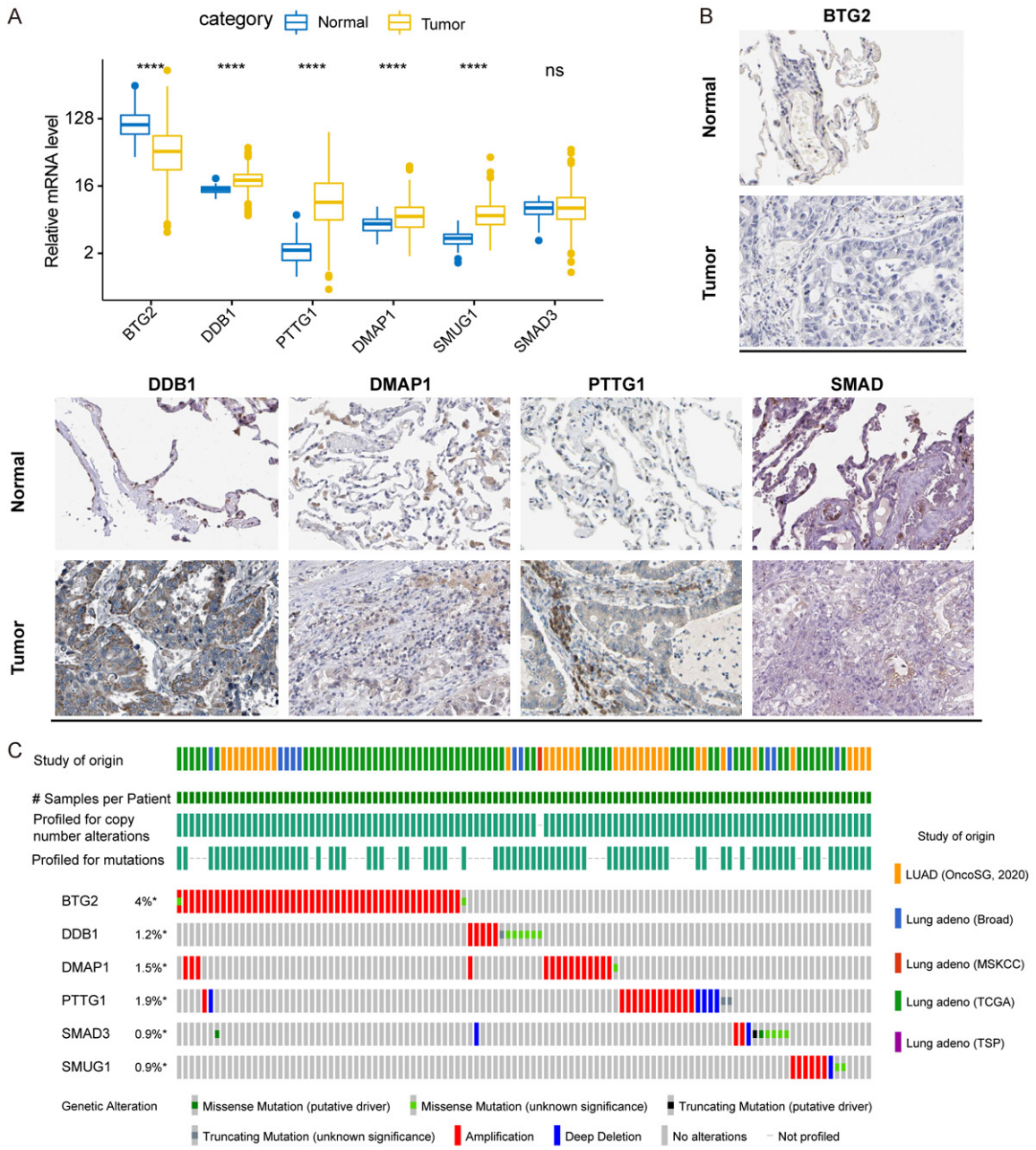


Figure 2. The mRNA, protein, and mutation landscape of the identified 6 DRG included in the DRG signature in LUAD. A. Difference in mRNA levels of identified DRGs between normal and tumor tissues in TCGA training set. B. Difference in protein levels of identified DRGs between normal and tumor tissues by using the Human Protein Atlas database. C. Mutational profiles of LUAD by using cBioPortal Cancer Genomics database.

the top 5 enriched pathways were cell cycle, DNA replication, p53 signaling pathway, proteasome, and RNA degradation (**Figure 3B**).

Prognostic value of the DRG signature in the training set

The risk scores for each case were calculated using the DRG signature formula mentioned above. Based on the median risk score, patients with LUAD in the training set were divided into low-risk (n = 245) and high-risk groups (n = 245) (**Figure 4A**). The survival status and survival times of patients with LUAD, ranked by risk score, are shown in **Figure 4B.** Additionally, a heatmap showing the expression profiles of the 6 DRGs was plotted (**Figure 4C**). The results of Kaplan-Meier survival analysis demonstrated that patients in the high-risk group had a worse OS than those in the low-risk group

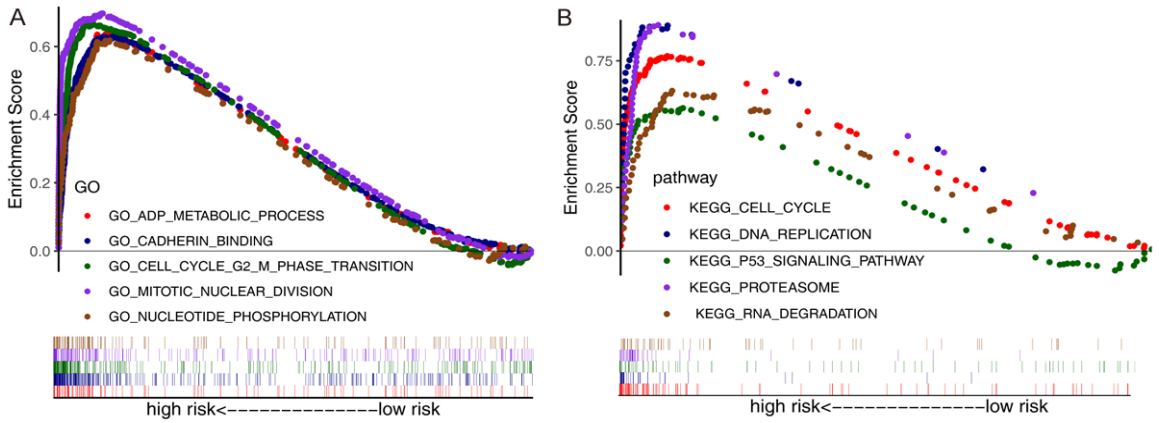


Figure 3. Gene set enrichment analysis between low- and high-risk groups. A. Top 5 representative gene ontology (GO) annotation terms in the high-risk group. B. Top 5 representative Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in the high-risk group.

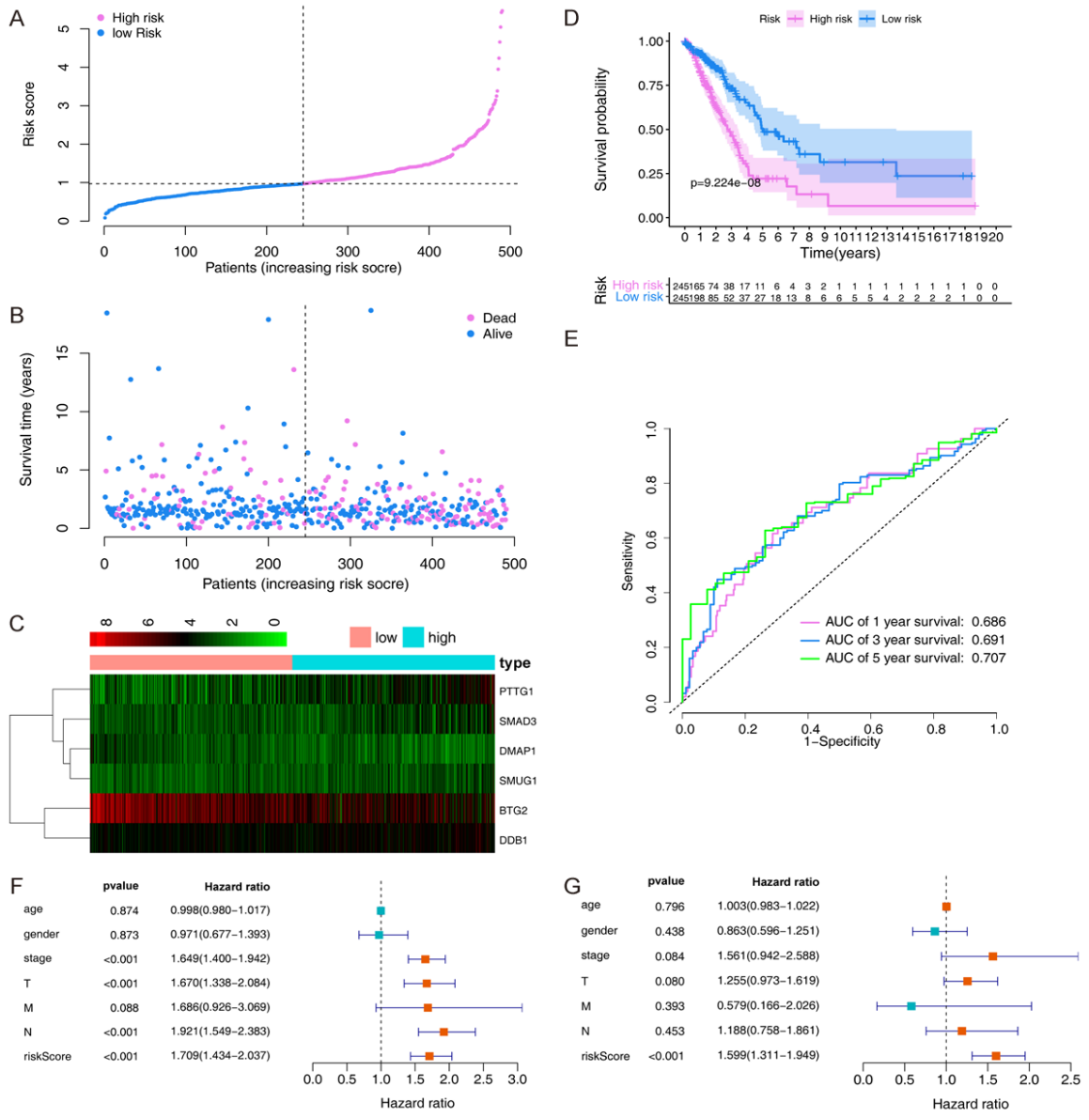


Figure 4. Prognostic value of identified DRG signature in TCGA training set. A. Rank of risk signature and score distribution. B. Survival status and times of patients with LUAD ranked by risk scores. C. Heatmap of the expression profiles of the 6 DRGs. D. Kaplan-Meier survival analysis between low- and high-risk groups. E. 1-, 3-, and 5-year OS ROC curves to assess the predictive capacity of the prognostic DRG signature. F. Univariate. G. Multivariate Cox regression analysis of the DRG signature and clinical features of LUAD.

