Received: 23 April 2020 | Revised: 7 June 2020 | Accepted: 3 July 2020

DOI: 10.1111/cas.14570

ORIGINAL ARTICLE

Cancer Science WILEY

Relapse-related molecular signature in early-stage lung adenocarcinomas based on base excision repair, stimulator of interferon genes pathway and tumor-infiltrating lymphocytes

Bo Yang | Wen Rao | Hao Luo | Liang Zhang | Dong Wang

Cancer Center, Daping Hospital & Army Medical Center of PLA, Third Military Medical University (Army Medical University), Chongqing, China

Correspondence
Dong Wang, Cancer Center, Daping Hospital & Army Medical Center of PLA, Third Military Medical University (Army Medical University) Chongqing, 400042, China.
Email: dongwang64@hotmail.com

Funding information
National Natural Science,CN, Grant/Award Number: NSFC 81772495

Abstract

Approximately 30% of patients with early-stage non–small cell lung cancer (NSCLC) relapse within 5 years after surgery. Therefore, it is necessary to identify a robust and reliable prognostic signature for early-stage NSCLC. Immunohistochemistry data from 147 patients with stage I lung adenocarcinoma (stage I-LUAD) were analyzed for the protein expression of base excision repair (BER), stimulator of interferon genes (STING) and tumor-infiltrating lymphocytes (TIL) to explore the relationship between protein expression and prognosis. A prediction model was further established by nomogram and externally verified using The Cancer Genome Atlas and Gene Expression Omnibus (GEO) databases. XRCC1 and H2AX are negative prognostic markers for relapse-free survival (RFS), while CD8, CD20 and STING are positive prognostic markers for RFS. Nomograms for RFS share common prognostic markers, including XRCC1, H2AX, STING, CD8 and CD20. The c-index was 0.724 and 0.698 in the training cohort and the internal validation cohort, respectively. It was externally verified that the nomogram model had a good prediction for recurrence of stage I-LUAD. Correlation analysis showed that APE1 and H2AX were negatively correlated with STING, while STING was positively correlated with TIL. BER, the STING pathway and TIL were associated with early recurrence and were correlated with the tissue expression of stage I-LUAD. Our nomogram model was a good predictor for recurrence of stage I-LUAD.

KEYWORDS

BER, lung adenocarcinoma, nomogram, predictive, recurrence, STING, TIL

1 | INTRODUCTION

Lung cancer is the leading cause of cancer morbidity and mortality worldwide.[1,2] However, many lung cancers can now be diagnosed at an early stage due to advances in diagnostic techniques. Surgery is currently the main treatment option for early-stage non–small cell lung cancer (NSCLC), with only stage IB patients with high-risk factors needing postoperative adjuvant therapy[3]

Bo Yang and Wen Rao contributed equally

and approximately 30% of early-stage NSCLC patients relapsing within 5 years after surgery.[4,5] However, the risk factors for treatment failure after surgery and the clinicopathologic characteristics of patients with a high risk of tumor recurrence have not been fully studied.

At present, most of the research on the prediction of lung cancer recurrence is screened through big data and the selected factors are verified. These predictive factors include protein, mRNA, DNA methylation and mutation.[6-9] Some key factors can be obtained using this method. The pathways in which these factors are located can also be easily ignored, while changes in these pathways may be based on the predicted values.

Recent studies have shown that tumor-infiltrating lymphocytes (TIL) are a promising biomarker for prognosis of lung cancer and high infiltrating lymphocytes are associated with a good prognosis.[10,11] TIL is mainly regulated by cytokines secreted by tumor cells in the tumor microenvironment (TME), in which process type I interferon (I-IFN) plays an important role.[12-15] The I-IFN signaling pathway is essential for production of cytokines, death of immunogenic tumor cells, antigen presentation and maturation of dendritic cells.[16] The STING pathway, which is considered to be the main source of the production of I-IFN in the TME,[17-19] is activated by cyclic dinucleotides (CDN) in the cytosol. Cytosolic CDN can be recognized by cyclic GMP-AMP synthase (cGAS) to form cyclic GMP-AMP (cGAMP), which then binds to STING. The complex subsequently facilitates the phosphorylation of interferon regulatory factor 3 (IRF3) by TANK-binding kinase 1 (TBK1) and then upregulates the production of I-IFN. The STING pathway is responsible for the activation of innate immunity after virus infection through the production of I-IFN. Recent studies have further found that activators of the STING pathway may also include leakage of DNA from the nucleus of the host cell, which may be the result of cell division or DNA damage.[20-22]

DNA damage repair has been thought to be associated with tumorigenesis for a long time. Recently, increasing research has been focused on DNA damage repair and tumor immunoregulation and immunotherapy. Some studies have shown that genomic DNA damage can affect the activation of certain immune factors and further participate in innate and adaptive immunity.[23-27] The base excision repair (BER) pathway, a highly conserved pathway in different species, is responsible for repairing endogenous DNA damage, including deamination, deanine, alkylation and excessive oxidative damage.[28,29] There are five major enzymes, including APE1, OGG1, XRCC1, pol β and NTH1, as well as other enzymes involved in this process.[30]

Increasing evidence suggests that DNA damage caused by carcinogens or radiation and leakage of mitochondrial DNA are the sources of cytoplasmic DNA, which can trigger the STING pathway and subsequent immune response in cancer cells. This clearly indicated a strong link between DNA damage and cancer immune response mediated by the STING pathway.[31] Based on this evidence, we assume that BER dysfunction affects the STING pathway by generating excessive dsDNA and further affects the recruitment of TIL in TME. Therefore, the protein expression level of BER, STING and TIL as a markers of relapse and the correlation of BER, STING pathway and TIL in stage I-LUAD were evaluated in this study.

2 | MATERIALS AND METHODS

2.1 | Study design and patients

A total of 178 patients with pathologically confirmed primary stage I-LUAD were enrolled in this retrospective study. Later, 29 patients were excluded due to loss of follow up and 2 were excluded because of unqualified paraffin samples. Finally, 147 patients were included in this study. A total of 100 patients from 2010 to 2012 were set as the training cohort, and 2013 patients were set as the internal validation cohort (47 cases). A postoperative follow up was conducted according to NCCN guidelines: medical history, physical examination and chest CT contrast examination were performed every 6 months, and then history, physical examination and a low-dose non–contrast-enhanced chest CT were performed every year for a total of 2-3 years. At the same time, patients were followed up by phone every 2 months and clinical data, including patients' medical history, therapeutic effect and overall survival time, were collected from the case system of the Oncology Department of Daping Hospital. Data were collected from Daping Hospital between March 2010 and December 2013. This study was conducted in accordance with the Declaration of Helsinki and the protocol was approved by the Ethics Committee of Daping Hospital (#2019-103).

