



OPEN ACCESS

Edited by:

Alison Taylor, University of Leeds, United Kingdom

Reviewed by:

Chengzhi Zhou, Clinical Management Department of National Respiratory Medical Center, China Weilai Dong, Yale University, United States

*Correspondence:

Yun Li liyun28@mail.sysu.edu.cn Yuhong Liang rainbowliang1230@163.com Bihui Huang bihui_huang@126.com

[†]ORCID:

Yun Li orcid.org/0000-0003-0949-6519 Yuhong Liang orcid.org/0000-0003-1157-984X Bihui Huang orcid.org/0000-0003-3084-7096

[‡]These authors have contributed equally to this work and share the first authorship

[§]These authors have contributed equally to this work and share the last authorship

Specialty section:

This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Oncology

Received: 30 June 2021 Accepted: 27 October 2021 Published: 02 December 2021

Citation:

Leng Y, Dang S, Yin F, Gao T, Xiao X, Zhang Y, Chen L, Qin C, Lai N, Zhan X-Y, Huang K, Luo C, Kang Y, Wang N, Li Y, Liang Y and Huang B (2021) GDPLichi: a DNA Damage Repair-Related Gene Classifier for Predicting Lung Adenocarcinoma Immune Checkpoint Inhibitors Response. Front. Oncol. 11:733533. doi: 10.3389/fonc.2021.733533

GDPLichi: a DNA Damage Repair-Related Gene Classifier for Predicting Lung Adenocarcinoma Immune Checkpoint Inhibitors Response

Yang Leng^{1‡}, Shiying Dang^{1‡}, Fei Yin², Tianshun Gao¹, Xing Xiao¹, Yi Zhang¹, Lin Chen¹, Changfei Qin¹, Nannan Lai¹, Xiao-Yong Zhan¹, Ke Huang¹, Chuanming Luo³, Yang Kang¹, Nan Wang¹, Yun Li^{2*†‡}, Yuhong Liang^{4*†§} and Bihui Huang^{1*†§}

¹ Scientific Research Center, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China, ² Department of Thoracic Surgery, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China, ³ Center for Clinical Neuroscience, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, China, ⁴ School of Medicine, Southern University Of Science And Technology, Shenzhen, China

Lung cancer is one of the most common and mortal malignancies, usually with a poor prognosis in its advanced or recurrent stages. Recently, immune checkpoint inhibitors (ICIs) immunotherapy has revolutionized the treatment of human cancers including lung adenocarcinoma (LUAD), and significantly improved patients' prognoses. However, the prognostic and predictive outcomes differ because of tumor heterogeneity. Here, we present an effective method, GDPLichi (Genes of DNA damage repair to predict LUAD immune checkpoint inhibitors response), as the signature to predict the LUAD patient's response to the ICIs. GDPLichi utilized only 7 maker genes from 8 DDR pathways to construct the predictive model and classified LUAD patients into two subgroups: low- and high-risk groups. The high-risk group was featured by worse prognosis and decreased B cells, CD8⁺ T cells, CD8⁺ central memory T cells, hematopoietic stem cells (HSC), myeloid dendritic cells (MDC), and immune scores as compared to the low-risk group. However, our research also suggests that the high-risk group was more sensitive to ICIs, which might be explained by increased TMB, neoantigen, immune checkpoint molecules, and immune suppression genes' expression, but lower TIDE score as compared to the lowrisk group. This conclusion was verified in three other LUAD cohort datasets (GSE30219, GSE31210, GSE50081).

Keywords: DNA damage repair (DDR), immune check inhibitor (ICI), GDPLichi, lung adenocarcinoma, gene classifier

INTRODUCTION

Lung cancer ranks the second in incidence and top in mortality among malignancies worldwide (1), of which lung adenocarcinoma is the most common subtype (2). The prognosis of advanced and recurrent lung cancer is usually poor because most standard treatments by cytotoxic anticancer drugs only have limited therapeutic effects. In recent years, with a better understanding of immune response regulation,

1

the immune checkpoint inhibitor (ICI) therapy showed improved survival rates in multiple cancers including non-small-cell lung cancer (NSCLC). The principle of ICIs is to reactivate immune cells by using specific antibodies against inhibitory signaling molecules such as CTLA-4 and PD-1 expressed on tumor and immune cells. Currently, the approved drugs of anti-CTLA-4 (ipilimumab), anti-PD-1 (nivolumab and pembrolizumab), anti-PD-L1 (atezolizumab, avelumab, and durvalumab), and their combinations have performed significant improvements in treating advanced NSCLC patients (3–6). Lung adenocarcinoma accounts for 80% of NSCLC and benefits most from ICIs therapy. However, it was also reported that there were still only partial LUAD patients responsive to ICIs (7).

PD-L1 expression has been widely used as an ICI response predictive marker, but the sensitivity and specificity are not very consistent due to different antibodies and cutoff values used for PD-L1 test (3, 8, 9). Meanwhile, PD-L1 expression cannot accurately reflect the complicated tumor immune microenvironment (10). Recent studies have also reported that tumor mutation burden (TMB) is closely related to the efficacy of ICIs response (11, 12) and can also be used as a predictive marker for the efficacy of ICI treatment. Like PD-L1, the cut-off value of TMB is controversial (13-15). Additionally, TMB alone does not directly produce neoantigen processing by major histocompatibility complex (MHC) class molecules, thus the accuracy of TMB as a predictor for ICI treatment is modest. Neoantigen expressed on tumor cells is one of the main targets for an effective antitumor T-cell response (16), but difficult to be identified. Therefore, identifying novel markers that can efficiently and accurately predict ICI responses is urgent. One promising area for this research is DNA damage repair (DDR). To ensure the integrity of the genome, cells activate DNA damage repair pathways to repair genetic lesions (SNP, Indel, etc.) during the process of DNA replication. DDR consists of eight pathways including miss match repair (MMR), base excision repair (BER), nucleotide excision repair (NER), direct damage repair (DR), homologous dependent recombination (HDR), nonhomologous end joining (NHEJ), fanconi anemia pathway (FA), and translesion DNA synthesis (TLS). Defects in DDR pathways lead to the accumulation of genomic aberrations and an elevated TMB (17-19), thus promoting tumor development (20). Many studies have shown that mutations in DDR pathway genes are associated with ICIs responses (20, 21). Patients who have DDR genomic alterations usually have a better clinical benefit after ICIs therapy (19, 21).

