A signature of tumor DNA repair genes associated with the prognosis of surgically-resected lung adenocarcinoma

Xiongtao Yang [1] , **Guohui Wang** [2] , **Runchuan Gu** [3] , **Xiaohong Xu** [1] , **Guangying Zhu** [Corresp. 4]

[1] Department of Radiation Oncology, Peking University China-Japan Friendship School of Clinical Medicine, Beijing, China

[2] Department of Radiotherapy, Second hospital of Hebei Medical University, Shijiazhuang, Hebei, China

[3] Department of Radiation Oncology, China-Japan Friendship Institute of Clinical Medicine, Beijing, China

[4] Department of Radiation Oncology, China-Japan Friendship Hospital, Beijing, China

Corresponding Author: Guangying Zhu
Email address: zryyfa@163.com

Background. Lung cancer has the highest morbidity and mortality of cancers worldwide. Lung adenocarcinoma (LUAD) is the most common pathological subtype of lung cancer and surgery is its most common treatment. The dysregulated expression of DNA repair genes is found in a variety of cancers and has been shown to affect the origin and progression of these diseases. However, the function of DNA repair genes in surgically-treated LUAD is unclear.

Methods. We sought to determine the association between the signature of DNA repair genes for patients with surgical LUAD and their overall prognosis. We obtained gene expression data and corresponding clinical information of LUAD from The Cancer Genome Atlas (TCGA) database. The differently expressed DNA repair genes of surgically-treated LUAD and normal tissues were identified using the Wilcoxon rank-sum test. We used uni- and multivariate Cox regression analyses to shrink the aberrantly expressed genes, which were then used to construct the prognostic signature and the risk score formula associated with the independent prognosis of surgically-treated LUAD. We used Kaplan–Meier and Cox hazard ratio analyses to confirm the diagnostic and prognostic roles. Two validation sets (GSE31210 and GSE37745) were downloaded from the Gene Expression Omnibus (GEO) and were used to externally verify the prognostic value of the signature. OSluca online database verifies the hazard ratio for the DNA repair genes by which the signature was constructed. We investigated the correlation between the signature of the DNA repair genes and the clinical parameters. The potential molecular mechanisms and pathways of the prognostic signature were explored using Gene Set Enrichment Analysis (GSEA).

Results. We determined the prognostic signature based on 6 DNA repair genes (PLK1, FOXM1, PTTG1, CCNO, HIST3H2A, and BLM) and calculated the risk score based on this formula. Patients with surgically-treated LUAD were divided into high-risk and low-risk groups according to the median risk score. The high-risk group showed poorer overall survival than the low-risk group; the signature was used as an independent prognostic indicator and had a greater prognostic value in surgically-treated LUAD. The prognostic value was replicated in GSE31210 and GSE37745. OSluca online database analysis shows that 6 DNA repair genes were associated with poor prognosis in most lung cancer datasets. The prognostic signature risk score correlated with the pathological stage and smoking status in surgically-treated LUAD. The GSEA of the risk signature in high-risk patients showed pathways associated with the cell cycle, oocyte meiosis, mismatch repair, homologous recombination, and nucleotide excision repair.

Conclusions. A 6-DNA repair gene signature was determined using TCGA data mining and GEO data

verification. The gene signature may serve as a novel prognostic biomarker and therapeutic target for surgically-treated LUAD.

A signature of tumor DNA repair genes associated with the prognosis of surgically-resected lung adenocarcinoma

Xiongtao Yang[1], Guohui Wang[2], Runchuan Gu[3], Xiaohong Xu[1], Guangying Zhu[4]

[1]Department of Radiation Oncology, Peking University China-Japan Friendship School of Clinical Medicine, Beijing, China

[2]Department of Radiotherapy, Second Hospital of Hebei Medical University, Shijiazhuang, Hebei, China

[3] Department of Radiation Oncology, China-Japan Friendship Institute of Clinical Medicine, Beijing, China

[4] Department of Radiation Oncology, China-Japan Friendship Hospital, Beijing, China

Corresponding Author: Guangying Zhu

Yinghua East Street, Chaoyang District, Beijing, 100000, China

Email address: zryyfa@163.com

Abstract

Background. Lung cancer has the highest morbidity and mortality of cancers worldwide. Lung adenocarcinoma (LUAD) is the most common pathological subtype of lung cancer and surgery is its most common treatment. The dysregulated expression of DNA repair genes is found in a variety of cancers and has been shown to affect the origin and progression of these diseases. However, the function of DNA repair genes in surgically-treated LUAD is unclear.