2.2 | Immunohistochemistry

Paraffin specimens were drilled in the tumor area by 3 mm trephine (Minicore) and re–embedded. Tissue microarrays were constructed as previously described.[32] We followed routine immunohistochemical methods in our laboratory[33]: The sections were deparaffinized, hydrated, subjected to HIER for 2 minutes with 10 mmol/L sodium EDTA (pH 9.0) and treated with 3% H_2O_2–methanol solution for 10 minutes to reduce endogenous peroxidase activity. Primary antibodies (Table S1) were incubated overnight at 4°C in a humidified chamber. Afterwards, biotinylated rabbit anti–mouse IgG/anti–rabbit IgG secondary antibody was applied for 30 minutes at 37°C. Diaminobenzidine (DAB) showed color, subsequently hematoxylin counterstain.

Immunoscore was independently assessed by two pathologists, blinded to clinical outcomes. If the results were inconsistent, the slides were re–examined under a multi-head microscope and discussed to determine the final score. Semi-quantitative score = number of staining cells × intensity of cell staining. An expression with a total score of 4 points is defined as low expression, and that above 5 points as high expression. The number of staining cells is as follows: 1 = number of staining cells is <1/3 of the observation cells; 2 = number of staining cells is between 1/3 and 2/3 of the observation cells; 3 = number of staining cells is

>2/3 of the observation cells. The intensity of cell staining is as follows: 3 = intense; 2 = moderate; 1 = faint staining in ≥ 10% of tumor cells; 0 = no staining. Lymphocytes were counted in two locations in each tumor: the intratumoral compartment (within the tumor cell nests) and within the adjacent stroma. A score of 1+ (<30%) was considered low abundance TIL; 2+ (30%-60%) was moderate; and 3+ (>60%) was a marked increase in TIL; an expression with a score above 2 is defined as high expression of TIL. Immunohistochemistry expression was divided into high-expression and low-expression groups according to our previous scoring method (Figure 1).[49]

2.3 | Analysis of The Cancer Genome Atlas and Gene Expression Omnibus data

The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO) datasets were used for external validation. TCGA data for LUAD were obtained from http://gdac.broadinstitute.org and https://genome-cancer.ucsc.edu, respectively. After removing patients with incomplete follow-up data, 236 stage I-LUAD patients were finally included. Gene expression data and corresponding clinical information data for East Asian NSCLC patients were obtained from the publicly available database Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). All microarray data were normalized using robust multi-array average (RMA) and microarray Suite 5 (MAS5) methods and calculated by log2 scale. In addition, two independent datasets (GSE31210 and GSE13213) were used in this study, with 226 LUAD patients and 117 LUAD patients included, respectively. Finally, only stage I-LUAD data were selected for analysis. Thus, 168 and 79 cases from GSE31210 and GSE13213 datasets, respectively, were used. The detailed clinical data of the patients are shown in Table 1.

The X-tile software (Yale University, New Haven, CT, USA) was used to provide the best cutoff points. According to the optimal cutoff value of X-tile, patients in the database is divided into high-expression and low-expression groups.

2.4 | Statistical analysis

Data analysis was performed using SPSS (Version 21.0, SPSS, Chicago, IL) or GraphPad Prism 7 (GraphPad Software, La Jolla, CA, USA). Cumulative survival probabilities (survival rates) were analyzed using the Kaplan-Meier method and significant differences between the survival rates were tested using the log-rank test. Univariate and multivariate Cox regression analyses and the receiver under the operator characteristic (ROC) curve were calculated by SPSS. Hazard ratios (HR) and 95% confidence intervals (95% CI) were presented for the associations between dependent and independent variables. The χ^2-test was used to assess differences between the two groups. All statistical assessments were two-tailed and considered significant at $P < 0.05$.

Prognostic models were computed with an adaptive elastic net approach for censored data using the R-package "rms" (Version 5.1-3.1) and were illustrated by nomograms. Hazard ratios (HR) and corresponding 95% confidence intervals (95% CI) were calculated using the R-package "forestplot" (Version 1.74). To validate the nomogram, discrimination was evaluated using the concordance index (C-index), which estimates the probability of concordance between the predicted and observed responses. The model predictor was further dichotomized into high-risk and low-risk groups to illustrate the impact