The Tumor Immune Dysfunction and Exclusion (TIDE) algorithm is a computational method that uses gene expression profiles to predict the ICIs response, particularly successful in NSCLC and melanoma (22). TIDE uses a specific set of marker genes to estimate dysfunction of tumor-infiltrating cytotoxic T lymphocyte and exclusion of CTL by an immunosuppressive factor to predict patients' response to ICIs. Patients with lower TIDE scores have a lower chance of antitumor immune escape, thus having a higher response rate of ICIs treatment (22). The TIDE score exhibited a higher accuracy than PD-L1 expression level and TMB in predicting the overall survival of patients treated with ICIs (20, 23, 24). Some studies also have reported its utility in predicting or evaluating the ICIs efficacy (24–28).

We identified seven significant genes (e.g., DUT, MGMT, POLH, RAD1, RAD17, TYMS, and YWHAG) strongly associated with prognosis from DDR pathways using Cox regression analysis.

Patients with lower expression of DUT, TYMS, and YWHAG but higher expression of MGMT, POLH, RAD1, and RAD17 had a better prognosis. Based on the expressions and weights calculated by Cox regression on these genes, we developed a classifier, GDPLichi (Genes of DDR to Predict LUAD immune checkpoint inhibitors), as the signature to predict the ICIs response. LUAD patients were classified into low- and highrisk groups based on the cutoff of the GDPLichi score. Many features, including PDCD1, CTLA4, PD-L1 expression, TMB, and neoantigen, displayed strong discerning abilities in the survival analysis of these two subgroups. Especially, the highrisk subgroup had a worse prognosis but is presumably more efficacious towards ICIs treatment.

MATERIALS AND METHOD

Data Source

To predict the LUAD ICIs response, we built a multi-step approach called GDPLichi described below (Figure 1 and Supplementary Figure S1). The transcriptome gene expression, genomic data, and clinical phenotype data of 526 TCGA-LUAD samples were downloaded from the website xenabrowser (https://xenabrowser.net/datapages/). TCGA raw RNA-Seq transcriptome count data including 526 LUAD samples were further transferred into a transcript per kilobase mullion (TPM). Three validation groups of raw data, including 438 LUAD samples from 3 cohorts [GSE31210 (29), GSE30219 (30), and GSE50081 (31)], were downloaded from Gene Expression Omnibus (GEO) repository. Then raw data were transferred to expression data using the "Oligo" package in R software. For genes with multiple probes, their expression levels were calculated as the maximum expression level of these probes. Finally, all expression data were normalized and converted to Z-scores.

GDPLichi Score

First, a univariate Cox regression model was used to assess the association of 276 DNA damage repair-related genes (**Supplementary File 1**) with the overall survival in the TCGA LUAD cohort. P-value was used to identify key genes and genes with P-value < 0.05 were considered as predictive genes (**Supplementary File 2**). Then, 63 predictive genes were selected for multivariate Cox regression and genes with P-value < 0.05 were considered as risk genes (**Supplementary File 3**). Finally, seven risk genes were obtained by multivariate Cox regression and combined to construct the GDPLichi classifier. By combining the expression values of risk genes and weighting by multivariate Cox regression coefficients, the GDPLichi score for each patient was calculated as follows:

GDPLichi score = $\sum_{i=1}^{n} expre_i \beta_i$



Here n is the number of prognostic genes, $expre_i$ meant the expression value of gene i, and β_i represented the regression coefficient of gene i in the multivariate Cox regression analysis. Using the median GDPLichi score as a cutoff value, TCGA and GEO LUAD patients could be classified into low and high-risk groups.

Gene Set Enrich Analysis, Survival Analysis, Principal Components Analysis, Tumor Microenvironment Analysis, and TIDE

R language 4.0 was applied in this study for the statistical analyses. GSEA was used to explore the pathway enrichment between low- and high-risk groups using the R package "clusterProfiler" (32) on the Reactome pathway database (33) with default parameters. The fold change of gene expressions between two groups was used to rank the genes. The absolute values of the normalized enrichment score (NES) >1 and P-value ≤ 0.05 were used to screen out significantly enriched pathways.

The "survivalROC" package was used to plot the survival ROC curve. The cutoff of survival time was set to 36 months. "Forest plot" was used in the "forestmodel" package and the "factor_separate_line" parameters were set as TRUE. Survival analysis of two groups was carried out by the R package "survminer". PCA was used by the R packages "FactoMineR" and "factoextra" with the values of all genes' expression as the input. We used the "xCell" package (34) to estimate relative subsets of immune cells. TIDE Score was calculated with the TIDE algorithm (22) from the website (http://tide.dfci.harvard. edu). All R package parameters can be found in the source analysis code in main_code.R (**Supplementary File 6**).

Patient Sample Collection

From the TCGA LUAD cohort downloaded from the xenabrowser website, samples with survival, and genomic data were collected. In datasets GSE31210, GSE50081, and GSE30219, lung squamous cell carcinoma samples could be excluded and lung adenocarcinoma with survival data were collected.