(**Figure 4D**). We then conducted 1-, 3-, and 5-year ROC curve analyses to assess the predictive capacity of the prognostic DRG signature. Our calculations showed that the AUCs for 1-, 3-, and 5-year OS predictions were 0.686, 0.691, and 0.707, respectively (**Figure 4E**).

To investigate whether the DRG signature is an independent risk factor for the OS of patients with LUAD, the DRG signature and clinical features (including age, gender, and stage) of patients in the training set were analyzed using the univariate and multivariate Cox proportional hazard regression models. Univariate Cox regression revealed that stage, tumor (T), node (N), and DRG signature were OS-related factors (**Figure 4F**). The multivariate Cox regression analysis revealed that the DRG signature was an independent risk factor for LUAD (**Figure 4G**). To further validate the prognostic value of the DRG signature in various demographic and clinical characteristics, we performed subgroup analysis of TCGA training set. The results suggested that our signature was useful in most of the subgroups (**Table 1**). **Figure 5** also shows the Kaplan-Meier subgroup analysis in which the DRG signature can distinguish the distinct prognosis associated with each subgroup.

Validation of the DRG signature and correlation with clinicopathological characteristics

We further validated the prediction ability of the DRG signature using LUAD samples from the validation set (GSE31210). As described above for the training set, the risk scores of patients with LUAD in the GSE31210 dataset were calculated using the DRG signature formula, and the 226 patients in this cohort were divided into low-risk (n = 113) and high-risk groups (n = 113) based on the median risk score (**Figure 6A**). A heatmap illustrating the expression profiles of the 6 DRGs in the validation set is shown in **Figure 6B**. The survival status (OS and relapse-free survival [RFS]) and survival times of patients with LUAD ranked by risk scores are shown in **Figure 6C** and **6D**. The Kaplan-Meier survival analysis demonstrated that patients in the high-risk group had a worse OS and RFS than those in the low-risk group (**Figure 6E**, **6F**). When we performed 1-, 3-, and 5-year ROC curve analyses to assess the predictive capacity of the prognostic DRG signature, our results showed that the AUCs for 1-, 3-, and 5-year OS predictions were 0.784, 0.727, and 0.8, respectively (**Figure 6G**). The AUCs for 1-, 3-, and 5-year RFS predictions were 0.72, 0.717, and 0.744, respectively (**Figure 6H**). To further validate the prognostic value of the DRG signature in another dataset, we performed subgroup analysis on the GSE31210 validation set. The results suggested that our signature was useful in most of the subgroups (**Table 2**).

We also examined whether there are correlations between the clinicopathological characteristics and DRG signatures in patients with LUAD. Clinical and demographic features, including age, sex, pathological tumor/node/metastasis (TNM) stage, and pathological tumor stage, were analyzed in the training cohort, and the relationships between the screened DRGs and clinical indices were explored. The results suggested that there is a differential expression of *PTTG1* and *DMAP1* in patients with various clinical features (**Figure 7**).

Immune landscape in patients with LUAD

To investigate the relationship between the two risk groups and the immune infiltrate, we next analyzed 22 immune cell phenotypes in the training set using CIBERSORT. **Figure 8A** shows the immune cell type percentages in the low- and high-risk groups from the training dataset. High-risk patients were found to be associated with significantly higher levels of CD4 memory activated T cells, resting NK cells, M0 and M1 macrophages, and activated mast cells. In contrast, low-risk patients exhibited higher levels of memory B cells, plasma cells, T cells CD4 memory resting, follicular helper T cells, activated NK cells, monocytes, resting and activated dendritic cells, and resting mast cells (**Figure 8B**). **Figure 8C** also demonstrates a correlation between the immune cell types.

Table 1. Prognostic roles of the DRGs signature with different demographic and clinical characteristics in TCGA training set

Characteristics	Number (high-/low-risk group)	%	HR (95% CI)	*P*-value
Age (years)				
≥ 65	134/137	55.3%	2.139 (1.395-3.280)	0.000
< 65	111/108	44.7%	2.733 (1.639-4.557)	0.000
Sex				
Female	126/140	54.3%	1.896 (1.219-2.951)	0.005
Male	119/105	45.7%	3.00 (1.830-4.919)	0.000
Stage				
I	106/105	53.3%	1.529 (0.892-2.619)	0.122
II	67/50	23.9%	4.192 (2.009-8.748)	0.000
III	56/23	16.1%	1.439 (0.720-2.876)	0.303
IV	13/12	5.1%	1.585 (0.528-4.757)	0.412
NA	3/5	1.6%	-	-
T stage				
T1	65/101	21.6%	2.489 (1.305-4.747)	0.006
T2	138/120	52.7%	2.077 (1.345-3.208)	0.001
T3	29/16	9.2%	2.457(0.803-7.520)	0.115
T4	12/6	3.7%	2.389 (0.494-11.566)	0.279
NA	1/2	0.6%	-	-
M stage				
M0	169/153	65.7%	2.494 (1.662-3.744)	0.000
M1	13/11	4.9%	1.840 (0.564-5.998)	0.312
NA	63/81	29.4%	-	-
N stage				
N0	143/174	64.7%	1.828 (1.148-2.909)	0.011
N1	53/39	18.8%	4.784 (2.203-10.391)	0.000
N2	45/23	13.9%	1.279 (0.638-2.562)	0.488
N3	2/0	0.4%	-	-
NA	2/9	2.2%	-	-

NA, not available; HR, hazard ratio; CI, confidence interval.

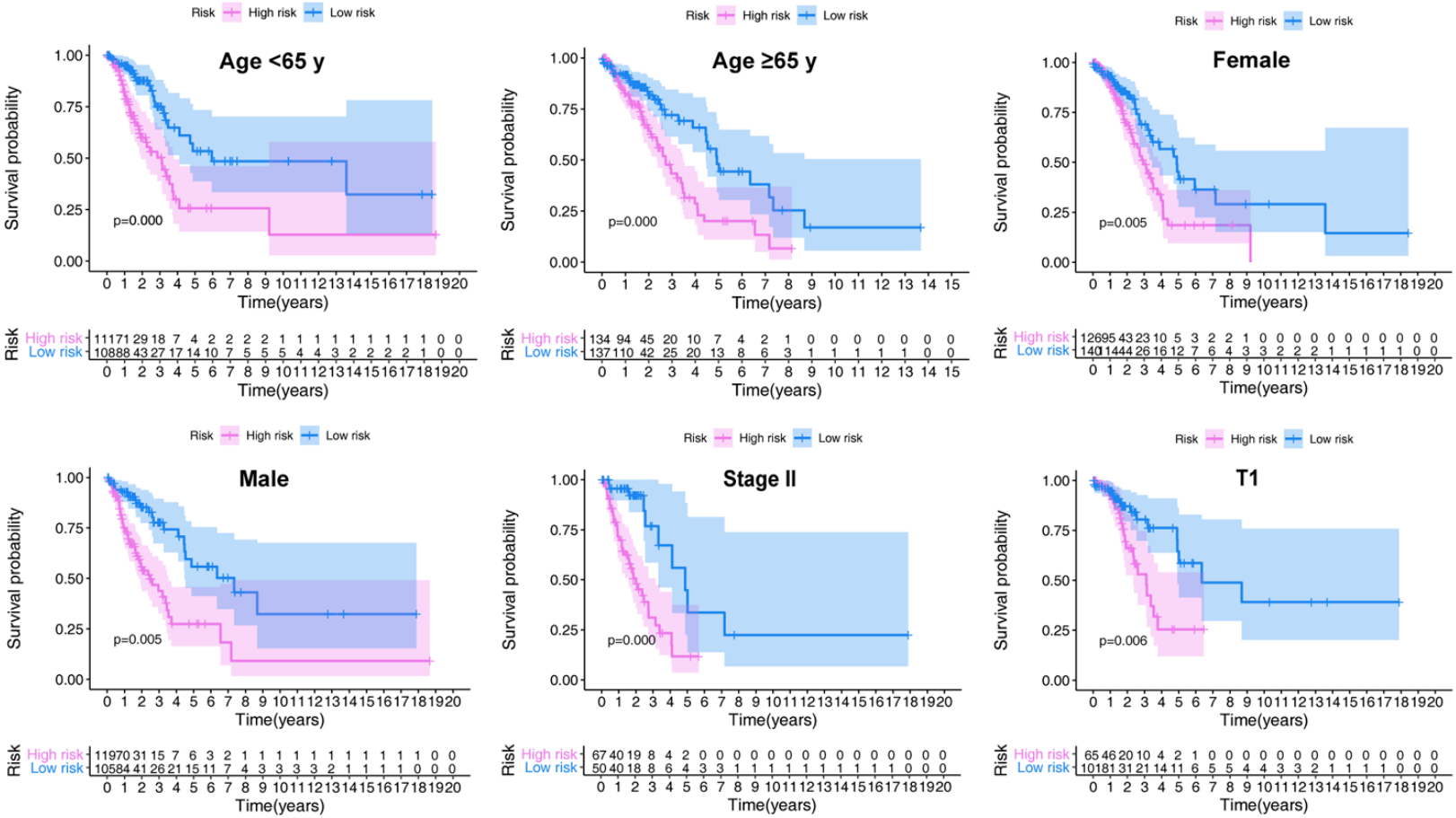
Construction and validation of DRG nomogram

In consideration of the clinical features that are commonly assessed in clinical practice, we combined data on age, stage, gender, and DRG signature to establish a nomogram to predict survival probability at 1-, 3-, and 5-years, based on data from TCGA training set (**Figure 9A**). We then conducted the 1-, 3-, and 5-year ROC curve analyses to assess the predictive capacity of the nomogram. We observed that the AUCs for 1-, 3-, and 5-year OS predictions were 0.755, 0.751, and 0.761, respectively (**Figure 9B**), and the AUCs for 1-, 3-, and 5-year RFS predictions were 0.901, 0.717, and 0.803, respectively (**Figure 9C**).