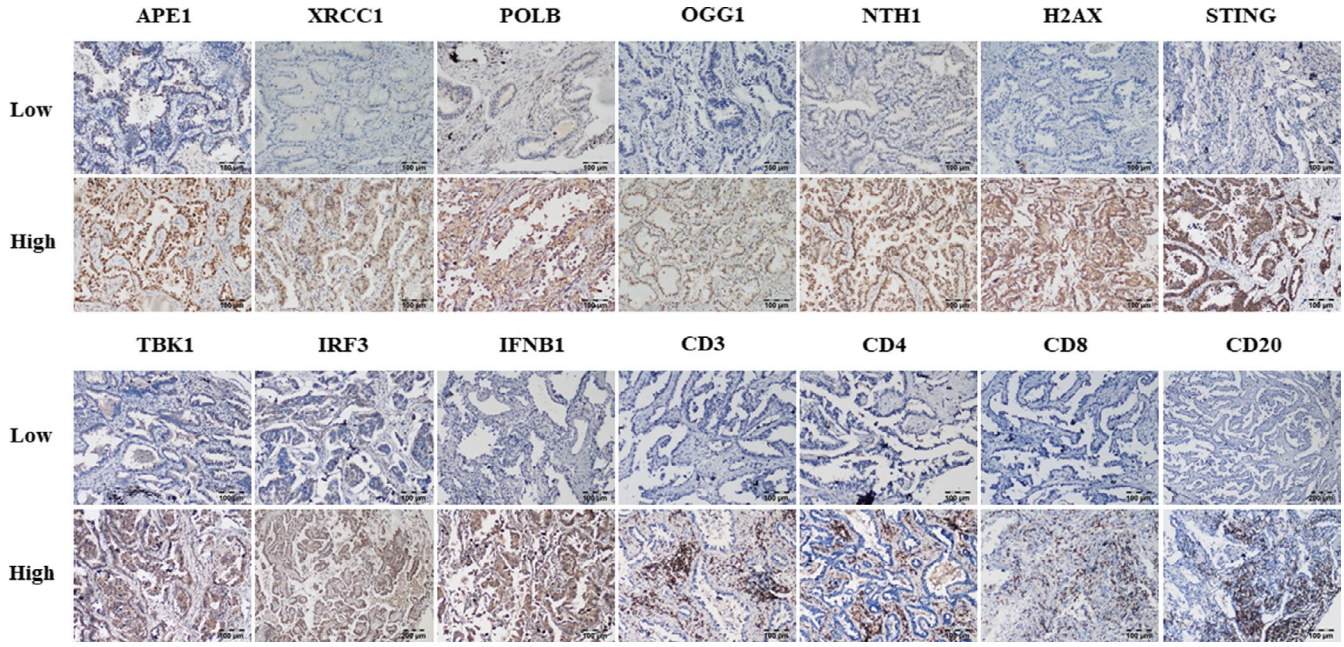

FIGURE 1 Immunohistochemical expression of APE1, XRCC1, Polβ, OGG1, NTH1, H2AX, STING, TBK1, IRF3, IFNB1, CD3, CD4, CD8 and CD20 in lung adenocarcinoma (LUAD) tissues. (bar = 100 μm)

TABLE 1 Clinical characteristics of the Daping, The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO) datasets

Clinicopathological characteristics	Daping Hospital cases n (%) n = 147	TCGA cases n (%) n = 236	GSE31210 cases n (%) n = 168	GSE13213 cases n (%) n = 79
Age median (range)	59 (37-80)	67 (38-88)	61 (30-76)	61 (35-84)
Gender				
Male	75 (51.02)	99 (41.95)	71 (42.26)	41 (51.90)
Female	72 (48.98)	137 (58.05)	97 (57.74)	38 (48.10)
Smoking history				
Smoking	42 (28.57)	198 (83.90)	74 (44.05)	40 (50.63)
Non–smoking	105 (71.43)	34 (16.10)	94 (55.95)	39 (49.37)
Primary location				
Left lung	63 (42.86)	92 (38.98)		
Right lung	84 (57.14)	141 (59.74)		
Unknown		3 (1.28)	168 (100)	79 (100)
Stage				
IA	58 (39.46)	118 (50.00)		42 (53.16)
IB	89 (60.54)	113 (47.88)		37 (46.84)
I		5 (2.12)	168 (100)	
Gene mutation				
ALK	10 (6.80)	2 (0.85)	3 (1.78)	/
KRAS	7 (4.76)	29 (12.29)	14 (8.33)	11 (13.92)
EGFR	44 (29.93)	14 (5.93)	103 (61.31)	29 (36.71)
EGFR/KRAS/ALK-	50 (34.02)	58 (24.58)	48 (28.58)	43 (54.43)
Unknown	36 (24.49)	133 (56.35)		

of each final model on clinical outcome through Kaplan-Meier survival estimates. Pearson correlation coefficient, for analysis of TCGA and GEO databases as well as immunohistochemistry (IHC) markers, was calculated using the intrinsic R commands and the R-package "corrplot" (Version 0.84). *P*-values of the "corrplot" were corrected in accordance with the false discovery rate.

3 | RESULTS

3.1 | Characteristics of the patients

A total of 147 patients were enrolled from Daping Hospital, including 75 men and 72 women, with a median age of 59 years. Among them, smokers accounted for 28.57% and stage IB patients 60.54%. Driver gene mutations were analyzed in 111 patients, including 44 patients with EGFR mutation, 10 patients with ALK fusion and 7 patients with KRAS mutation.

A total of 236 stage I-LUAD patients were examined from TCGA database, including 99 men and 137 women, with ages ranging from 38 to 88 years (mean age: 67). Stage IA patients accounted for 50% of patients. The median follow-up time was 20.4 (0.5-224.4) months. A total of 168 and 79 patients were analyzed from GSE31210 and GSE13213 datasets, respectively. In the GSE13213 dataset, female patients accounted for 48.1%, and stage IA patients 53.16%. In the GSE31210 dataset, female patients accounted for 57.7%, but there were no stage IA and IB patients. In addition, the driver gene mutation rates of the patients were 71.42% and 45.47% in GSE31210 and GSE13213 datasets, respectively. The median follow-up time in the GSE31210 and GSE13213 was 58.9 (6.6-115.5) months and 68.9 (6.2-109.8) months, respectively. The descriptive statistics are shown in Table 1.

3.2 | Protein expression pattern of base excision repair, stimulator of interferon genes pathway and tumor-infiltrating lymphocytes by immunohistochemistry-based tissue microarray

We first compared the expression of these makers in tumor and adjacent normal tissues in TCGA database. The results showed that the expression of markers, except for XRCC1, TBK1 and CD8A, in tumor and adjacent normal tissues was statistically different (Figure S1).

The protein expression of BER, STING pathway and TIL in 147 early-stage lung cancer tissues was analyzed by IHC-based tissue microarray (Figure 1). APE1, OGG1, XRCC1, pol β, NTH1 and H2AX were expressed in tumor cells and stromal cells, among which APE1 was expressed in nucleus and/or cytoplasm, and the rest in nucleus. The expression of STING, IRF3 and TBK1 in tumor cells

was significantly higher than that in stromal cells, among which IRF3 was expressed in nucleus and cytoplasm, and STING and TBK1 in cytoplasm. IHC results showed that APE1 (92/147), NTH1 (97/147), STING (88/147) and IRF3 (79/147) were highly expressed, while polβ (60/147), OGG1 (62/147), H2AX (57/147) and TBK1 (58/147) were downregulated. Afterward, the correlation between clinical characteristics and expression profiles was analyzed. The results showed that XRCC1 and STING were differentially expressed in the age and smoking groups. The proportion of upregulated XRCC1 in patients over 60 years old was higher than that in younger patients, while the proportion of highly expressed STING in smoking patients was lower than that in non–smoking patients. We also found different expressions of NTH1, H2AX, TBK1, CD8 and CD20 in stage IA and stage IB patients, among which NTH1, CD8 and CD20 were highly expressed in stage IB patients, while H2AX and TBK1 had low expression in stage IB patients (Table S2).