RESULT

Construction of GDPLichi

A univariate Cox regression model was used to assess the association of 276 DNA damage repair genes with the overall survival in the TCGA LUAD cohort. There were 63 predictive genes screened out with an initial significance (P <0.05). By using these 63 predictive genes as input for multivariate Cox regression, seven risk genes (DUT, MGMT, POLH, RAD1, RAD17, TYMS, and YWHAG) were screened out (Figure 2A) and Kaplan-Meier analysis further confirmed the prognostic value of the individual genes (Suppelementary Figure S2). The multivariate Cox regression analysis of the above seven risk genes showed high accuracy in predicting the survival of LUAD patients (Figure 2B). By combining the expression values of seven risk genes and weighted by COX regression coefficients, the GDPLichi score for each patient was calculated (Described in 2.2). To further facilitate the application of the GDPLichi, the patients were divided into low- and high-risk groups according to the median value of the GDPLichi score. PCA showed that patients could be distinctively clustered according to the selected signatures (the seven risk

genes) in the TCGA LUAD cohort (Figure 2C) and three GEO validation cohorts (Supplementary Figure S3A). In addition, Spearman's correlation test indicated that GDPLichi was significantly correlated with the selected genes in the TCGA LUAD cohort (Figure 2D) and three GEO validation cohorts (Figure S3B).

Identification of LUAD Subgroups With Prognostic Significance According to GDPLichi

LUAD patients were classified into low- or high-risk groups based on the median GDPLichi score described above. The overall survival analysis for these two subgroups showed a significant difference in the TCGA cohort (**Figure 3A**, P<0.0001) and three GEO validation cohorts (**Suppelementary Figure S4A**).

The hazard ratio of the two subgroups in the TCGA cohort is 1.912 (GSE30219: 2.99, GSE31210: 3.79, GSE50081: 2.43). The 95% confidence interval of two subgroups of TCGA cohort is 1.421-2.573 (GSE30219: 1.585-5.641, GSE31210: 1.72-8.351, GSE50081: 1.356-4.356). The difference remained statistically significant after adjusting for age, gender, stage, and smoking



FIGURE 2 | DDR signature accurately predicts the prognosis of LUAD patients. (A) Univariate and multivariate Cox regression analyses screened out seven risk genes. The point number represents a score for the relation between the expression of each selected gene and the predicted survival calculated by Cox regression. High and low represent the highest and lowest expression levels of the gene, respectively. Total points were the sum of the individual points from the seven selected genes. Based on the total points, 1-year and 3-year predicted overall survival rates of each LUAD patient were calculated. The higher the number, the lower the predicted survival. (B) Multivariate Cox regression analysis of the seven risk genes in predicting the survival of LUAD patients. (C) PCA based on the expression profile of the seven risk genes from different risk groups. (D) Correlation between the GDPLichi and the seven risk genes in the TCGA LUAD cohort.



status in the TCGA cohort (**Figure 3C**) and three GEO validation cohorts (**Suppelementary Figure S4C**). To test the practicality of the GDPLichi classifier, we applied ROC (Receiver Operating Characteristic) analyses to the TCGA cohort and found that the GDPLichi score could function as a better prognostic index than any risk gene alone (**Figure 3B**). This result was also validated in the three GEO cohorts (**Suppelementary Figure S4B**). Therefore, the GDPLichi could be a good model to predict the prognosis of LUAD patients.

GSEA Explored the Pathway Enrichment Between Low- and High-Risk Groups

To further investigate the difference of biological mechanisms between low- and high-risk groups divided by GDPLichi, we performed GSEA on the TCGA LUAD cohort. It revealed that cell proliferation-related pathways such as cellular response to hypoxia, MAPK signaling, and noncanonical NF-KB signaling were significantly enriched in the high-risk group (Figures 4, B). Meanwhile, cell cycle pathways were also significantly enriched in the high-risk group (Figures 4B, D). The results also showed that immune-related pathways such as antigen procession, crosspresentation, interleukin-10 signaling, and MHC class II antigen presentation were significantly enriched (Figure 4C). By examining the expression of HLA genes, it was revealed that the expression of MHC II genes in the low-risk group was significantly higher than in the high-risk group (Figure 4E). MHC II genes are only expressed in antigen-presenting cells. This may indicate a higher tumor-infiltrating lymphocyte (TIL) in the low-risk group, and ultimately a poorer prognosis.

The Difference in Tumor Immune Microenvironment Between Low- and High-Risk Groups

The "xCell" algorithm was employed to estimate the immune cells in malignant tumor tissues between two subgroups using

RNA sequencing data. Our results showed that the immune scores, B cells, hematopoietic stem cells (HSC), myeloid dendritic cells (MDC), CD8⁺ T cells, and CD8⁺ central memory T cells were significantly higher in low-risk groups compared to high-risk groups, suggesting a higher TIL in the low-risk group (Figures 5A-F). CD8⁺ T cells, also named cytotoxic T cells, are one of the major tumor killer cells and CD8⁺ cell exclusion is strongly associated with tumor immune escape. Therefore, we examined the expression of several immune-suppression genes such as TIM-3 (HAVCR2), IDO1, LAG3, PD-L2 (PDCD1LG2), TIGIT, CD276, CD160, VEGFA, VEGFB, SLAMF7, KIR2DL3, and IL1B between low- and high-risk groups. As shown in Figure 5G, the high-risk group had a higher expression of immunosuppression genes than the low-risk group, which might account for higher sensitivity to ICIs in the high-risk subgroup of LUAD patients.

The Difference in TMB, Neoantigen, and ICIs-Target Expression Between Low- and High-Risk Subgroups

To predict the sensitivity to ICIs between low- and high-risk groups as classified by the GDPLichi model, we further examined immunotherapy-related markers such as tumor mutation burden (TMB), neoantigen, and expression of PDCD1 (PD-1), CD274 (PD-L1), and CTLA4. The degree of TMB and neoantigen in the high-risk group was significantly higher as compared to the low-risk group in the TCGA-Cohort (**Figures 6A, B**). A significantly higher expression of PD-L1, PDCD1, and CTLA4 was also observed in the high-risk group as compared to the low-risk group in the TCGA-Cohort (**Figure 6C**) as well as the three GEO validation cohorts (**Supplementary Figures S5A-C**). These results indicated that the high-risk group might be more sensitive to immunotherapy.