To further evaluate the predictive performance and clinical usefulness of the prognostic nomogram, calibration curves and DCA were performed. Calibration plots based on TCGA training set showed that the nomogram could accurately predict 1-, 3-, and 5-year OS and RFS (**Figure 10A**). In addition, DCA showed that the nomogram benefit was significantly higher than the extreme curve, and the nomogram is also higher than the cancer Stage (**Figure 10B**). Collectively, these results show good performance of the predictive model as a clinical prognostic tool.

Discussion

The accurate prediction of LUAD prognosis is particularly important in decision-making regarding more aggressive treatment, earlier intervention, and delayed tumor progression [16]. However, reliable prognostic gene bio-


Risk High risk Low risk
Age <65 y
Survival probability
p=0.000
Time(years)
Risk
High risk 111 71 29 18 7 4 2 2 2 2 1 1 1 1 1 1 1 1 0 0
Low risk 108 88 43 27 17 14 10 7 5 5 5 4 4 3 2 2 2 2 1 0 0
Age ≥65 y
p=0.000
High risk 134 94 45 20 10 7 4 2 1 0 0 0 0 0 0 0
Low risk 137 110 42 25 20 13 8 6 3 1 1 1 1 1 0 0
Female
p=0.005
High risk 126 95 43 23 10 5 3 2 2 1 0 0 0 0 0 0 0 0 0 0
Low risk 140 114 44 26 16 12 7 6 4 3 3 2 2 2 1 1 1 1 1 0 0
Male
p=0.005
High risk 119 70 31 15 7 6 3 2 1 1 1 1 1 1 1 1 1 1 0 0
Low risk 105 84 41 26 21 15 11 7 4 3 3 3 3 2 1 1 1 1 0 0 0
Stage II
p=0.000
High risk 67 40 19 8 4 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Low risk 50 40 18 8 6 4 3 3 1 1 1 1 1 1 1 1 1 1 0 0 0
T1
p=0.006
High risk 65 46 20 10 4 2 1 0 0 0 0 0 0 0 0 0 0 0 0 0
Low risk 101 81 31 21 14 11 6 5 5 4 4 3 3 2 1 1 1 1 0 0 0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

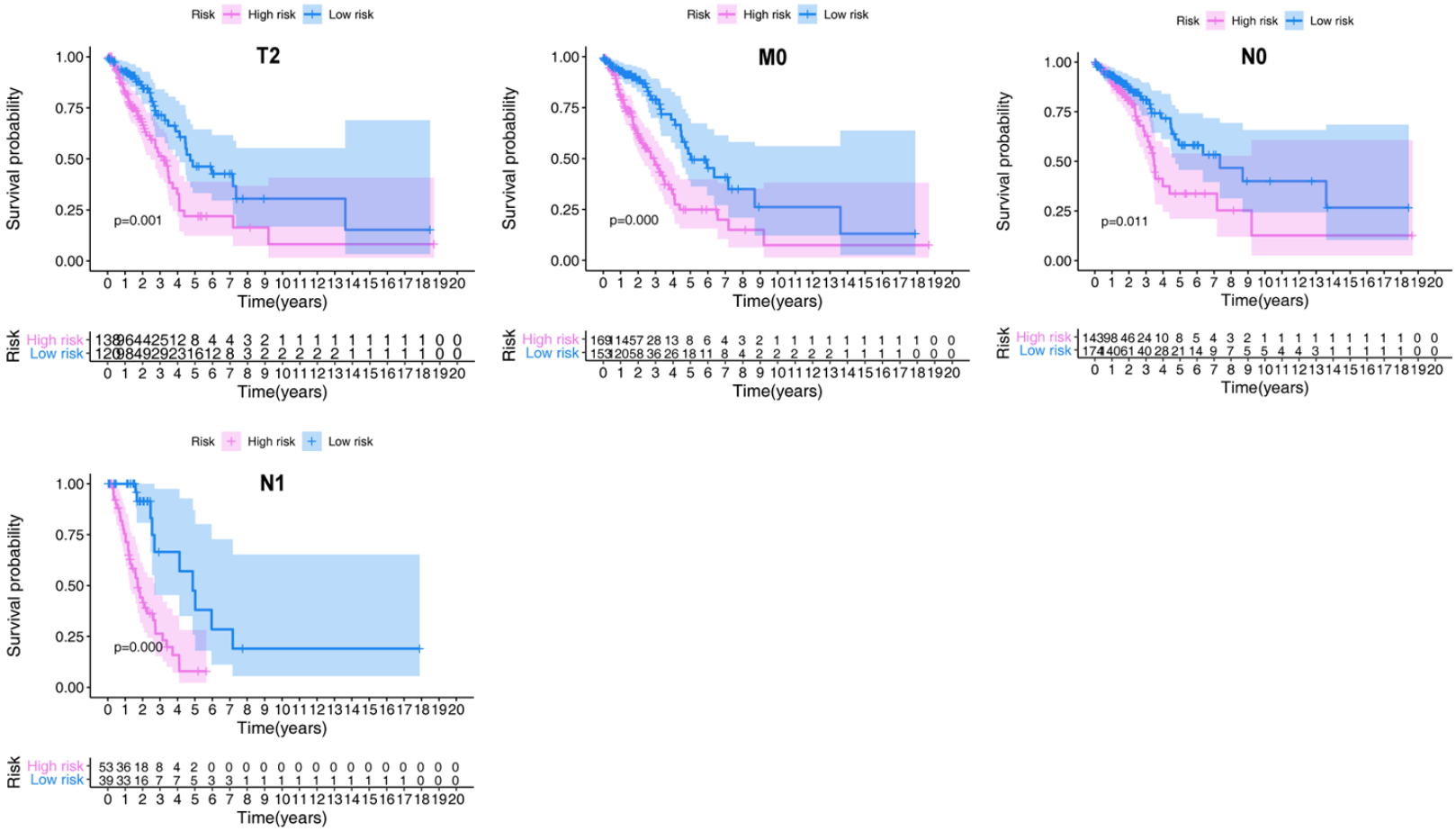


Figure 5. Confirmation of the DRG signature via stratification of patients based on specific demographic and clinical features in the TCGA training set. The different subgroups including age < 65 y, age ≥ 65 y, female, male, stage II, T1, T2, M0, N0, and N1.

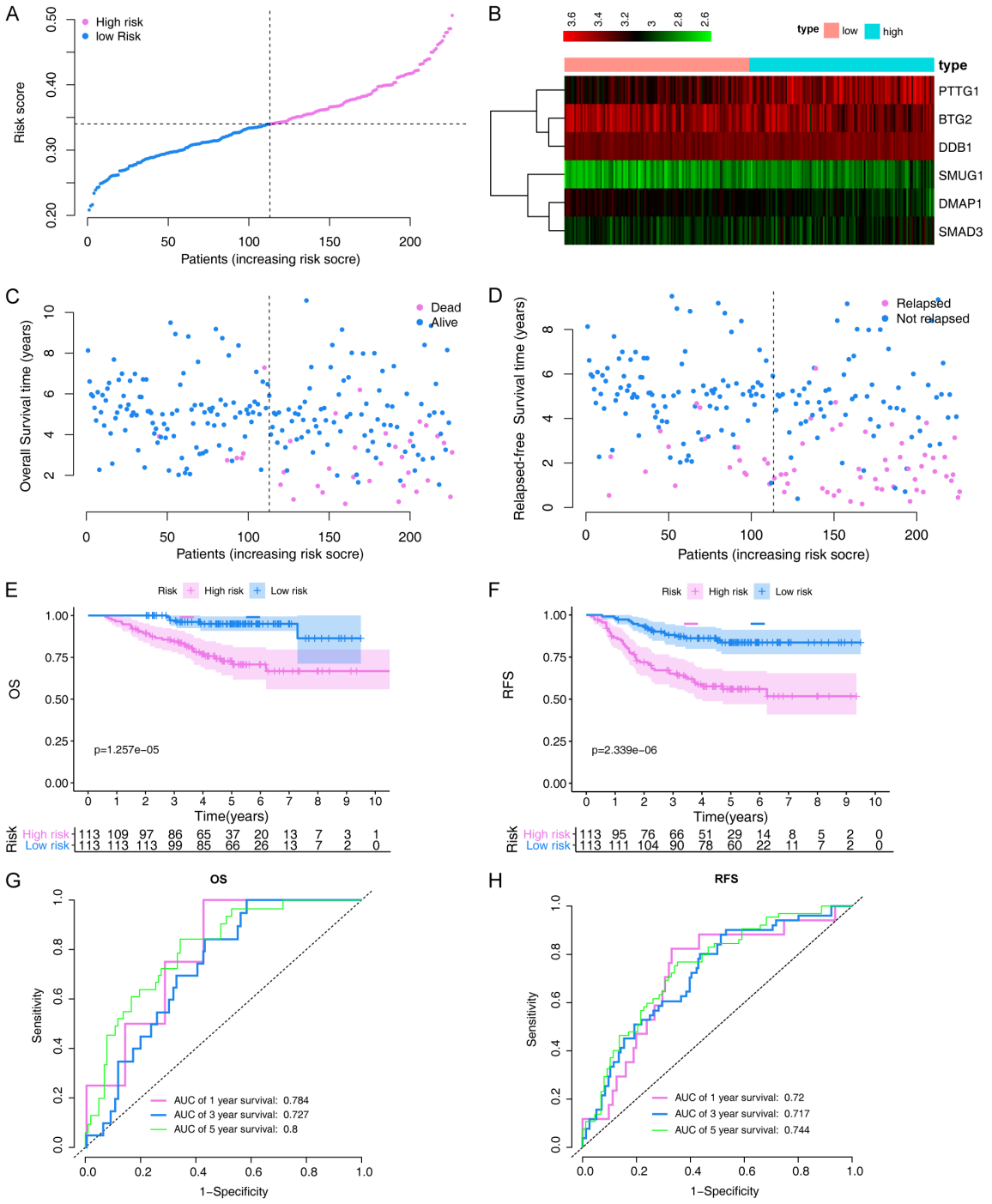


Figure 6. Validation of the DRG signature in GSE31210 validation set. (A) Rank of risk signature and score distribution. (B) Heatmap of the expression profiles of the 6 DRGs. (C) OS status and times, and (D) RFS status and times, of LUAD patients ranked by risk scores. Kaplan-Meier analysis of (E) OS, and (F) RFS, between low- and high-risk groups. 1-, 3-, and 5-year (G) OS, and (H) RFS, ROC curves to assess the predictive capacity of the prognostic DRG signature.

markers for LUAD are still very rare [17]. Recently, an increasing number of studies have begun focusing on research into DNA damage, and there are strong links between malignant tumors and DNA damage [18]. As DRGs play important roles in maintaining the integrity of