3.3 | Impact of immunohistochemistry markers on the prognosis of relapse-free survival

Overall, 73 patients relapsed at the end of follow up. The median follow-up time for RFS was 49 months (95% CI: 33.10-64.90). Among them, 23 patients (15.65%) relapsed within 1 year, 43 patients (29.25%) relapsed within 2 years and 59 patients (40.13%) relapsed within 3 years. The median follow-up time was 46 months, ranging from 5 to 87 months. A total of 73 patients relapsed or died at the end of the follow up.

The Kaplan-Meier (K-M) method and log-rank test were used to compare the RFS between high-expression and low-expression groups divided based on IHC scores. The results showed that the low expression of CD3 ($P < 0.001$), CD4 ($P = 0.008$) and CD8 ($P = 0.0081$) and the high expression of polβ ($P = 0.0079$), XRCC1 ($P = 0.029$), H2AX ($P = 0.0036$) and TBK1 ($P = 0.021$) were significantly associated with poor RFS (Figure 2).

3.4 | Development of prognostic nomogram for relapse-free survival and validation of predictive accuracy

The relationship between prognostic markers and RFS was investigated by univariate and multivariate COX regression. The results of univariate COX demonstrated that XRCC1 (HR: 1.865, 95% CI: 1.145-3.152, $P = 0.024$), polβ (HR: 1.755, 95% CI: 0.984-2.857, $P = 0.060$), H2AX (HR: 2.000, 95% CI: 1.131-3.542, $P = 0.015$), STING (HR: 0.657, 95% CI: 0.387-1.144, $P = 0.041$), CD3 (HR: 0.565, 95% CI: 0.324-0.993, $P = 0.048$), CD8 (HR: 0.465, 95% CI: 0.274-0.794, $P = 0.0046$) and CD20

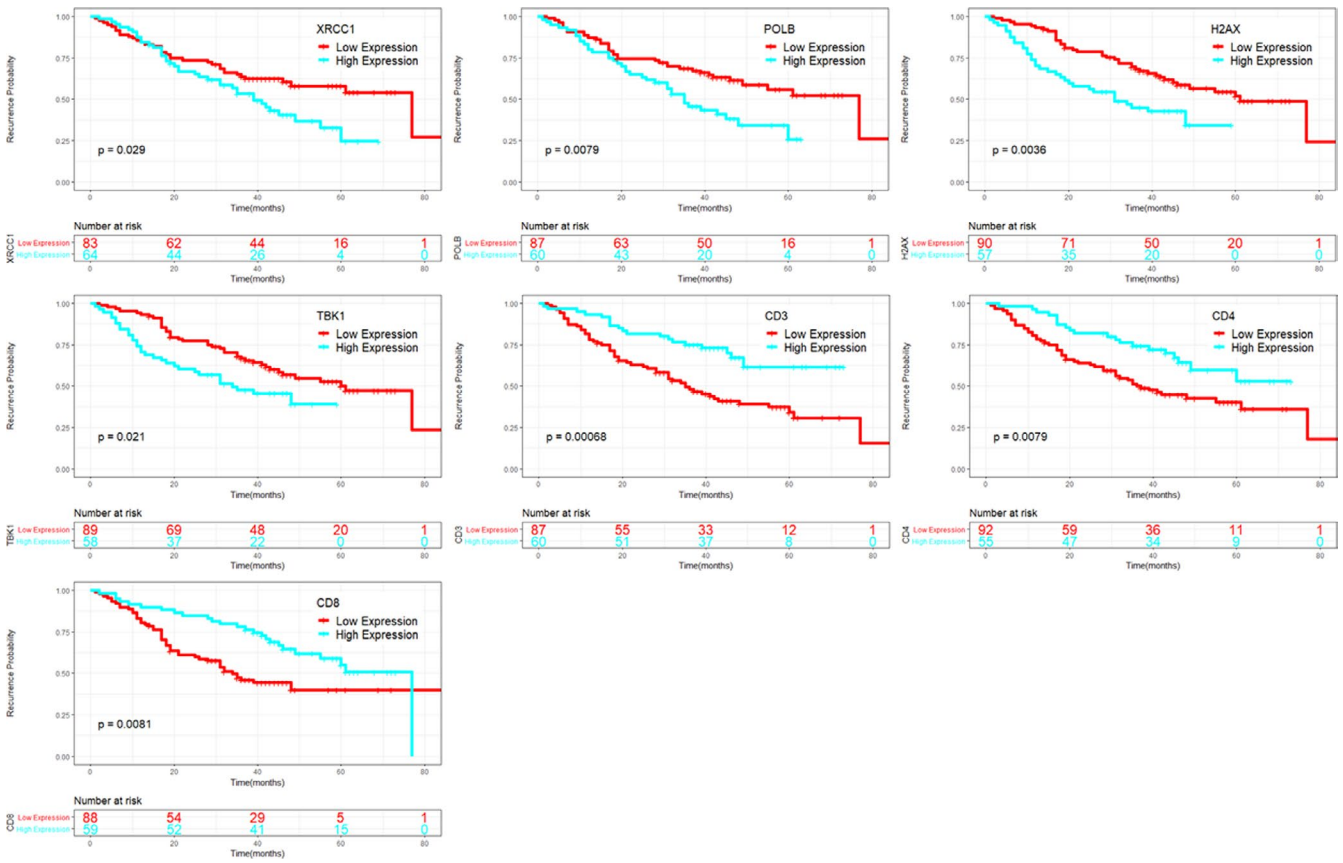

FIGURE 2 Kaplan-Meier survival analysis of relapse-free survival (RFS) for high and low expression of prognostic makers. Blue line indicates high expression, and red line indicates low expression

(HR: 0.564, 95% CI: 0.330-0.973, $P = 0.04$) were associated with RFS. Multivariate analysis further revealed that XRCC1, H2AX, STING, CD8 and CD20 were independent prognostic factors for PFS. The detailed prognostic markers are shown in Table 2.