It also suggested that the GDPLichi model may help to predict the response to ICIs of LUAD patients.

Gene Mutation Pattern Between Lowand High-Risk Groups Classified by GDPLichi Model

We identified 12 candidates from the top 25 frequently mutated genes in the TCGA cohort of LUAD patients, which exhibited a significant difference between the low- and the high-risk group classified by GDPLichi (**Figure 7**). Recent studies reported that mutations in the TTN and MUC16 genes indicated high TMB (35, 36) and could be used to predict immunotherapy efficacy (37). TTN mutation status can independently predict immunotherapy prognosis in lung adenocarcinoma patients after ICIs (38). MUC16 mutation was associated with greater response rates

associated with ICIs response and overall survival (39). TP53 is a well-known tumor suppressor gene with mutation occurring in more than 50 percent of all malignancies. TP53 mutation is usually associated with a poor prognosis. Some studies also reported that loss of CSMD3 results in increased proliferation of airway epithelial cells in the LUAD (40). Ovarian carcinoma patients with CSMD3 mutation had sustained responses to anti-PD1 without prior chemotherapy (41). Somatic mutations in the ZFHX4 gene are associated with poor overall survival in Chinese lung cancer patients (42). The mutation of RYR2 is a significant biomarker associated with high TMB in LUAD (43). Patients with lung adenocarcinoma with the ADAMTS12 mutation would have a worse prognosis (44). Taken together, these results suggested that the high-risk group might be more sensitive to ICIs and GDPLichi model may predict the response to ICIs.



FIGURE 5 | The high-risk LUAD group exhibits relatively lower infiltration of B, HSC, MDC, CD8⁺ T, and memory T cells, and lower immune scores than the lowrisk group, but has higher expression of immuno-suppression genes. (A–F) Comparison of infiltrating immune cells (xCell) between low- and high-risk groups using xCell algorism. (G) Statistical analysis of the expression of immunosuppression genes between low- and high-risk groups. *P < 0.05; **P < 0.01; ***P < 0.001; ***P < 0.001; ns, no significant difference.





ICIs Response Prediction Between Lowand High-Risk Groups of LUAD Patients Classified by GDPLichi Model

The TIDE algorithm has been proved to help predict ICIs response of LUAD patients with high accuracy (22). Therefore, we calculated the TIDE scores of both low- and high-risk groups of the TCGA LUAD cohort as classified by the GDPLichi model. The results revealed that the TIDE score in the high-risk group was significantly lower than the low-risk group (**Figure 8A**). Similar results were observed in the other three external GEO datasets (**Figures 8B-D**). These results suggested that the high-risk group has a lower chance of antitumor immune escape and exhibiting a higher response rate of ICIs treatment.

DISCUSSION

Lung cancer ranks second in incidence and top in mortality among malignancies worldwide (1). Recently, immunotherapy has become an important new therapeutic approach in treating multiple types of cancer with promising results. It has greatly changed the landscape of cancer care. Many studies have shown that mutations in DDR pathway genes are associated with the prognosis of LUAD patients (20, 21), however, using the DDR gene expression profile as a molecular signature to predict the

response to ICIs of LUAD patients has not been reported yet. In this study, we constructed a GDPLichi model based on seven DDR genes (DUT, MGMT, POLH, RAD1, RAD17, TYMS, and YWHAG), to classify LUAD into two distinct subgroups: lowand high-risk groups. Thymidylate synthase (TYMS) is a critical target for cancer chemotherapy (45). Tyrosine 3-Monooxygenase/ Tryptophan 5-Monooxygenase Activation Protein Gamma (YWHAG) is also known as 14-3-3 γ . A recent study reported that knockdown of YWHAG suppresses epithelial-mesenchymal transition (EMT) and reduces the metastatic potential of human NSCLC (46). O-6-Methylguanine-DNA Methyltransferase (MGMT) catalyzes the transfer of methyl groups from O(6)alkylguanine and other methylated moieties of the DNA to its molecule. A low protein expression of MGMT was found in the bronchial epithelium of patients with lung cancer as compared to healthy controls, suggesting that there is a negative correlation between MGMT expression and lung cancer risk (47). DNA polymerase eta (POLH) is a DNA polymerase belonging to a subset of tumor suppressor proteins required for maintaining genome integrity (48). RAD1 encodes a component of a heterotrimeric cell cycle checkpoint complex, known as the 9-1-1 complex, that is activated to stop cell cycle progression in response to DNA damage or incomplete DNA replication. RAD17 is a cell cycle checkpoint gene required for cell cycle arrest and DNA damage repair in response to DNA damage. This protein recruits



the RAD1-RAD9-HUS1 checkpoint protein complex onto chromatin after DNA damage and initiates DNA repair.

Gene Set Enrichment Analysis (GSEA) is a computational method that determines whether a priori defined set of genes shows statistically significant, concordant differences between two biological states. GSEA showed that cellular response to hypoxia, MAPK signaling, noncanonical NF-KB signaling, and cell cycle pathways relating to cell proliferation was significantly enriched in the high-risk group, which might account for a higher malignancy and poorer prognosis of LUAD patients in the high-risk group. Oxygen deprivation (hypoxia) is a feature of solid tumors that promotes genomic instability, enhanced aggressiveness, and metastases and is an important factor in treatment resistance and poor survival (49). The MAPK pathways converge in the amplification of key molecules that sustain cell proliferation, growth, and survival processes (50, 51). Noncanonical NF-KB signaling contains NIK phosphorylates IKK/and helps IKK/to phosphorylate p100. Mutations in various upstream regulators (TRAF2, TRAF3, cIAP1&2, CD40) lead to increased stability of NIK and subsequent activation of the noncanonical NF-kB pathway, and this mechanism of activation appears to be important for different cancer types including DLBC and lung cancer (52). The human cell cycle is a tightly regulated process with checkpoints in place to ensure genomic integrity. Recent studies have shown that CDK4 and CDK6 inhibitors can promote T cell activation (53) and reverse T

cell exclusion, thus leading to a better response to ICIs (54). Taken together, this suggests that tumor cells in the high-risk group proliferated faster, leading to increased malignancy.