Table 2. Prognostic roles of the DRGs signature with different demographic and clinical characteristics in GSE31210 validation set

Characteristics	Number (high-/low-risk group)	%	OS		RFS	
			HR (95% CI)	*P*-value	HR (95% CI)	*P*-value
Age (years)						
≥ 65	35/27	72.6%	3.122 (0.868-11.227)	0.081	3.072 (1.133-8.333)	0.027
< 65	78/86	27.4%	7.920 (2.330-26.924)	0.001	3.593 (1.836-7.032)	0.000
Sex						
Female	54/67	53.5%	2.133 (0.774-5.878)	0.143	2.396 (1.171-4.902)	0.017
Male	59/46	46.5%	64.543 (1.597-2609.022)	0.027	5.725 (2.197-14.921)	0.000
Smoking status						
Ever smoker	63/48	49.1%	8.107 (1.879-34.966)	0.005	4.533 (1.883-11.007)	0.001
Never smoker	50/65	50.9%	4.101 (1.303-12.908)	0.016	2.738 (1.301-5.763)	0.008
Stage						
I	69/99	74.3%	5.051 (1.645-15.507)	0.005	3.039 (1.546-5.974)	0.001
II	44/14	25.7%	3.244 (0.744-14.151)	0.117	(2.478 (0.856-7.178)	0.094
Mutation						
ALK fusion	8/3	4.9%	34.288 (0.000-14481016.11)	0.593	35.919 (0.000-11933069.55)	0.581
EGFR mutation	46/81	30.1%	4.540 (1.424-14.478)	0.011	3.068 (1.477-6.372)	0.003
KRAS mutation	17/3	56.2%	26.101 (0.000-26324973)	0.644	27.725 (0.004-214550.001)	0.467
Wild-type *EGFR/KRAS/ALK*	42/26	8.8%	6.333 (1.428-28.091)	0.015	3.744 (1.403-9.991)	0.008

HR, hazard ratio; CI, confidence interval.

native DNA, these genes affect the development, progression, and relapse of various types of tumors [19-21]. Furthermore, the predictive power of a multi-gene-based signature to predict biological characteristics and prognosis is considered to be significantly higher compared to the use of a single marker [22, 23]. Therefore, in this study, we focused upon DRGs and constructed a prognostic signature based on data from patients with LUAD. To the best of our knowledge, this is the first study to focus on DRGs associated with LUAD patient survival.

In the present study, we systematically investigated the implications of DRG expression in LUAD prognosis. By analyzing the RNA seq data of TCGA-LUAD patients, we obtained 93 DRGs, which were associated with the OS of patients with LUAD. A prognostic signature was constructed by LASSO and multivariate Cox analyses, based on six prognosis-related DRGs, namely *BTG2*, *DDB1*, *DMAP1*, *PTTG1*, *SMAD3*, and *SMUG1*. After dividing the LUAD patients into two categories according to the median risk score, Kaplan-Meier analysis was performed both in the training and validation sets, and the results suggested that high-risk patients have a worse OS and RFS than the low-risk group. Altogether, our analyses suggest that the DRG signature may be powerful tool in forecasting the survival of patients with LUAD.

DNA damage often triggers an immune response in the tumor microenvironment [24]. Understanding the nature of the immune cell types in the tumor immune microenvironment is important in administering effective therapies for LUAD. In our study, we investigated the difference in immune infiltrate fractions between the two risk groups. We determined that high-risk patients were associated with significantly higher levels of CD4 memory activated T cells, resting NK cells, M0 and M1 macrophages, and activated mast cells. However, lower levels of memory B cells, plasma cells, CD4 memory resting T cells, follicular helper T cells, activated NK cells, monocytes, resting and activated dendritic cells, and resting mast cells were also observed in the high-risk group. Therefore, targeting these DRGs may alter the tumor microenvironment and immune responses. However, the specific mechanisms underlying the variation of the tumor immune microenvironment in LUAD require further research.

To date, many studies have developed models based on sequencing data and clinical features to stratify LUAD patients [25]. However, almost none of them have been applied in clinical practice. Nomograms, an easy-to-use tool for predicting prognosis and disease incidence, are currently widely used for predicting the survival of patients with malignant tumors [26]. In the present study, we also constructed a compre-

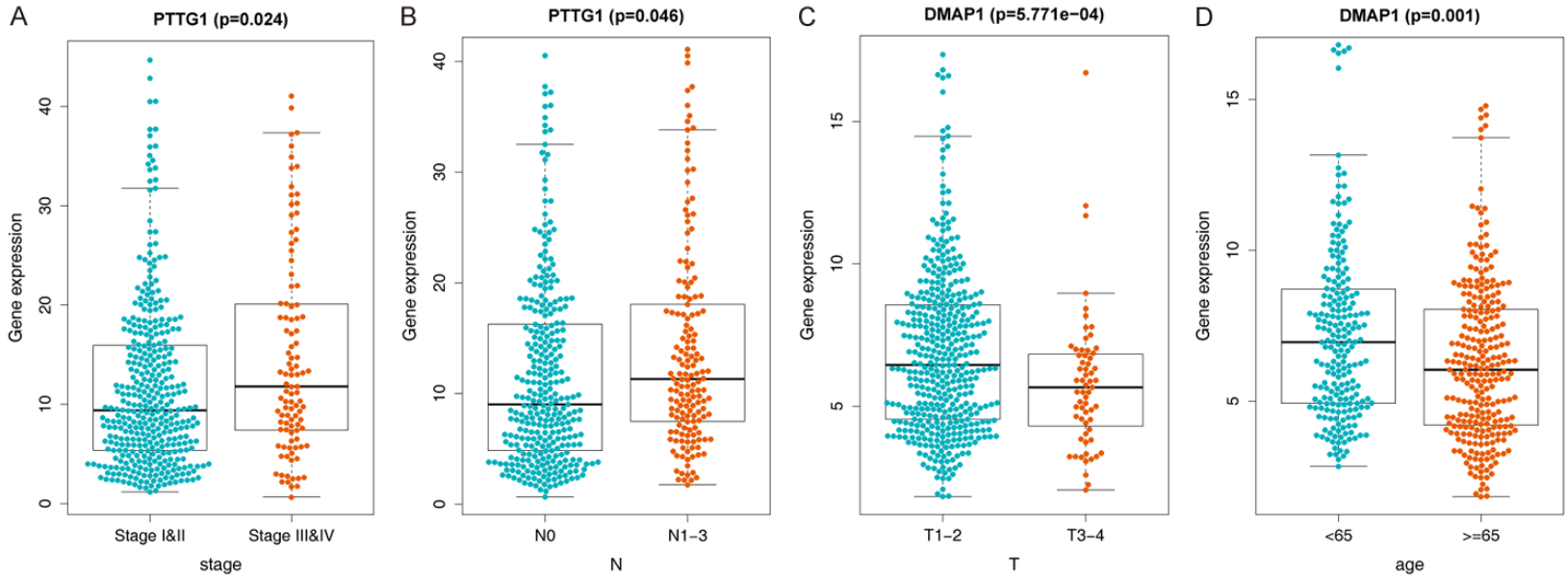


Figure 7. Correlation between the DRGs in the signature and clinicopathological characteristics. A. *PTGG1* and TNM stage. B. *DMAP1* and T stage. C. *PTGG1* and T stage. D. *DMAP1* and age.

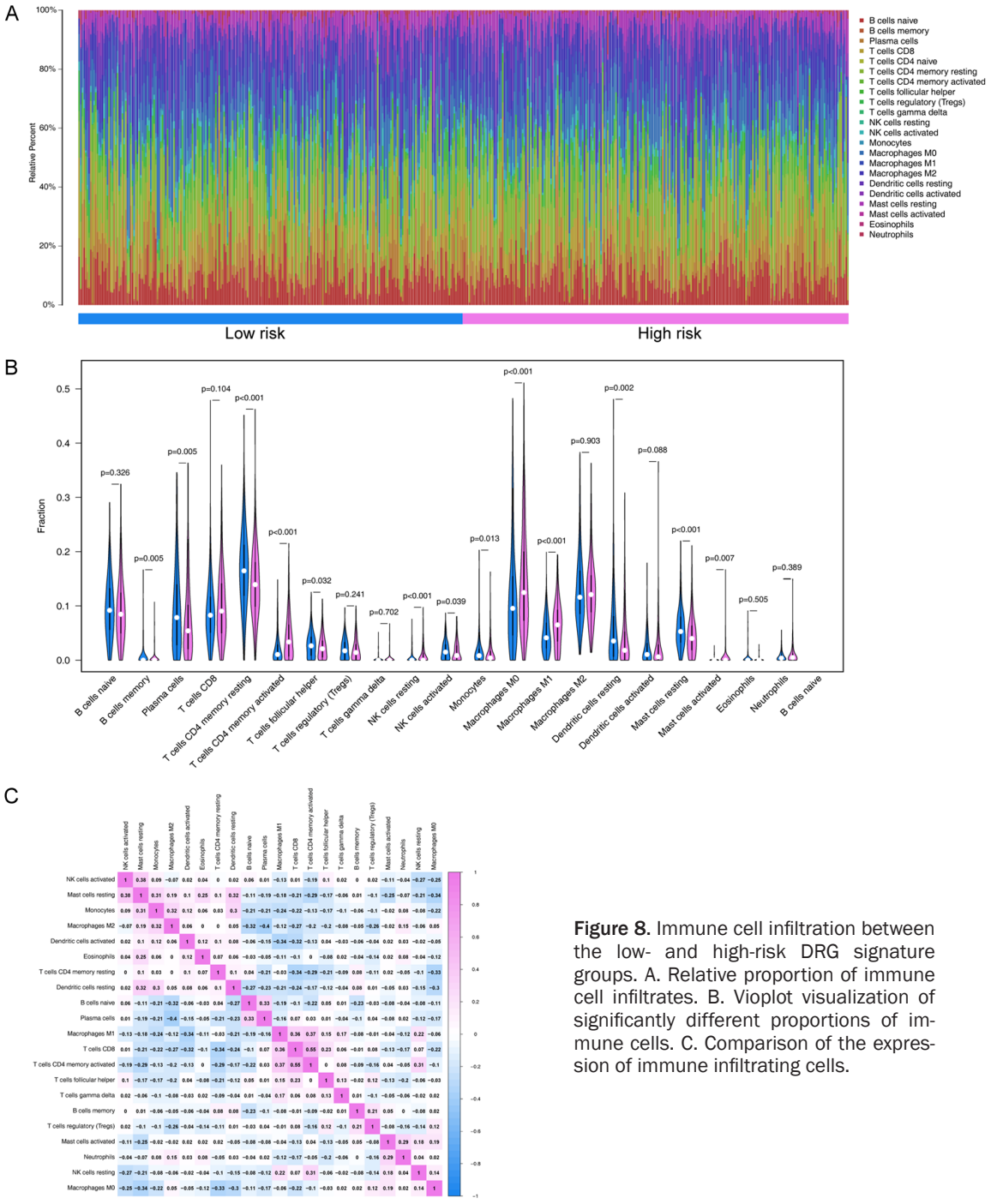


Figure 8. Immune cell infiltration between the low- and high-risk DRG signature groups. A. Relative proportion of immune cell infiltrates. B. Violplot visualization of significantly different proportions of immune cells. C. Comparison of the expression of immune infiltrating cells.

hensive nomogram with satisfactory AUCs based on the DRG signature and other variables to assess the deterioration and survival of patients. The performance of the nomogram was also evaluated via calibration curve analysis and DCA in both datasets. Our results suggest that the nomogram could be a cost-effective tool to predict LUAD prognosis and assist with clinical decision-making.