Then a prognostic nomogram was developed to predict RFS based on the results of multivariate COX regression of the training cohort, including XRCC1, H2AX, STING, CD8 and CD20 (Figure 3A). A nomogram integrating the prognostic factors was constructed based on C-index values. Risk points of the nomogram were added to calculate the total score. A calibration curve was then generated to calculate the 1-year, 3-year and 5-year RFS (Figure 3B-D), the C-index value of this model was 0.724. The bootstrap-corrected c-indexes (0.698) of the internal validation cohort were close to those of the nomograms. Both models exhibited good validation. Patients of the internal validation cohort were further divided into high-risk and low-risk groups according to the statistically significant model, which indicated that the RFS probability between the two groups was significantly different (Figure 4A). The risk scores of each patient in the external validation cohort were calculated according to the constructed model, where risk score = XRCC1 × 88 + H2AX × 57 + (1-STING) × 80 + (1-CD 8) × 100 + (1-CD 20) × 84. Patients were divided into high-risk and low-risk groups based on the best cutoff value of X-tile. Survival analysis performed using the Kaplan-Meier method and the log-rank test showed significant differences in TCGA and GSE31210. (Figure 4B-D). Univariate

TABLE 2 Univariate and multivariate COX regression analysis of the prognosis of immunohistochemistry (IHC) markers for relapse-free survival (RFS) in the training cohort

	Univariate			Multivariate		
Variables	HR	95% CI	*P*-value	HR	95% CI	*P*-value
APE1	1.313	0.714-2.221	0.43			
XRCC1	1.865	1.145-3.152	0.024	1.948	1.123-3.378	0.017
POLB	1.755	0.984-2.857	0.06	1.287	0.722-2.293	0.391
NTH1	0.852	0.461-1.457	0.44			
OGG1	0.881	0.533-1.623	0.67			
STING	0.657	0.387-1.144	0.041	0.554	0.323-0.949	0.031
TBK1	1.635	0.932-2.821	0.089	0.437	0.148-1.292	0.134
IRF3	0.883	0.519-1.509	0.63			
IFNB1	1.297	0.745-2.113	0.39			
CD3	0.565	0.324-0.993	0.048			
CD4	0.664	0.382-1.238	0.15			
CD8	0.465	0.274-0.794	0.0046	0.438	0.238-0.807	0.008
CD20	0.564	0.330-0.973	0.04	0.523	0.297-0.920	0.024
H2AX	2.000	1.131-3.542	0.015	2.918	0.956-8.909	0.021

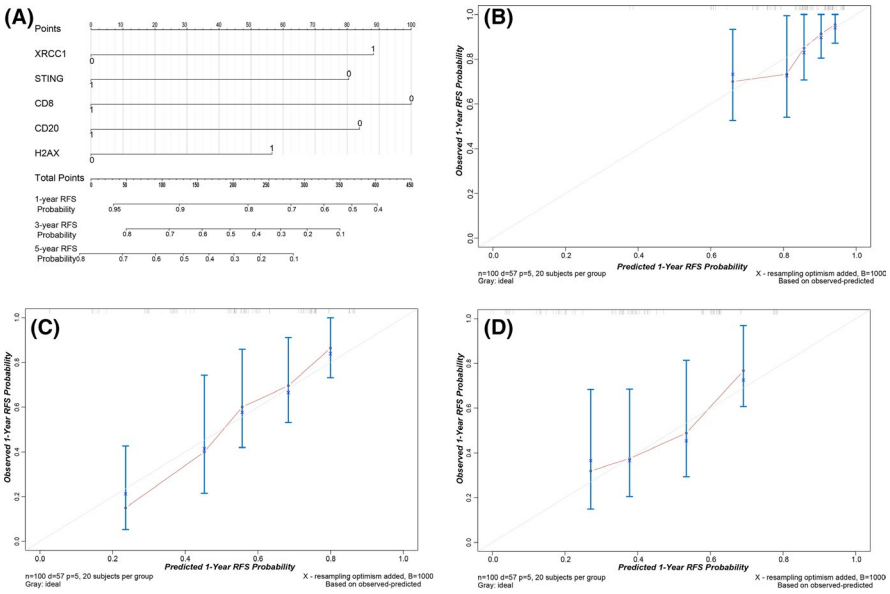

FIGURE 3 3Nomogram prediction model for estimating the 1-y, 3-y and 5-y relapse-free survival (RFS). A, Internal cross–validated adaptive elastic-net model for RFS was illustrated by a nomogram. A straight line is drawn pointing to the "point" axis for each patient's clinical features and expression markers, then follow a line pointing to the "total point" axis, the "linear predictor" and the corresponding "1-, 3-, 5-" straight-line annual RFS probability. B, Calibration plots show predictions for 1-y RFS. C, Calibration plots show predictions for 3-y RFS. D, Calibration plots show predictions for 5-y RFS. C-index = 0.724

and multivariate COX analysis revealed an association between risk score and disease recurrence as well as demonstrating the role of risk score as an independent prognostic factor (Table 3).

3.5 | Correlation analysis among base excision repair, the stimulator of interferon genes pathway and tumor-infiltrating lymphocytes

To investigate the biological relationship between the BER, the STING pathway and TIL, we generated a correlogram on TCGA data (Figure 5). We found that APE1 and H2AX was negatively correlated with STING ($r = -0.29$, $r = -0.37$), CD4 ($r = -0.3$, $r = -0.18$) and CD20 ($r = -0.2$, $r = -0.16$); STING was positively correlated with CD3 ($r = 0.16$), CD4 ($r = 0.4$) and CD20 ($r = 0.24$); XRCC1 was positively correlated with IRF3 ($r = 0.4$).

4 | DISCUSSION

The 5-year RFS after surgical resection of early-stage lung cancer is only 70%,[4,5] which, therefore, requires the identification of useful prognosis predictors. In this study, the protein expression levels of BER, the STING pathway and TIL were studied to predict the recurrence of early-stage lung cancer, and the prediction model was verified by external mRNA data. The results revealed that the BER and STING pathway played a crucial role in the recurrence of early-stage lung cancer, and there was an important relationship between BER, the STING pathway and TIL.