In addition, there were decreased B cells, CD8⁺ T cells, CD8⁺ central memory T cells, HSC, MDC, and immune scores found in the high-risk group. The CD8 T cell-dependent killing of cancer cells could produce interferon-gamma (IFN- γ) and then activate antitumor immunity (55). Myeloid dendritic cells (MDC) are crucial for the activation of antigen-specific CD8 T Cells. A recent study reported that anti-tumor effects of DCs can be reduced by a low DC count, low antigen presentation efficiency of tumor-infiltrating DCs, and a weak ability of DC to migrate into tumor (56). Many studies reported that B cell infiltration was associated with a favorable prognosis in NSCLC (57-60). Hematopoietic stem cells (HSC) are a very small group of source cells that can self-renew and generate various blood cells and immune cells. Tumor immune infiltrating cells migrate from blood to tumor tissues and play an important role in immune regulation. Lots of studies have shown that tumor immune infiltrating cells are closely related to the efficacy of ICIs and prognosis (61, 62).

Interestingly, we noticed that there were increased TMB, neoantigen, immune checkpoint molecules, and immunosuppression genes' expression in the high-risk group. Meanwhile, the expression of MHC II genes that express on antigen-presenting cells only in the low-risk group was significantly lower than in the high-risk group. It has been widely studied that higher TMB, neoantigen, and immune checkpoint molecules are indicators implicated in a better response to ICI treatment (8, 11, 12). Therefore, it is suggested that the high-risk group might be more sensitive to immunotherapies as compared to the low-risk group classified by the GDPLichi model.

We further examined genomic mutations in both the lowand high-risk groups and identified 12 candidates from the first top 25 mutated genes, whose mutation frequency has a significant difference between low- and high-risk groups classified by the GDPLichi model. Most of these genes are associated with TMB (35, 36), which could be used to predict the efficacy of immunotherapy.

The TIDE algorithm is a computational method that uses the expression profile of immune-related genes to predict the ICIs response. It is particularly successful in NSCLC and melanoma (22) and has exhibited a higher accuracy than PD-L1 expression level or TMB alone in predicting overall survival of patients treated with ICIs (13, 19, 20). Further analysis revealed that the TIDE scores in the high-risk group were significantly lower than the low-risk group, suggesting that patient of the high-risk group is more sensitive to response for ICIs. This conclusion was verified in the other external datasets (GSE31210, GSE50081, GSE30219).

In conclusion, we firstly identified two prognostically and clinically relevant subgroups of LUAD using the GDPLichi model which was constructed from seven DDR-risk genes. Patients from the high-risk group showed lower TIDE scores, and are thus more responsive to ICIs. The limit of this research was that it was retrospective, and results should thus be further confirmed by prospective studies.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

BH, YHL, and YLi designed the study. YLe and SD collected the data and performed the analysis. FY, TG, XX, YZ, LC, CFQ, NL, XZ, CL, YK, KH, and NW helped to analyze the data. YLe and SD wrote the paper. BH, TG, and XZ revised the paper. All the authors read and approved the final manuscript.

FUNDING

This research was supported by the National Natural Science Foundation of China (grant number 81971470 and 32170913) and Shenzhen Science and Technology Commission Fund (program number JCYJ20190809143803732 and JCYJ20210324120602008) to BH and Chinese Postdoctoral Science Foundation (grant number 2020M683062) to NL.

ACKNOWLEDGMENTS

We are grateful for Dr. ShiQiang Zhang's advice on organizing the results.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.733533/full#supplementary-material

Supplementary Figure S1 | Detailed flowchart of this study.

Supplementary Figure S2 | (A–G) Kaplan-Meier analysis of the genes included for GDPLichi construction (DUT, MGMT, POLH, RAD1, RAD17, TYMS, and YWHAG). (H) Example of calculating predicted survival rate related to .

Supplementary Figure S3 | (A) PCA based on the expression profile of the seven risk genes according to different risk groups and (B) correlation between the GDPLichi and the seven risk genes in the three GEO validation cohorts.

Supplementary Figure S4 | (A) Kaplan-Meier plots of the survival probability for low- and high-risk subgroups, (B) ROC curves of the GDPLichi score and seven risk genes of the classifier, (C) Forest plot representation of multivariate Cox model depicted association between overall survival and LUAD subgroups with other clinical factors considered in the three GEO validation cohorts.

Supplementary Figure S5 | (A–C) Statistical analysis of the expression of PD-L1, PDCD1, CTLA4 between low- and high-risk groups in the three GEO validation cohorts

Supplementary Figure S6 | (A) ROC curves for the performance of the GDPLichi score, TMB, PD-L1 expression, and neoantigen in predicting prognosis.
(B–E) Correlation between the TIDE score and GDPLichi score, TMB, PD-L1 expression, and neoantigen, respectively. (F) ROC curves for the performance of the original seven-gene score and three-gene combination (DUT, TYMS, and YWHAG) in predicting prognosis.

Supplementary File 1 | Gene list of DDR-related genes.

Supplementary File 2 | Univariate Cox regression result of DDR-related genes.

Supplementary File 3 | Multivariate Cox regression result of DDR-related genes.

Supplementary File 4 | Jianguoyun source data download link.

Supplementary File 5 | All GSEA significant pathways.

Supplementary File 6 | main_code.R (source analysis R code).