To understand the potential mechanisms through which these DRGs may mediate differential prognoses for LUAD, we also performed GSEA between the low- and high-risk groups based on our DRG signature. These analyses revealed that ADP metabolic process, cadherin binding, cell cycle G2M phase transition, cell cycle, DNA replication, and the p53 signaling pathway might play essential roles in the dis-

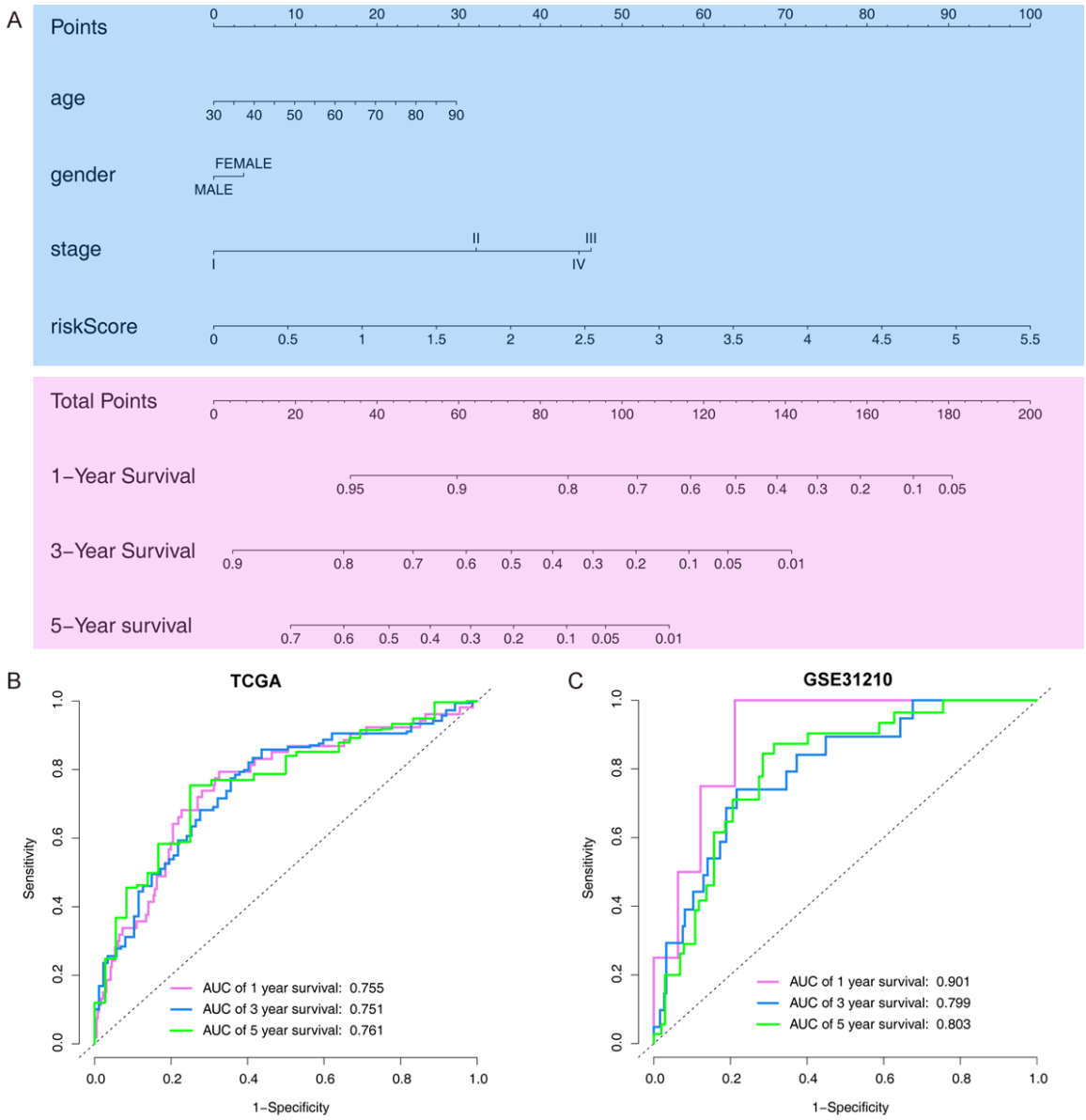


Figure 9. Construction and validation of a nomogram for predicting LUAD survival time based on the DRG signature and other clinical features. (A) Nomogram based on the DRG signature and clinical information of patients with LUAD. AUCs of the nomogram in ROC analysis were calculated at 1-, 3- and 5-years OS time in (B) TCGA training set and (C) the validation set.

tinct DRG-associated risk. It will be interesting, in future studies, to address whether these biological processes and pathways participate in the functions of the identified DRGs.

In this study, six DRGs were selected and included in the prognostic signature. *BTG2* is a member of the anti-proliferative (APR) gene family and has often been considered a tumor inhibitor in other cancers [27, 28]. However, to date, very little is known about the critical role that *BTG2* may play in LUAD. Our results revealed that, consistent with its previously described role as a tumor suppressor, high levels of *BTG2* were associated with a better prognosis. *DDA1* is an evolutionarily conserved gene that is linked to the ubiquitin proteasome pathway and facilitates the degradation of target proteins. A previous study reported that *DDA1* promotes lung cancer cell progression, indicating that it acts as a cancer-promoting gene [29], which is consistent with our data.

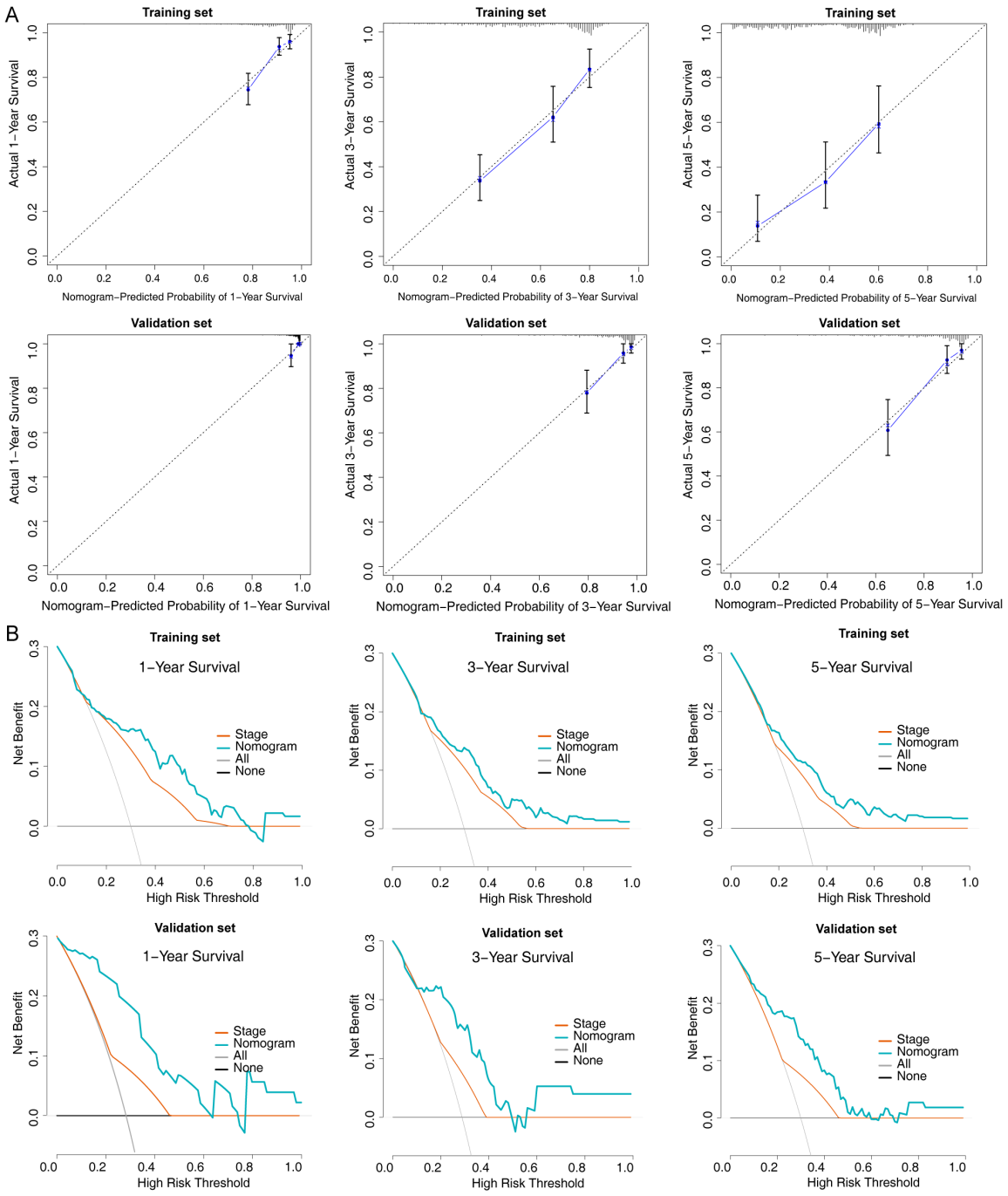


Figure 10. Evaluation of the nomogram in the training (TCGA) and validation (GSE31210) sets. A. Calibration curves of the nomogram used for evaluating the 1-, 3-, and 5-year AUCs in the training and validation sets. B. Decision curve analyses assessing clinical utility in the training and validation sets.