DNA damage, a hallmark of cancer, is the cause of genetic mutation.[34-36] BER dysfunction will increase the damage to nucleic acid and DNA double-strand breaks, which participate in tumor immunity through innate immunity. Johannes et al found that DNA damage was negatively associated with immunogenic cell death and T cell infiltrates, and, thus, might be a therapeutic target for metastatic colorectal liver cancer.[37] Morgan *et al.* found that DNA damage was able to trigger the innate immune response and improved the efficacy of immune checkpoint blockade (ICB) in pancreatic cancer.[38] Some studies also found that BER was associated with tumor prognosis and treatment response. High expression of APE1 in tumors was associated with the resistance to platinum and radiotherapy.[39,40] XRCC1 downregulated in gastric cancer, lung cancer and cervical cancer was associated with tumor invasion and poor prognosis.[41-43] Furthermore, polβ mutations were found to be associated with tumor susceptibility.[44,45]

Activation of the host cGAS-STING pathway contributes to antitumor immunity. Activation of STING in tumor cells and dendritic cells can promote the cross–presentation of tumor antigens, thereby activating CD8 + T cells to control the tumor.[46] Although no direct association was found between STING and recurrence of early-stage lung cancer, TBK1, the downstream molecule of STING, was found to be associated with recurrence of early-stage lung cancer in our study. Furthermore, the expression of STING

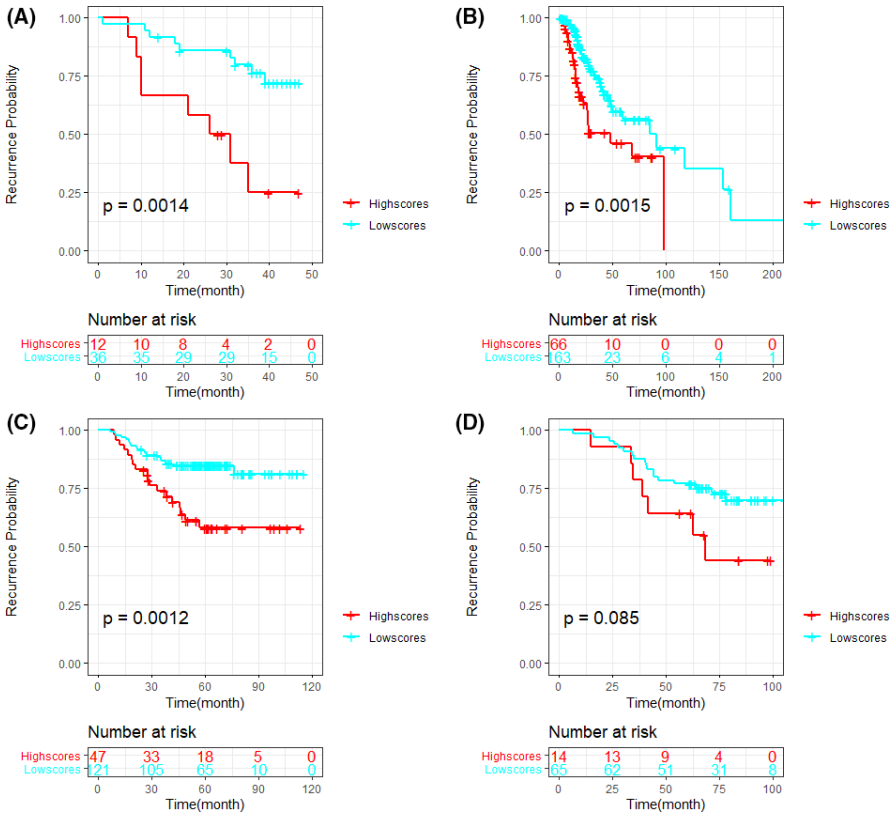

FIGURE 4 4Kaplan-Meier survival analysis of relapse-free survival (RFS) for high-risk and low-risk groups based on the nomogram of Daping immunohistochemistry (IHC) data, The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO) database. A, internal validation cohort data. B, TCGA dataset. C, GSE31210 dataset. D, GSE13213 dataset

TABLE 3 Univariate analysis of the prediction model for RFS in patients in TCGA and GEO database

	TCGA			GSE31210			GSE13213		
	HR	95% CI	*P*-value	HR	95% CI	*P*-value	HR	95% CI	*P*-value
Prediction model	2.653	1.471-4.388	0.012	2.667	1.439-5.447	0.025	2.281	0.984-5.849	0.043
Sex	0.942	0.575-1.545	0.814	1.110	0.579-2.128	0.753	2.028	0.895-4.594	0.090
Age	1.144	0.640-2.046	0.650	1.706	0.857-3.398	0.128	1.019	0.440-2.362	0.965
Smoking	1.231	0.745-1.984	0.564	1.167	0.611-2.230	0.639	1.478	0.670-3.259	0.333
Stage	1.157	0.713-1.878	0.555	/	/	/	1.576	0.715-3.476	0.259
Mutation	/	/	/	0.441	0.230-0.846	0.014	1.063	0.485-2.331	0.879

Note: Sex: male vs female.

Age: >60 vs <60.

Smoking: Yes vs no.

Stage: IB vs IA.

Gene signature: High risk vs low risk.

Positive vs negative.

Mutation positive: Mutations in any of the EGFR, KRAS or ALK fusion were defined as mutation positive; mutation negative: mutations not in any of the EGFR, KRAS and no ALK fusion driver genes were defined as mutation negative.

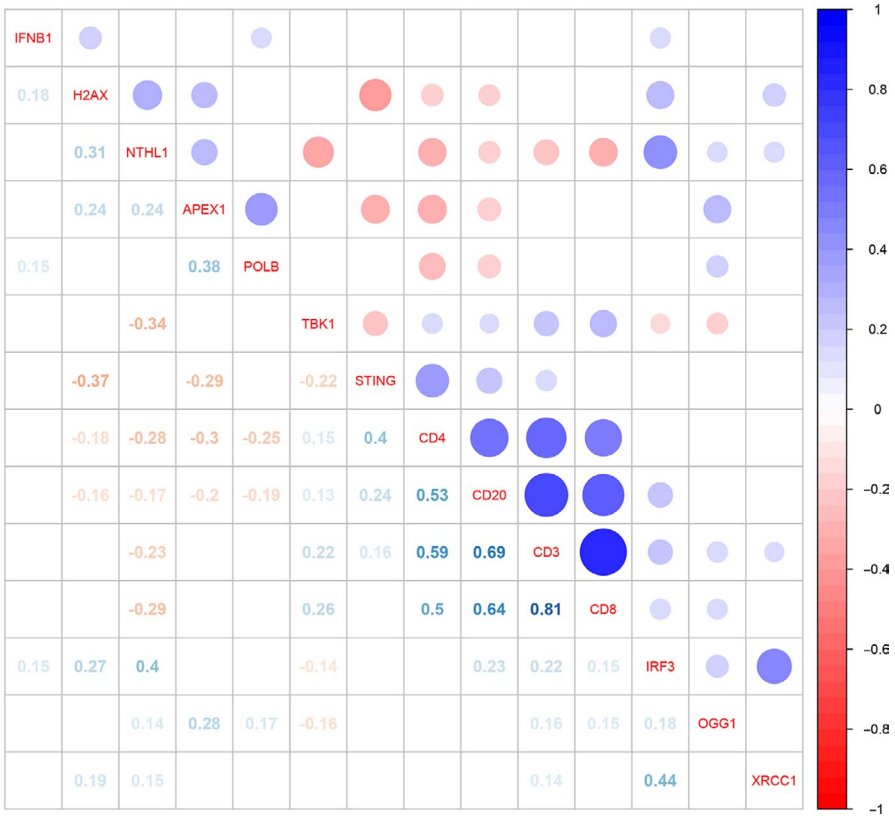

FIGURE 5 A correlogram of base excision repair, stimulator of interferon genes and tumor-infiltrating lymphocyte panel values. Pearson's correlation coefficients (*r*) for all markers are given in the lower triangle. Colored circles indicate statistically significant correlations. The size and color intensity of the circles are related to the correlation coefficients. Blue color indicates a positive correlation and red color indicates a negative correlation