REFERENCES

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21660
- Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2020. CA Cancer J Clin (2020) 70:7–30. doi: 10.3322/caac.21590
- Reck M, Rodriguez-Abreu D, Robinson AG, Hui R, Csoszi T, Fulop A, et al. Pembrolizumab Versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. N Engl J Med (2016) 375:1823–33. doi: 10.1056/ NEJMoa1606774
- Herbst RS, Baas P, Kim DW, Felip E, Perez-Gracia JL, Han JY, et al. Pembrolizumab Versus Docetaxel for Previously Treated, PD-L1-Positive, Advanced non-Small-Cell Lung Cancer (KEYNOTE-010): A Randomised Controlled Trial. *Lancet Lond Engl* (2016) 387:1540–50. doi: 10.1016/S0140-6736(15)01281-7
- Wei SC, Duffy CR, Allison JP. Fundamental Mechanisms of Immune Checkpoint Blockade Therapy. *Cancer Discov* (2018) 8:1069–86. doi: 10.1158/2159-8290.CD-18-0367
- Akinleye A, Rasool Z. Immune Checkpoint Inhibitors of PD-L1 as Cancer Therapeutics. J Hematol Oncol J Hematol Oncol (2019) 12:92. doi: 10.1186/ s13045-019-0779-5
- Niu Y, Lin A, Luo P, Zhu W, Wei T, Tang R, et al. Prognosis of Lung Adenocarcinoma Patients With NTRK3 Mutations to Immune Checkpoint Inhibitors. Front Pharmacol (2020) 11:1213. doi: 10.3389/fphar.2020.01213
- Taube JM, Klein A, Brahmer JR, Xu H, Pan X, Kim JH, et al. Association of PD-1, PD-1 Ligands, and Other Features of the Tumor Immune Microenvironment With Response to Anti–PD-1 Therapy. *Clin Cancer Res* (2014) 20:5064–74. doi: 10.1158/1078-0432.CCR-13-3271
- Doroshow DB, Bhalla S, Beasley MB, Sholl LM, Kerr KM, Gnjatic S, et al. PD-L1 as a Biomarker of Response to Immune-Checkpoint Inhibitors. *Nat Rev Clin Oncol* (2021) 18:345–62. doi: 10.1038/s41571-021-00473-5
- Mehnert JM, Monjazeb AM, Beerthuijzen JMT, Collyar D, Rubinstein L, Harris LN. The Challenge for Development of Valuable Immuno-Oncology Biomarkers. *Clin Cancer Res* (2017) 23:4970–9. doi: 10.1158/1078-0432.CCR-16-3063
- Hollern DP, Xu N, Thennavan A, Glodowski C, Garcia-Recio S, Mott KR, et al. B Cells and T Follicular Helper Cells Mediate Response to Checkpoint Inhibitors in High Mutation Burden Mouse Models of Breast Cancer. *Cell* (2019) 179:1191–206.e21. doi: 10.1016/j.cell.2019.10.028
- Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational Landscape Determines Sensitivity to PD-1 Blockade in non–Small Cell Lung Cancer. Science (2015) 348:124–8. doi: 10.1126/science.aaa1348
- Hellmann MD, Ciuleanu TE, Pluzanski A, Lee JS, Otterson GA, Audigier-Valette C, et al. Nivolumab Plus Ipilimumab in Lung Cancer With a High Tumor Mutational Burden. N Engl J Med (2018) 378:2093–104. doi: 10.1056/ NEJMoa1801946
- Hellmann MD, Nathanson T, Rizvi H, Creelan BC, Sanchez-Vega F, Ahuja A, et al. Genomic Features of Response to Combination Immunotherapy in Patients With Advanced Non-Small-Cell Lung Cancer. *Cancer Cell* (2018) 33:843–52.e4. doi: 10.1016/j.ccell.2018.03.018
- Strickler JH, Hanks BA, Khasraw M. Tumor Mutational Burden as a Predictor of Immunotherapy Response: Is More Always Better? *Clin Cancer Res* (2021) 27:1236–41. doi: 10.1158/1078-0432.CCR-20-3054
- Bulik-Sullivan B, Busby J, Palmer CD, Davis MJ, Murphy T, Clark A, et al. Deep Learning Using Tumor HLA Peptide Mass Spectrometry Datasets Improves Neoantigen Identification. *Nat Biotechnol* (2018) 37, 55–63. doi: 10.1038/nbt.4313
- Teo MY, Bambury RM, Zabor EC, Jordan E, Al-Ahmadie H, Boyd ME, et al. DNA Damage Response and Repair Gene Alterations Are Associated With Improved Survival in Patients With Platinum-Treated Advanced Urothelial Carcinoma. *Clin Cancer Res* (2017) 23:3610–8. doi: 10.1158/1078-0432.CCR-16-2520
- Parikh AR, He Y, Hong TS, Corcoran RB, Clark JW, Ryan DP, et al. Analysis of DNA Damage Response Gene Alterations and Tumor Mutational Burden Across 17,486 Tubular Gastrointestinal Carcinomas: Implications for Therapy. *Oncologist* (2019) 24:1340–7. doi: 10.1634/theoncologist.2019-0034