PTTG1 is known as an oncogene associated with tumorigenesis [30, 31]. Previous studies have demonstrated that *PTTG1* promotes cancer cell proliferation, migration, and invasion [32-34]. However, relatively little information is available on its role in LUAD. SMUG1, as a DNA repair protein, removes uracil residues from double-stranded and single-stranded DNA to maintain genomic stability following oxidative attacks [35]. It has been reported that *SMUG1* is associated with adverse clinicopathological features and may serve as a biomarker response to adjuvant therapy in breast cancer [36]. However, only a limited number of studies

have focused on the role of *SMUG1* in LUAD. DMAP1 was initially identified as a protein associated with the N-terminal domain of DNA methyltransferase 1, which is a key enzyme that mediates mammalian DNA methylation [37, 38]. In our study, we observed that high levels of *DMAP1* were associated with better survival. SMAD3 is a well-known transcription factor that plays an essential role in carcinogenesis, including LUAD [39, 40]. Consistent with previous studies, our results reveal that high *SMAD3* levels were significantly linked to a worse prognosis of LUAD [17].

It should be noted that there are certain limitations to this study. First, only three clinical features were included in the nomogram, while other features (such as drug treatments), which may act as independent risk factors for LUAD survival were not considered in this model. Second, the critical mechanisms underlying the prognostic capacity of the DRGs in LUAD still need to be explored in vitro and in vivo.

In summary, we developed a prognostic signature based on 6 DRGs, which independently predicts the survival of patients with LUAD and reflects the overall intensity of immunity in the LUAD microenvironment. For clinical application, we also constructed a prognostic nomogram by integrating the DRG signature with age, sex, and TNM staging. This nomogram will improve the ability of clinicians to evaluate the prognosis of patients with LUAD and will play a role in guiding follow-up and treatment processes.

Acknowledgements

The authors thank Yunliang Tan for his technical guidance on the analysis.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Ying Zhou, Departmend of Histology and Embryology, Jiangxi Medical College, Nanchang University, 461 Bayi, Nanchang 330006, Jiangxi, China. E-mail: zhouy1@ncu.edu.cn

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68: 394-424.

[2] Herbst RS, Heymach JV and Lippman SM. Lung cancer. N Engl J Med 2008; 359: 1367-1380.

[3] Rami-Porta R, Bolejack V, Crowley J, Ball D, Kim J, Lyons G, Rice T, Suzuki K, Thomas CF Jr, Travis WD and Wu YL. The IASLC lung cancer staging project: proposals for the revisions of the T descriptors in the forthcoming eighth edition of the TNM classification for lung cancer. J Thorac Oncol 2015; 10: 990-1003.

[4] Rusch VW, Chansky K, Kindler HL, Nowak AK, Pass HI, Rice DC, Shemanski L, Galateau-Sallé F, McCaughan BC, Nakano T, Ruffini E, van Meerbeeck JP and Yoshimura M. The IASLC mesothelioma staging project: proposals for the M descriptors and for revision of the TNM stage groupings in the forthcoming (Eighth) edition of the TNM classification for mesothelioma. J Thorac Oncol 2016; 11: 2112-2119.

[5] Sharma A and Almasan A. USP14 regulates DNA damage response and is a target for radiosensitization in non-small cell lung cancer. Int J Mol Sci 2020; 21: 6383.

[6] Rodriguez-Rocha H, Garcia-Garcia A, Panayiotidis MI and Franco R. DNA damage and autophagy. Mutat Res 2011; 711: 158-166.

[7] Tubbs A and Nussenzweig A. Endogenous DNA damage as a source of genomic instability in cancer. Cell 2017; 168: 644-656.

[8] Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. Nature 2001; 411: 366-374.

[9] Xiao W, Zheng J, Zhou B and Pan L. Replication protein A 3 is associated with hepatocellular carcinoma tumorigenesis and poor patient survival. Dig Dis 2018; 36: 26-32.

[10] Dai Z, Wang S, Zhang W and Yang Y. Elevated expression of RPA3 is involved in gastric cancer tumorigenesis and associated with poor patient survival. Dig Dis Sci 2017; 62: 2369-2375.

[11] Zhang D, Yang S, Li Y, Yao J, Ruan J, Zheng Y, Deng Y, Li N, Wei B, Wu Y, Zhai Z, Lyu J and Dai Z. Prediction of overall survival among female patients with breast cancer using a prognostic signature based on 8 DNA repair-related genes. JAMA Netw Open 2020; 3: e2014622.

[12] Santarpia L, Iwamoto T, Di Leo A, Hayashi N, Bottai G, Stampfer M, André F, Turner NC, Symmans WF, Hortobágyi GN, Pusztai L and Bianchini G. DNA repair gene patterns as prognostic and predictive factors in molecular breast cancer subtypes. Oncologist 2013; 18: 1063-1073.

[13] Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVSK and Varambally S. UALCAN:

a portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia 2017; 19: 649-658.
[14] Smith AL, Alirezaie N, Connor A, Chan-Seng-Yue M, Grant R, Selander I, Bascuñana C, Borgida A, Hall A, Whelan T, Holter S, McPherson T, Cleary S, Petersen GM, Omeroglu A, Saloustros E, McPherson J, Stein LD, Foulkes WD, Majewski J, Gallinger S and Zogopoulos G. Candidate DNA repair susceptibility genes identified by exome sequencing in high-risk pancreatic cancer. Cancer Lett 2016; 370: 302-312.
[15] Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M and Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015; 12: 453-457.
[16] Wang J, Hu ZG, Li D, Xu JX and Zeng ZG. Gene expression and prognosis of insulin-like growth factor-binding protein family members in non-small cell lung cancer. Oncol Rep 2019; 42: 1981-1995.
[17] Zeng Z, Yang Y, Qing C, Hu Z, Huang Y, Zhou C, Li D and Jiang Y. Distinct expression and prognostic value of members of SMAD family in non-small cell lung cancer. Medicine (Baltimore) 2020; 99: e19451.
[18] Ijsselsteijn R, Jansen JG and de Wind N. DNA mismatch repair-dependent DNA damage responses and cancer. DNA Repair (Amst) 2020; 93: 102923.
[19] Yin B, Delwel R, Valk PJ, Wallace MR, Loh ML, Shannon KM and Largaespada DA. A retroviral mutagenesis screen reveals strong cooperation between Bcl11a overexpression and loss of the Nf1 tumor suppressor gene. Blood 2009; 113: 1075-1085.
[20] Dutta D, Abarna R, Shubham M, Subbiah K, Duraisamy S, Chinnusamy R and Anbalagan M. Effect of Arg399Gln single-nucleotide polymorphism in XRCC1 gene on survival rate of Indian squamous cell head-and-neck cancer patients. J Cancer Res Ther 2020; 16: 551-558.
[21] Sample KM. DNA repair gene expression is associated with differential prognosis between HPV16 and HPV18 positive cervical cancer patients following radiation therapy. Sci Rep 2020; 10: 2774.
[22] Tang Y, Zeng Z, Wang J, Li G, Huang C, Dong X and Feng Z. Combined signature of nine immune-related genes: a novel risk score for predicting prognosis in hepatocellular carcinoma. Am J Transl Res 2020; 12: 1184-1202.
[23] Tang Y, Hu Y, Wang J and Zeng Z. A novel risk score based on a combined signature of 10 immune system genes to predict bladder cancer prognosis. Int Immunopharmacol 2020; 87: 106851.
[24] Sang W, Zhang Z, Dai Y and Chen X. Recent advances in nanomaterial-based synergistic combination cancer immunotherapy. Chem Soc Rev 2019; 48: 3771-3810.
[25] Li J, Li Q, Su Z, Sun Q, Zhao Y, Feng T, Jiang J, Zhang F and Ma H. Lipid metabolism genewide profile and survival signature of lung adenocarcinoma. Lipids Health Dis 2020; 19: 222.
[26] Tang XR, Li YQ, Liang SB, Jiang W, Liu F, Ge WX, Tang LL, Mao YP, He QM, Yang XJ, Zhang Y, Wen X, Zhang J, Wang YQ, Zhang PP, Sun Y, Yun JP, Zeng J, Li L, Liu LZ, Liu N and Ma J. Development and validation of a gene expression-based signature to predict distant metastasis in locoregionally advanced nasopharyngeal carcinoma: a retrospective, multicentre, cohort study. Lancet Oncol 2018; 19: 382-393.
[27] Tsui KH, Chiang KC, Lin YH, Chang KS, Feng TH and Juang HH. BTG2 is a tumor suppressor gene upregulated by p53 and PTEN in human bladder carcinoma cells. Cancer Med 2018; 7: 184-195.
[28] Shang D, Xie C, Hu J, Tan J, Yuan Y, Liu Z and Yang Z. Pancreatic cancer cell-derived exosomal microRNA-27a promotes angiogenesis of human microvascular endothelial cells in pancreatic cancer via BTG2. J Cell Mol Med 2020; 24: 588-604.
[29] Cheng L, Yang Q, Li C, Dai L, Yang Y, Wang Q, Ding Y, Zhang J, Liu L, Zhang S, Fan P, Hu X, Xiang R, Yu D, Wei Y and Deng H. DDA1, a novel oncogene, promotes lung cancer progression through regulation of cell cycle. J Cell Mol Med 2017; 21: 1532-1544.
[30] Wondergem B, Zhang Z, Huang D, Ong CK, Koeman J, Hof DV, Petillo D, Ooi A, Anema J, Lane B, Kahnoski RJ, Furge KA and Teh BT. Expression of the PTTG1 oncogene is associated with aggressive clear cell renal cell carcinoma. Cancer Res 2012; 72: 4361-4371.
[31] Liang HQ, Wang RJ, Diao CF, Li JW, Su JL and Zhang S. The PTTG1-targeting miRNAs miR-329, miR-300, miR-381, and miR-655 inhibit pituitary tumor cell tumorigenesis and are involved in a p53/PTTG1 regulation feedback loop. Oncotarget 2015; 6: 29413-29427.
[32] Li H, Yin C, Zhang B, Sun Y, Shi L, Liu N, Liang S, Lu S, Liu Y, Zhang J, Li F, Li W, Liu F, Sun L and Qi Y. PTTG1 promotes migration and invasion of human non-small cell lung cancer cells and is modulated by miR-186. Carcinogenesis 2013; 34: 2145-2155.
[33] Huang S, Liao Q, Li L and Xin D. PTTG1 inhibits SMAD3 in prostate cancer cells to promote their proliferation. Tumour Biol 2014; 35: 6265-6270.
[34] Zhou M, Chen J, Zhang H, Liu H, Yao H, Wang X, Zhang W, Zhao Y and Yang N. KLF10 inhibits

cell growth by regulating PTTG1 in multiple myeloma under the regulation of microRNA-106b-5p. Int J Biol Sci 2020; 16: 2063-2071.