plays a vital role in our recurrence risk prediction model. We initially assumed that good prognosis was related to the high expression of STING pathway-related proteins, but our results showed that the low expression of TBK1 played a role in good prognosis. The reason may be that TBK1 does not represent the activation state of the STING pathway, whose activation, as we know, requires the phosphorylation activation of STING/TBK1/IRF3, because the total protein expression level does not necessarily reflect the phosphorylation level.

Tumor-infiltrating lymphocytes are associated with tumor prognosis,[47,48] and a large number of tumor-infiltrating TIL cells are associated with good prognosis of lung cancer.[10] Thus, based on the

impact of TIL on prognosis, some researchers have proposed that TNM-I staging can be used to better predict the prognosis.[49] In our study, TIL were labeled with CD3, CD4, CD8 and CD20, respectively. Results showed that good prognosis was associated with high expressions of CD3, CD4 and CD8 in tumor tissues, but there was no association between the expression of CD20 in TIL and prognosis.

In this study, we explored the impact of BER-STING-TIL on the recurrence of early-stage lung cancer and established a prediction nomogram model. We found that the expression of XRCC1 and H2AX are negative prognostic markers for RFS, while CD8, CD20 and STING are positive prognostic markers for RFS. Nomograms for RFS share common prognostic markers, including XRCC1, H2AX, STING, CD8 and CD20. The c-index was 0.724 and 0.698 in the training cohort and the internal validation cohort, respectively. The model had a good prediction effect after the external data were verified. At the same time, we analyzed the correlations among BER, STING and TIL, and the results showed that the BER and STING pathways were correlated with TIL. There were some limitations in this study: (a) a prediction nomogram was established by protein expression, but validation of this nomogram was performed using external mRNA expression data; and (b) the correlations among BER, STING and TIL were calculated by statistical analysis but had not been confirmed by basic experimental studies.

In conclusion, we found that BER, the STING pathway and TIL were associated with the recurrence of early-stage lung cancer. APE1 and H2AX were negatively correlated with STING, CD4 and CD20. Our nomogram model of recurrence prediction established based on BER, STING and TIL expression showed a good prediction function for the recurrence of stage I LUAD. The results of this study provided further clinical evidence for recruiting and activating TIL through the STING pathway.

ACKNOWLEDGMENTS

The authors thank the Department of Pathology and Director Hualiang Xiao for their help in specimen preparation and IHC scoring. The results shown here are in part based upon data generated by TCGA Research Network and GEO database. This study was supported by grant NSFC 81772495 from the National Natural Science Foundation of China.

DISCLOSURE

The authors declare that they have no competing interests.

ORCID

Dong Wang https://orcid.org/0000-0001-6585-2704

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, 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. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69:7-34.
3. NCCN guidelines version 1.2019 non–small cell lung cancer. National Comprehensive Cancer Network, Inc. 2019.
4. Guerrera F, Errico L, Evangelista A, et al. Exploring Stage I non–small-cell lung cancer: development of a prognostic model predicting 5-year survival after surgical resectiondagger. *Eur J Cardiothorac Surg*. 2015;47:1037-1043.
5. Kelsey CR, Marks LB, Hollis D, et al. Local recurrence after surgery for early stage lung cancer: an 11-year experience with 975 patients. *Cancer*. 2009;115:5218-5227.
6. Gold KA, Kim ES, Liu DD, et al. Prediction of survival in resected non–small cell lung cancer using a protein expression-based risk model: implications for personalized chemoprevention and therapy. *Clin Cancer Res*. 2014;20:1946-1954.
7. Wistuba II, Behrens C, Lombardi F, et al. Validation of a proliferation-based expression signature as prognostic marker in early stage lung adenocarcinoma. *Clin Cancer Res*. 2013;19:6261-6271.
8. Brock MV, Hooker CM, Ota-Machida E, et al. DNA methylation markers and early recurrence in stage I lung cancer. *N Engl J Med*. 2008;358:1118-1128.
9. Lu Y, Govindan R, Wang L, et al. MicroRNA profiling and prediction of recurrence/relapse-free survival in stage I lung cancer. *Carcinogenesis*. 2012;33:1046-1054.
10. Bremnes RM, Busund LT, Kilvaer TL, et al. The role of tumor-infiltrating lymphocytes in development, progression, and prognosis of non–small cell lung cancer. *J Thorac Oncol*. 2016;11:789-800.
11. Yang X, Shi Y, Li M, et al. Identification and validation of an immune cell infiltrating score predicting survival in patients with lung adenocarcinoma. *J Transl Med*. 2019;17:217.
12. Musella M, Manic G, De Maria R, Vitale I, Sistigu A. Type-I-interferons in infection and cancer: Unanticipated dynamics with therapeutic implications. *Oncoimmunology*. 2017;6:e1314424.
13. Wang YJ, Fletcher R, Yu J, Zhang L. Immunogenic effects of chemotherapy-induced tumor cell death. *Genes Dis*. 2018;5:194-203.
14. Sistigu A, Yamazaki T, Vacchelli E, et al. Cancer cell-autonomous contribution of type I interferon signaling to the efficacy of chemotherapy. *Nat Med*. 2014;20:1301-1309.
15. Pepin G, Gantier MP. cGAS-STING activation in the tumor microenvironment and its role in cancer immunity. *Adv Exp Med Biol*. 2017;1024:175-194.
16. Chen K, Liu J, Cao X. Regulation of type I interferon signaling in immunity and inflammation: a comprehensive review. *J Autoimmun*. 2017;83:1-11.
17. Ishikawa H, Ma Z, Barber GN. STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature*. 2009;461:788-792.
18. Corrales L, McWhirter SM, Dubensky TW Jr, Gajewski TF. The host STING pathway at the interface of cancer and immunity. *J Clin Investig*. 2016;126:2404-2411.
19. Deng L, Liang H, Xu M, et al. STING-dependent cytosolic DNA sensing promotes radiation-induced Type I interferon-dependent antitumor immunity in immunogenic tumors. *Immunity*. 2014;41:843-852.
20. Zhang C, Shang G, Gui X, Zhang X, Bai XC, Chen ZJ. Structural basis of STING binding with and phosphorylation by TBK1. *Nature*. 2019;567:394-398.
21. Shang G, Zhang C, Chen ZJ, Bai XC, Zhang X. Cryo-EM structures of STING reveal its mechanism of activation by cyclic GMP-AMP. *Nature*. 2019;567:389-393.
22. Ablasser A, Goldeck M, Cavlar T, et al. cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. *Nature*. 2013;498:380-384.
23. Chatzinikolaou G, Karakasilioti I, Garinis GA. DNA damage and innate immunity: links and trade-offs. *Trends Immunol*. 2014;35:429-435.