- Wang Z, Zhao J, Wang G, Zhang F, Zhang Z, Zhang F, et al. Comutations in DNA Damage Response Pathways Serve as Potential Biomarkers for Immune Checkpoint Blockade. *Cancer Res* (2018) 78:6486–96. doi: 10.1158/0008-5472.CAN-18-1814
- Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniack AD, et al. Genomic and Molecular Landscape of DNA Damage Repair Deficiency Across The Cancer Genome Atlas. *Cell Rep* (2018) 23:239– 54.e6. doi: 10.1016/j.celrep.2018.03.076
- 21. Teo MY, Seier K, Ostrovnaya I, Regazzi AM, Kania BE, Moran MM, et al. Alterations in DNA Damage Response and Repair Genes as Potential Marker of Clinical Benefit From PD-1/PD-L1 Blockade in Advanced Urothelial Cancers. J Clin Oncol (2018) 36:1685–94. doi: 10.1200/JCO.2017.75.7740
- 22. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T Cell Dysfunction and Exclusion Predict Cancer Immunotherapy Response. Nat Med (2018) 24:1550–8. doi: 10.1038/s41591-018-0136-1
- Kaderbhaï C, Tharin Z, Ghiringhelli F. The Role of Molecular Profiling to Predict the Response to Immune Checkpoint Inhibitors in Lung Cancer. Cancers (2019) 11(2):201. doi: 10.3390/cancers11020201
- Wang S, He Z, Wang X, Li H, Liu X-S. Antigen Presentation and Tumor Immunogenicity in Cancer Immunotherapy Response Prediction. *eLife* (2019) 8:e49020. doi: 10.7554/eLife.49020
- Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating Single-Cell Transcriptomic Data Across Different Conditions, Technologies, and Species. *Nat Biotechnol* (2018) 36:411–20. doi: 10.1038/nbt.4096
- Bretz AC, Parnitzke U, Kronthaler K, Dreker T, Bartz R, Hermann F, et al. Domatinostat Favors the Immunotherapy Response by Modulating the Tumor Immune Microenvironment (TIME). J Immunother Cancer (2019) 7:294. doi: 10.1186/s40425-019-0745-3
- Pallocca M, Angeli D, Palombo F, Sperati F, Milella M, Goeman F, et al. Combinations of Immuno-Checkpoint Inhibitors Predictive Biomarkers Only Marginally Improve Their Individual Accuracy. J Transl Med (2019) 17:131. doi: 10.1186/s12967-019-1865-8
- George AP, Kuzel TM, Zhang Y, Zhang B. The Discovery of Biomarkers in Cancer Immunotherapy. *Comput Struct Biotechnol J* (2019) 17:484–97. doi: 10.1016/j.csbj.2019.03.015
- Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, et al. Identification of Genes Upregulated in ALK-Positive and EGFR/KRAS/ALK-Negative Lung Adenocarcinomas. *Cancer Res* (2012) 72:100–11. doi: 10.1158/ 0008-5472.CAN-11-1403
- Rousseaux S, Debernardi A, Jacquiau B, Vitte AL, Vesin A, Nagy-Mignotte H, et al. Ectopic Activation of Germline and Placental Genes Identifies Aggressive Metastasis-Prone Lung Cancers. *Sci Transl Med* (2013) 5:186ra66. doi: 10.1126/scitranslmed.3005723
- Der SD, Sykes J, Pintilie M, Zhu CQ, Strumpf D, Liu N, et al. Validation of a Histology-Independent Prognostic Gene Signature for Early-Stage, non-Small-Cell Lung Cancer Including Stage IA Patients. J Thorac Oncol (2014) 9:59–64. doi: 10.1097/JTO.000000000000042
- Yu G, Wang L-G, Han Y, He Q-Y. Clusterprofiler: An R Package for Comparing Biological Themes Among Gene Clusters. Omics J Integr Biol (2012) 16:284–7. doi: 10.1089/omi.2011.0118
- Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res* (2020) 48:D498–503. doi: 10.1093/nar/gkz1031
- Aran D, Hu Z, Butte AJ. Xcell: Digitally Portraying the Tissue Cellular Heterogeneity Landscape. *Genome Biol* (2017) 18:220. doi: 10.1186/s13059-017-1349-1
- Oh J-H, Jang SJ, Kim J, Sohn I, Lee JY, Cho EJ, et al. Spontaneous Mutations in the Single TTN Gene Represent High Tumor Mutation Burden. NPJ Genomic Med (2020) 5:33. doi: 10.1038/s41525-019-0107-6
- Li X, Pasche B, Zhang W, Chen K. Association of MUC16 Mutation With Tumor Mutation Load and Outcomes in Patients With Gastric Cancer. JAMA Oncol (2018) 4:1691–8. doi: 10.1001/jamaoncol.2018.2805
- Yang Y, Zhang J, Chen Y, Xu R, Zhao Q, Guo W, et al. MUC4, MUC16, and TTN Genes Mutation Correlated With Prognosis, and Predicted Tumor Mutation Burden and Immunotherapy Efficacy in Gastric Cancer and Pan-Cancer. *Clin Transl Med* (2020) 10:e155. doi: 10.1002/ctm2.155
- Wang Z, Wang C, Lin S, Yu X. Effect of TTN Mutations on Immune Microenvironment and Efficacy of Immunotherapy in Lung

Adenocarcinoma Patients. Front Oncol (2021) 11:725292. doi: 10.3389/fonc. 2021.725292