[35] Pettersen HS, Sundheim O, Gilljam KM, Slupphaug G, Krokan HE and Kavli B. Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. Nucleic Acids Res 2007; 35: 3879-3892.

[36] Abdel-Fatah TM, Albarakati N, Bowell L, Agarwal D, Moseley P, Hawkes C, Ball G, Chan S, Ellis IO and Madhusudan S. Single-strand selective monofunctional uracil-DNA glycosylase (SMUG1) deficiency is linked to aggressive breast cancer and predicts response to adjuvant therapy. Breast Cancer Res Treat 2013; 142: 515-527.

[37] Muromoto R, Sugiyama K, Takachi A, Imoto S, Sato N, Yamamoto T, Oritani K, Shimoda K and Matsuda T. Physical and functional interactions between Daxx and DNA methyltransferase 1-associated protein, DMAP1. J Immunol 2004; 172: 2985-2993.

[38] Arand J, Spieler D, Karius T, Branco MR, Meilinger D, Meissner A, Jenuwein T, Xu G, Leonhardt H, Wolf V and Walter J. In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. PLoS Genet 2012; 8: e1002750.

[39] Tang X, Li G, Su F, Cai Y, Shi L, Meng Y, Liu Z, Sun J, Wang M, Qian M, Wang Z, Xu X, Cheng YX, Zhu WG and Liu B. HDAC8 cooperates with SMAD3/4 complex to suppress SIRT7 and promote cell survival and migration. Nucleic Acids Res 2020; 48: 2912-2923.

[40] Tong X, Wang S, Lei Z, Li C, Zhang C, Su Z, Liu X, Zhao J and Zhang HT. MYOCD and SMAD3/SMAD4 form a positive feedback loop and drive TGF-β-induced epithelial-mesenchymal transition in non-small cell lung cancer. Oncogene 2020; 39: 2890-2904.

Table S1. Prognosis related DRGs in TCGA

DRGs	HR	95% CI (low)	95% CI (high)	*P*-value
AATF	1.037225392	1.007372634	1.067962815	0.01416888
BCCIP	1.031253369	1.003297561	1.059988135	0.028181813
BLM	1.103854196	1.022799207	1.191332646	0.011107053
BRCA2	1.274605092	1.014639347	1.60117794	0.037083833
BRIP1	1.178695564	1.00863714	1.37742621	0.038626214
BTG2	0.993805361	0.989759635	0.997867625	0.002830197
BUB1	1.03920025	1.009404662	1.069875343	0.009579911
BUB1B	1.050485314	1.014646519	1.087589987	0.005419809
CCNA2	1.029304454	1.012915839	1.045958232	0.000420165
CCNB1	1.016187228	1.007380129	1.025071324	0.000299646
CDC25A	1.127936933	1.039215728	1.224232554	0.003973728
CDC25C	1.179182259	1.082668036	1.284300222	0.00015493
CDK1	1.017358242	1.003253342	1.031661445	0.015694374
CDKN1A	1.006672731	1.000845161	1.012534232	0.024757799
CEBPG	1.014097743	1.000223416	1.028164524	0.046398834
CHAF1B	1.091583285	1.016329876	1.172408779	0.016198513
CHEK1	1.083785534	1.03289481	1.137183643	0.001042039
CLSPN	1.077148609	1.009973834	1.148791273	0.023694984
CRY2	0.955097379	0.919980575	0.991554635	0.016229823
CSNK1E	1.027122154	1.017144618	1.037197563	7.74E-08
CUL4A	1.002066268	1.000177374	1.003958729	0.032016748
CUL4B	1.054942214	1.014300876	1.097211983	0.00762206
DDB1	1.025458264	1.001411434	1.050082528	0.037851203
DEK	1.010125165	1.000783244	1.01955429	0.033576485
DMAP1	0.940246359	0.895002551	0.987777315	0.014336119
DMC1	1.729237486	1.16365055	2.569725321	0.006730722
DTL	1.100272771	1.047667916	1.155518989	0.00013188
EGFR	1.005200229	1.000393919	1.010029629	0.033920321
ERCC1	1.028245681	1.003684012	1.053408411	0.023941324
ESCO2	1.35015392	1.098545616	1.659389998	0.004328267
EXO1	1.071810665	1.030426886	1.11485649	0.000556682
FANCC	1.172360785	1.008401511	1.362978731	0.038563991
FANCD2	1.156266655	1.034043281	1.292936767	0.010857136
FANCI	1.062266176	1.021849925	1.104280971	0.002272449
FANCL	1.047001862	1.001110099	1.094997344	0.044592309
FEN1	1.024095691	1.008563737	1.039866839	0.002261406
FHIT	0.726644409	0.541210525	0.975613135	0.033653092
FOXM1	1.029594035	1.013354082	1.04609425	0.000323996
GADD45A	1.011503061	1.002244984	1.020846659	0.014770352
GADD45G	0.987287125	0.97573451	0.998976522	0.033132222
GTF2H4	0.801432391	0.658538019	0.975333022	0.027157866
H2AFX	1.010576626	1.003656283	1.017544686	0.002691374
HDAC2	1.060355721	1.017351298	1.105177984	0.005531594
KPNA2	1.009774727	1.004481805	1.015095539	0.000285997
MMS22L	1.591860586	1.083017567	2.339777492	0.017991063
MYC	1.007949423	1.001358333	1.014583897	0.018006296
NBN	1.050716725	1.02144005	1.080832534	0.000600776

NEIL1	0.880787028	0.805374482	0.96326095	0.005442798
NHEJ1	6.304524237	1.535901815	25.87862419	0.010602043
NUDT1	1.036524286	1.002427588	1.071780752	0.035549773
PARP1	1.013549326	1.002104113	1.025125256	0.020194045
PLK1	1.047994045	1.023840975	1.072716902	8.13E-05
PRKDC	1.018120898	1.006208383	1.030174445	0.002783881
PTTG1	1.024191068	1.011541056	1.036999276	0.000163503
RAD23B	1.016984405	1.007100352	1.026965463	0.000725276
RAD51	1.095149904	1.030775559	1.163544577	0.003275296
RBBP7	1.038923387	1.015827481	1.062544402	0.00087152
RECQL	1.057316946	1.01873171	1.097363628	0.003299272
REV1	0.897381312	0.81699624	0.985675551	0.023741256
RFC3	1.040469025	1.008753644	1.073181543	0.012012253
RFC4	1.025769266	1.002357516	1.049727837	0.030783542
RFWD3	1.053796152	1.004420466	1.105599067	0.032345946
RHNO1	1.034833789	1.007949412	1.062435237	0.010786935
RNF168	1.053375015	1.013975899	1.094305026	0.007504965
RUVBL1	1.03764496	1.005606529	1.070704129	0.020923853
RUVBL2	1.018182781	1.001525069	1.035117549	0.032271515
SMAD2	1.175932642	1.033497806	1.337997595	0.013888347
SMAD3	1.035299325	1.010958498	1.060226205	0.004265674
SMC1A	1.031966949	1.00423164	1.060468263	0.023589408
SMC2	1.053412805	1.017176251	1.090940274	0.0035738
SMC3	1.021408397	1.00408715	1.039028449	0.015208739
SMUG1	1.04306962	1.002562858	1.085212986	0.036922386
SOD1	1.003683538	1.000488164	1.006889117	0.023824779
SP1	1.035806519	1.003279803	1.069387764	0.030686779
SPATA22	3.145132466	1.616742689	6.11838748	0.000738287
SSRP1	1.017518841	1.00482407	1.030373996	0.006702825
STAT1	1.002850545	1.00007099	1.005637824	0.044421335
SWSAP1	0.835967472	0.703015087	0.994063468	0.042627787
TDG	1.067909078	1.023036713	1.114749633	0.002701232
TICRR	1.290415308	1.102805327	1.509941623	0.001468854
TRIP13	1.025143107	1.003225472	1.04753958	0.024321769
TYMS	1.02358626	1.010696628	1.036640276	0.000311506
UBC	1.003226516	1.000182232	1.006280067	0.037757503
UBE2T	1.009058005	1.000509627	1.01767942	0.037770271
UBE2V2	1.04652333	1.010464988	1.08386841	0.011025146
UHRF1	1.064544905	1.015703902	1.115734471	0.009048234
UNG	1.019879863	1.003591761	1.036432318	0.016555041
USP1	1.033106693	1.011292354	1.055391584	0.002778597
WWP2	0.945664362	0.898965311	0.994789314	0.030606467
XAB2	0.958137991	0.91864167	0.999332427	0.046474687
XRCC4	1.054025333	1.012386648	1.097376585	0.010509699
XRCC5	1.014444545	1.006909923	1.022035547	0.000163011
XRCC6	1.005880526	1.001702352	1.010076129	0.005764675
YY1	1.041228728	1.000999093	1.083075171	0.044469761

Table S2. Prognosis related DRGs in GSE13210

DRGs	HR	95% CI (low)	95% CI (high)	*p* value
APEX2	3.195563754	1.191115151	8.573165825	0.021038027
ASCC3	2.192648831	1.107718517	4.340190058	0.024220212
ATRX	0.298224184	0.091679055	0.970097954	0.04438805
ATXN3	0.153836114	0.052935371	0.447064973	0.000583843
BARD1	1.813410669	1.007701145	3.263326899	0.047082671
BAX	2.156122966	1.023338888	4.542841377	0.043317116
BLM	2.724144691	1.604969983	4.623740245	0.000205083
BRCA1	1.892155313	1.248066574	2.868638424	0.002667203
BRCA2	2.353584342	1.241656589	4.461265136	0.008707116
BRCC3	2.101573313	1.283269829	3.441684898	0.003167799
BRIP1	1.707450969	1.228525579	2.373079456	0.001445662
BTG2	0.621186806	0.411751014	0.937151423	0.023247272
BUB1	1.709439962	1.225516595	2.384451582	0.001590641
BUB1B	1.626534339	1.268885942	2.084989573	0.000123219
CASP3	4.494625815	1.960438783	10.30466312	0.000385037
CCNA2	1.630434958	1.128158493	2.356333945	0.009274851
CCNB1	1.602899327	1.193275913	2.153136776	0.001727227
CCNE1	1.57485557	1.218917258	2.034732096	0.000511991
CDC25A	1.929127999	1.285780847	2.894377255	0.001501853
CDC25C	1.619991291	1.203915884	2.179863076	0.001446061
CDC45	1.474832128	1.222416908	1.779368226	4.97E-05
CDC6	1.663586603	1.253177708	2.208402183	0.00042936
CDK1	1.674059503	1.241104232	2.258049848	0.000739004
CDK2	2.264920614	1.166521876	4.397573241	0.015737544
CDKN1B	0.35418991	0.163243945	0.76848481	0.008632181
CHAF1A	2.879540378	1.606655254	5.160878644	0.000381293
CHEK1	1.742621653	1.214884597	2.499603857	0.002548504
CIB1	2.212941006	1.124673271	4.354249385	0.021436949
CRB2	1.42211592	1.05189953	1.922630093	0.022089132
CREB1	0.072041279	0.013300779	0.390198643	0.002274939
CREBBP	0.213792772	0.073659841	0.620519246	0.004543705
CRY2	0.46294763	0.28857104	0.742695832	0.001405911
CUL4B	5.660856439	1.822017254	17.58781128	0.002724511
CYP1A1	0.612597782	0.394856466	0.950411289	0.028746919
DAPK1	0.577642816	0.37230428	0.896232572	0.014332748
DCLRE1B	3.424121007	1.641470958	7.14274268	0.00103415
DDB1	5.202798777	1.127542652	24.0071762	0.034530341
DMAP1	0.444992913	0.227806553	0.869240547	0.017779755
DTL	1.661450996	1.167270777	2.364849242	0.004822244
DTX3L	2.432631761	1.206847974	4.903432257	0.012930941
E2F2	1.824264974	1.13786242	2.924732057	0.012552058
EME1	2.182024427	1.110593426	4.287104972	0.023550803
EP300	0.324543031	0.142829036	0.737442345	0.007204125
EPC2	0.302830337	0.124177201	0.738510868	0.008629332
ERCC3	0.18895305	0.059390074	0.601165357	0.004776365
ERCC4	0.428582141	0.200668222	0.915354955	0.028640305
ERCC6L2	0.313402396	0.13294233	0.73882458	0.00800719