24. Mouw KW, Goldberg MS, Konstantinopoulos PA, D'Andrea AD. DNA damage and repair biomarkers of immunotherapy response. *Cancer Discov*. 2017;7:675-693.
25. Hartlova A, Erttmann SF, Raffi FA, et al. DNA damage primes the type I interferon system via the cytosolic DNA sensor STING to promote anti–microbial innate immunity. *Immunity*. 2015;42:332-343.
26. Samstein RM, Riaz N. The DNA damage response in immunotherapy and radiation. *Adv Radiat oncol*. 2018;3:527-533.
27. Xu Y. DNA damage: a trigger of innate immunity but a requirement for adaptive immune homeostasis. *Nat Rev Immunol*. 2006;6:261-270.
28. Krokan HE, Nilsen H, Skorpen F, Otterlei M, Slupphaug G. Base excision repair of DNA in mammalian cells. *FEBS Lett*. 2000;476:73-77.
29. Wallace SS. Base excision repair: a critical player in many games. *DNA Repair*. 2014;19:14-26.
30. Poletto M, Legrand AJ, Dianov GL. DNA base excision repair: the Achilles' heel of tumour cells and their microenvironment? *Curr Pharm Des*. 2017;23:4758-4772.
31. Bose D. cGAS/STING pathway in cancer: Jekyll and hyde story of cancer immune response. *Int J Mol Sci*. 2017;18.
32. Kallioniemi OP, Wagner U, Kononen J, Sauter G. Tissue microarray technology for high-throughput molecular profiling of cancer. *Hum Mol Genet*. 2001;10:657-662.
33. Wang LA, Yang B, Rao W, Xiao H, Wang D, Jiang J. The correlation of BER protein, IRF3 with CD8+ T cell and their prognostic significance in upper tract urothelial carcinoma. *OncoTargets and therapy*. 2019;12:7725-7735.
34. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144:646-674.
35. Chatterjee N, Walker GC. Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen*. 2017;58:235-263.
36. Tubbs A, Nussenzweig A. Endogenous DNA damage as a source of genomic instability in cancer. *Cell*. 2017;168:644-656.
37. Laengle J, Stift J, Bilecz A, et al. DNA damage predicts prognosis and treatment response in colorectal liver metastases superior to immunogenic cell death and T cells. *Theranostics*. 2018;8:3198-3213.
38. Zhang Q, Green MD, Lang X, et al. Inhibition of ATM increases interferon signaling and sensitizes pancreatic cancer to immune checkpoint blockade therapy. *Can Res*. 2019;79:3940-3951.
39. Wang D, Xiang DB, Yang XQ, et al. APE1 overexpression is associated with cisplatin resistance in non–small cell lung cancer and targeted inhibition of APE1 enhances the activity of cisplatin in A549 cells. *Lung Cancer*. 2009;66:298-304.
40. Yuan CL, He F, Ye JZ, et al. APE1 overexpression is associated with poor survival in patients with solid tumors: a meta-analysis. *Oncotarget*. 2017;8:59720-59728.
41. Lee KJ, Piett CG, Andrews JF, Mann E, Nagel ZD, Gassman NR. Defective base excision repair in the response to DNA damaging agents in triple negative breast cancer. *PLoS One*. 2019;14:e0223725.
42. Yousaf S, Khan AU, Akram Z, et al. Expression deregulation of DNA repair pathway genes in gastric cancer. *Cancer Genet*. 2019;237:39-50.
43. Albertella MR, Lau A, O'Connor MJ. The overexpression of specialized DNA polymerases in cancer. *DNA Repair*. 2005;4:583-593.
44. Zhao S, Klattenhoff AW, Thakur M, Sebastian M, Kidane D. Mutation in DNA polymerase beta causes spontaneous chromosomal instability and inflammation-associated carcinogenesis in mice. *Cancers*. 2019;11:1160.
45. Guo Z, Zheng L, Dai H, Zhou M, Xu H, Shen B. Human DNA polymerase beta polymorphism, Arg137Gln, impairs its polymerase activity and interaction with PCNA and the cellular base excision repair capacity. *Nucleic Acids Res*. 2009;37:3431-3441.
46. Woo SR, Fuertes MB, Corrales L, et al. STING-dependent cytosolic DNA sensing mediates innate immune recognition of immunogenic tumors. *Immunity*. 2014;41:830-842.
47. Tsoutsou PG, Bourhis J, Coukos G. Tumor-infiltrating lymphocytes in triple-negative breast cancer: a biomarker for use beyond prognosis? *J Clin Oncol*. 2015;33:1297-1298.
48. Dirican A, Ekinci N, Avci A, et al. The effects of hematological parameters and tumor-infiltrating lymphocytes on prognosis in patients with gastric cancer. *Cancer Biomark*. 2013;13:11-20.
49. Galon J, Mlecnik B, Bindea G, et al. Towards the introduction of the 'Immunoscore' in the classification of malignant tumours. *J Pathol*. 2014;232:199-209.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Yang B, Rao W, Luo H, Zhang L, Wang D. Relapse-related molecular signature in early-stage lung adenocarcinomas based on base excision repair, stimulator of interferon genes pathway and tumor-infiltrating lymphocytes. *Cancer Sci*. 2020;111:3493–3502. https://doi.org/10.1111/cas.14570