- Zhang L, Han X, Shi Y. Association of MUC16 Mutation With Response to Immune Checkpoint Inhibitors in Solid Tumors. JAMA Netw Open (2020) 3: e2013201. doi: 10.1001/jamanetworkopen.2020.13201
- Liu P, Morrison C, Wang L, Xiong D, Vedell P, Cui P, et al. Identification of Somatic Mutations in non-Small Cell Lung Carcinomas Using Whole-Exome Sequencing. *Carcinogenesis* (2012) 33:1270–6. doi: 10.1093/carcin/bgs148
- Terzic J, Seipel A, Dubuisson J, Tille JC, Tsantoulis P, Addeo A, et al. Sustained Response to Pembrolizumab Without Prior Chemotherapy in High-Grade Serous Ovarian Carcinoma With CSMD3 Mutation. *Gynecol* Oncol Rep (2020) 33:100600. doi: 10.1016/j.gore.2020.100600
- 42. Qing T, Zhu S, Suo C, Zhang L, Zheng Y, Shi L, et al. Somatic Mutations in ZFHX4 Gene are Associated With Poor Overall Survival of Chinese Esophageal Squamous Cell Carcinoma Patients. *Sci Rep* (2017) 7:4951. doi: 10.1038/s41598-017-04221-7
- 43. Wang C, Liang H, Lin C, Li F, Xie G, Qiao S, et al. Molecular Subtyping and Prognostic Assessment Based on Tumor Mutation Burden in Patients With Lung Adenocarcinomas. Int J Mol Sci (2019) 20. doi: 10.3390/ijms20174251
- Mohamedi Y, Fontanil T, Cal S, Cobo T, Obaya ÁJ. ADAMTS-12: Functions and Challenges for a Complex Metalloprotease. *Front Mol Biosci* (2021) 8:686763. doi: 10.3389/fmolb.2021.686763
- Rose MG, Farrell MP, Schmitz JC. Thymidylate Synthase: A Critical Target for Cancer Chemotherapy. *Clin Colorectal Cancer* (2002) 1:220–9. doi: 10.3816/ CCC.2002.n.003
- 46. Raungrut P, Wongkotsila A, Champoochana N, Lirdprapamongkol K, Svasti J, Thongsuksai P. Knockdown of 14-3-3γ Suppresses Epithelial–Mesenchymal Transition and Reduces Metastatic Potential of Human Non-Small Cell Lung Cancer Cells. *Anticancer Res* (2018) 38:3507–14. doi: 10.21873/anticanres.12622
- Povey AC, Margison GP, Santibáñez-Koref MF. Lung Cancer Risk and Variation in MGMT Activity and Sequence. DNA Repair (2007) 6:1134–44. doi: 10.1016/j.dnarep.2007.03.022
- Barnes RP, Tsao W-C, Moldovan G-L, Eckert KA. DNA Polymerase Eta Prevents Tumor Cell-Cycle Arrest and Cell Death During Recovery From Replication Stress. *Cancer Res* (2018) 78:6549–60. doi: 10.1158/0008-5472.CAN-17-3931
- Ren W, Mi D, Yang K, Cao N, Tian J, Li Z, et al. The Expression of Hypoxia-Inducible Factor-1α and its Clinical Significance in Lung Cancer: A Systematic Review and Meta-Analysis. Swiss Med Wkly (2013) 143:w13855. doi: 10.4414/smw.2013.13855
- Plotnikov A, Flores K, Maik-Rachline G, Zehorai E, Kapri-Pardes E, Berti DA, et al. The Nuclear Translocation of ERK1/2 as an Anticancer Target. *Nat Commun* (2015) 6:6685. doi: 10.1038/ncomms7685
- Chapnick DA, Warner L, Bernet J, Rao T. & Liu, X. Partners in Crime: The Tgfβ and MAPK Pathways in Cancer Progression. *Cell Biosci* (2011) 1:42. doi: 10.1186/2045-3701-1-42
- Cildir G, Low KC, Tergaonkar V. Noncanonical NF-κb Signaling in Health and Disease. Trends Mol Med (2016) 22:414–29. doi: 10.1016/j.molmed.2016.03.002
- Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and Genetic Properties of Tumors Associated With Local Immune Cytolytic Activity. *Cell* (2015) 160:48–61. doi: 10.1016/j.cell.2014.12.033

- 54. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su MJ, Melms JC, et al. A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. *Cell* (2018) 175:984–97.e24. doi: 10.1016/ j.cell.2018.09.006
- Ribas A, Hu-Lieskovan S. What Does PD-L1 Positive or Negative Mean? J Exp Med (2016) 213:2835–40. doi: 10.1084/jem.20161462
- Shang N, Figini M, Shangguan J, Wang B, Sun C, Pan L, et al. Dendritic Cells Based Immunotherapy. Am J Cancer Res (2017) 7:2091–102.
- Wang S, Liu SS, Ly W, Xu D, Qu H, Zhang L, et al. Tumor-Infiltrating B Cells: Their Role and Application in Anti-Tumor Immunity in Lung Cancer. *Cell Mol Immunol* (2019) 16:6–18. doi: 10.1038/s41423-018-0027-x
- Liu X, Xu J, Zhang B, Liu J, Liang C, Meng Q, et al. The Reciprocal Regulation Between Host Tissue and Immune Cells in Pancreatic Ductal Adenocarcinoma: New Insights and Therapeutic Implications. *Mol Cancer* (2019) 18:184. doi: 10.1186/s12943-019-1117-9
- Gottlin EB, Bentley RC, Campa MJ, Pisetsky DS, Herndon JE 2nd, Patz EF Jr, et al. The Association of Intratumoral Germinal Centers With Early-Stage non-Small Cell Lung Cancer. J Thorac Oncol (2011) 6:1687–90. doi: 10.1097/ JTO.0b013e3182217bec
- 60. Germain C, Gnjatic S, Tamzalit F, Knockaert S, Remark R, Goc J, et al. Presence of B Cells in Tertiary Lymphoid Structures is Associated With a Protective Immunity in Patients With Lung Cancer. Am J Respir Crit Care Med (2014) 189:832–44. doi: 10.1164/rccm.201309-1611OC
- Al-Shibli KI, Donnem T, Al-Saad S, Persson M, Bremnes RM, Busund LT. Prognostic Effect of Epithelial and Stromal Lymphocyte Infiltration in non-Small Cell Lung Cancer. *Clin Cancer Res* (2008) 14:5220–7. doi: 10.1158/1078-0432.CCR-08-0133
- 62. Kinoshita T, Muramatsu R, Fujita T, Nagumo H, Sakurai T, Noji S, et al. Prognostic Value of Tumor-Infiltrating Lymphocytes Differs Depending on Histological Type and Smoking Habit in Completely Resected non-Small-Cell Lung Cancer. Ann Oncol (2016) 27:2117–23. doi: 10.1093/annonc/ mdw319

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Leng, Dang, Yin, Gao, Xiao, Zhang, Chen, Qin, Lai, Zhan, Huang, Luo, Kang, Wang, Li, Liang and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.