ESCO2	1.887230082	1.227659548	2.901160496	0.003793236
ETS1	0.479405709	0.250386318	0.917900925	0.02652371
EXO1	1.391158656	1.13808033	1.700514766	0.001270821
FAM175A	0.316165862	0.14882783	0.671654299	0.002741911
FANCD2	2.711614233	1.591174546	4.621021474	0.000244736
FANCE	2.632040021	1.357466499	5.103355901	0.004175174
FANCI	1.956500756	1.404644719	2.725169686	7.20E-05
FEN1	2.054571613	1.42084881	2.970945596	0.000129916
FIGNL1	2.250965994	1.38799778	3.650472628	0.001005344
FOXM1	1.654449843	1.290993219	2.120231341	6.95E-05
FZR1	3.566000401	1.141509342	11.13995164	0.028692015
GADD45A	1.677009253	1.06263432	2.646592513	0.02635552
GSTP1	2.143527185	1.128511314	4.07147783	0.019842386
GTF2H3	0.152127898	0.042096378	0.549759823	0.004070321
H2AFX	1.96292585	1.24117724	3.104373629	0.003928909
HDAC2	2.610587294	1.203483443	5.662866459	0.015149995
HELQ	0.150592295	0.044011019	0.515280942	0.002558091
HERC2	0.318852349	0.13779056	0.737835888	0.007580281
IFI16	1.903447045	1.070922759	3.383167109	0.028273779
IKBKG	6.408472338	2.838715456	14.46728929	7.77E-06
INIP	3.312632513	1.120318027	9.795017041	0.030360092
INO80D	0.311023443	0.099079838	0.976339729	0.045393534
INTS3	0.345001873	0.145438938	0.818393572	0.015748068
JMY	0.219057279	0.09832332	0.488043848	0.000203134
KIAA0101	2.166054375	1.543541033	3.039628657	7.79E-06
KIN	4.547514522	1.432856332	14.43263213	0.010159618
KPNA2	2.10351708	1.069958517	4.135472579	0.031082051
MBD4	3.370698998	1.391211143	8.166705533	0.00711891
MC1R	1.846848809	1.064813076	3.203238767	0.029000727
MDC1	0.384670206	0.154726381	0.956340908	0.039779159
MDM4	0.242484139	0.109476305	0.537089347	0.000479469
MGME1	2.145613318	1.007584018	4.569005092	0.047753545
MRE11A	3.706464385	1.504806278	9.129333415	0.004392156
MSH2	2.003162459	1.009752037	3.973906156	0.046841446
MSH3	0.239328961	0.081790865	0.700302554	0.009046911
MSH6	2.467040812	1.029404999	5.912435212	0.042872337
MUM1	0.232659834	0.094490836	0.572866115	0.001515276
NEIL3	1.391545914	1.131764493	1.71095669	0.001724394
NEK1	0.351431449	0.129989556	0.950107588	0.039319866
NME1	1.828670746	1.170236452	2.857573522	0.008043999
NSMCE2	2.752716181	1.392315613	5.442333837	0.003595289
NUDT1	2.279356722	1.38886622	3.740797343	0.001115875
PALB2	2.855105896	1.243162082	6.557173676	0.013396698
PARP9	2.416080204	1.248112259	4.67701804	0.008854535
PARPBP	1.894086956	1.209818227	2.965375558	0.005226053
PCNA	2.54918411	1.402593194	4.633089377	0.002141681
PLK1	1.657531402	1.191633949	2.305582475	0.002688656
PMS1	0.243133816	0.067836537	0.871419071	0.029909594
PNKP	2.756509735	1.364798232	5.567376728	0.004697146

POLD4	1.883687684	1.052078497	3.372637404	0.033106026
POLE3	3.293028195	1.329529625	8.156294144	0.010010585
POLE4	2.347341808	1.179756009	4.670468742	0.015061105
POLI	0.435396951	0.228422552	0.829911513	0.011522676
POLK	0.255730163	0.121208417	0.5395493	0.000343943
POLQ	1.7244413	1.259505034	2.361005091	0.000675713
POLR2B	0.116608733	0.017289943	0.786445448	0.027338649
POLR2G	3.679418925	1.471143877	9.202447045	0.005347329
POLR2H	2.154183483	1.064492049	4.359362273	0.032864516
PPM1D	0.313964827	0.116247102	0.847968771	0.022296495
PPP2R5A	0.3538784	0.179064113	0.699358011	0.002800544
PPP2R5C	0.195896566	0.053849159	0.712647427	0.013356687
PPP4C	3.943861144	1.818161484	8.554818075	0.000514375
PTTG1	1.764449413	1.290277071	2.412878443	0.000376634
RAD17	0.267245699	0.077937937	0.916373543	0.035828485
RAD51	2.052826755	1.430489093	2.945913889	9.51E-05
RAD51AP1	1.975821053	1.422811292	2.743771333	4.81E-05
RAD54L	1.347564611	1.003436055	1.809712111	0.047393802
RBBP4	0.21134587	0.063076809	0.708137858	0.011756114
RBBP8	1.880169913	1.066896811	3.313384075	0.028965434
RBM14	0.054479142	0.010825873	0.274155885	0.000416221
RDM1	1.830050903	1.250070626	2.679117675	0.001885298
RECQL4	1.679421614	1.168275328	2.414205699	0.005111839
RECQL5	3.567936569	1.606417258	7.924573329	0.001782914
REV1	0.233157896	0.091969808	0.59109186	0.002156822
REV3L	0.302888014	0.165613369	0.553947724	0.00010547
RFC2	1.87987297	1.10038405	3.21153545	0.02088519
RFC3	2.269588379	1.184572007	4.348432496	0.013491125
RFC4	1.938628808	1.275672094	2.946118892	0.001933884
RFC5	2.894200774	1.302761714	6.429723894	0.00907047
RNASEH2A	1.935194037	1.226189061	3.054158679	0.004570563
RPA3	1.7359802	1.062733672	2.835731411	0.02759639
RPAIN	0.210206993	0.083566127	0.528766638	0.000920185
RPS27A	0.345572131	0.127371585	0.937572514	0.036929103
RPS3	0.039147322	0.002369298	0.646821484	0.023548134
RRM2B	0.416153381	0.246626355	0.702210584	0.001022188
SFPQ	0.132929365	0.047595899	0.371255012	0.000117715
SHFM1	2.980363633	1.402040016	6.335459248	0.004536221
SHPRH	0.240182366	0.098115743	0.587954256	0.001791944
SIRT1	0.172866452	0.065259476	0.457907603	0.000413217
SIRT6	2.812020319	1.003427943	7.880444558	0.049244027
SLC30A9	0.197018315	0.078125585	0.496843852	0.000577194
SMAD3	2.33251135	1.150092003	4.730586063	0.018895779
SMAD4	0.467296373	0.248624282	0.878296757	0.018125974
SMARCA2	0.299375535	0.148708796	0.602692733	0.00072935
SMARCA5	0.124507069	0.029792655	0.520329935	0.004299538
SMC2	2.290605944	1.252504475	4.189107264	0.007124849
SMUG1	3.523258197	1.936546153	6.410045177	3.72E-05
SPP1	1.759898058	1.299829375	2.382805957	0.000256084

SSRP1	2.971005698	1.187623459	7.432385066	0.019937899
STAT1	1.776198013	1.042558659	3.026092925	0.034576662
STRA13	1.844334099	1.112547015	3.057460243	0.017619869
SWI5	2.386009816	1.340394326	4.247289574	0.003119624
TDP1	2.29410751	1.254876695	4.193981201	0.006985371
TICRR	1.516388601	1.018455418	2.257766369	0.040363478
TOP2A	1.534441433	1.199882614	1.962284047	0.000644551
TOPBP1	2.272160573	1.096337935	4.709053208	0.027291087
TRIP13	1.353677342	1.072587228	1.708432003	0.010770916
TWIST1	1.338185222	1.094638892	1.635918204	0.004480578
TYMS	1.475956342	1.072586501	2.031022322	0.016839917
UBA1	2.080407219	1.013251269	4.271491511	0.045952514
UBC	70.57559366	4.70913429	1057.713396	0.002057633
UBE2A	3.554076519	1.156114475	10.92578648	0.026888155
UBE2I	0.361061571	0.144610321	0.901494836	0.029102604
UBE2NL	0.510619954	0.310185538	0.840570256	0.008220739
UBE2T	1.588125858	1.188899516	2.121410354	0.001740424
UBE2V2	2.696959309	1.225228702	5.936515774	0.013718273
UHRF1	1.445908469	1.10809252	1.886711859	0.006608237
USP7	0.263681998	0.078326238	0.887674399	0.031369778
UVSSA	0.403985418	0.204058329	0.799791992	0.009293152
WRNIP1	2.753878451	1.24375956	6.097518177	0.012495185
WWP2	0.381322716	0.173568319	0.837750893	0.016358689
ZNF350	0.333109602	0.1375437	0.806740017	0.014858035