Methods. We sought to determine the association between the signature of DNA repair genes for patients with surgical LUAD and their overall prognosis. We obtained gene expression data and corresponding clinical information of LUAD from The Cancer Genome Atlas (TCGA) database. The differently expressed DNA repair genes of surgically-treated LUAD and normal tissues were identified using the Wilcoxon rank-sum test. We used uni- and multivariate Cox regression analyses to shrink the aberrantly expressed genes, which were then used to construct the prognostic signature and the risk score formula associated with the independent prognosis of surgically-treated LUAD. We used Kaplan–Meier and Cox hazard ratio analyses to confirm the diagnostic and prognostic roles. Two validation sets (GSE31210 and GSE37745) were downloaded from the Gene Expression Omnibus (GEO) and were used to externally verify the prognostic value of the signature. OSluca online database verifies the hazard ratio for the DNA repair genes by which the signature was constructed. We investigated the correlation between the signature of the DNA repair genes and the clinical parameters. The potential molecular mechanisms and pathways of the prognostic signature were explored using Gene Set Enrichment Analysis (GSEA).

Results. We determined the prognostic signature based on 6 DNA repair genes (PLK1, FOXM1, PTTG1, CCNO, HIST3H2A, and BLM) and calculated the risk score based on this formula. Patients with surgically-treated LUAD were divided into high-risk and low-risk groups according to the median risk score. The high-risk group showed poorer overall survival than the

40 low-risk group; the signature was used as an independent prognostic indicator and had a greater
41 prognostic value in surgically-treated LUAD. The prognostic value was replicated in GSE31210
42 and GSE37745. OSluca online database analysis shows that 6 DNA repair genes were associated
43 with poor prognosis in most lung cancer datasets. The prognostic signature risk score correlated
44 with the pathological stage and smoking status in surgically-treated LUAD. The GSEA of the
45 risk signature in high-risk patients showed pathways associated with the cell cycle, oocyte
46 meiosis, mismatch repair, homologous recombination, and nucleotide excision repair.
47 **Conclusions.** A 6-DNA repair gene signature was determined using TCGA data mining and
48 GEO data verification. The gene signature may serve as a novel prognostic biomarker and
49 therapeutic target for surgically-treated LUAD.

50 ## Introduction

51 Lung cancer has the highest morbidity and mortality of cancers worldwide (Siegel et al. 2020).
52 Lung adenocarcinoma (LUAD) is the most common pathological subtype of lung cancer,
53 accounting for approximately 40% of all cases (Denisenko et al. 2018). Surgery is the most
54 common treatment for LUAD (Hirsch et al. 2017) although there is a high rate of relapse with
55 30- 40% of patients dying from stage I surgically-treated LUAD (Oskarsdottir et al. 2016;
56 Westaway et al. 2013). The 5-year overall survival (OS) rate decreases to 36% for those with
57 stage III--IV surgically-treated LUAD (Mansuet-Lupo et al. 2014). The patients with surgically
58 resected LUAD had varied prognoses, but patients with partially surgically-resected cancers had
59 poor prognoses, regardless of whether the patients were early-LUAD treated with surgery alone
60 or advanced-LUAD undergoing aggressive postoperative therapy (including platinum-based
61 chemotherapy and radiation therapy). The biological mechanism of these differences is not fully
62 understood (De Miguel-Perez et al. 2019). It is important to determine the prognoses of
63 surgically-treated LUAD patients to provide specific treatments for improved OS.
64 Previous studies have shown that DNA repair systems play an important role in cancer(Gavande
65 et al. 2016; Jeggo et al. 2016; Mei et al. 2019; Tessoulin et al. 2018). Five major DNA repair
66 pathways are vital to maintaining the genetic stability of cells (Chatterjee & Walker 2017): base
67 excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), homologous
68 recombination (HR) and non-homologous end joining (NHEJ), two specific lesion repair
69 pathways (direct chemical reversal and interstrand crosslink (ICL) repair) are important
70 supplements for DNA damage repair. DNA repair genes play an essential role in tumorigenesis,
71 tumor progression, and response to therapy(Lima et al. 2019). Efforts are being made to identify
72 DNA damage repair genes as therapeutic targets and to develop agents against these targets. For
73 example, poly ADP-ribose polymerase 1 (PARP1) is involved in multiple DNA repair pathways
74 to maintain genomic stability so PARP1 inhibitors have been approved to treat ovarian cancer,
75 breast cancer, and other DNA repair-deficient tumors (Ray Chaudhuri & Nussenzweig 2017).
76 LUAD exists in multiple DNA repair gene mutations but no studies have been conducted on
77 DNA damage repair pathways in LUAD to date. We studied surgically-resected LUAD and the
78 prognostic and predictive significance of the expression of DNA repair genes for this cancer.

79   We constructed a multi-gene signature suitable for clinical application in surgical LUAD
80   patients, based on multi-gene signatures that indicate a poor prognosis in high-risk populations of
81   specific tumor patients. This approach increases our knowledge of tumor cell generation, growth,
82   and metastasis, and creates new avenues for targeted therapy. We obtained 201 surgically-treated
83   LUAD cases from The Cancer Genome Atlas (TCGA) database and extracted and identified key
84   mRNAs associated with DNA repair and established a 6-gene risk signature that could accurately
85   predict the prognosis of these specific patients.

Materials & Methods

Data source and preprocessing

88   We extracted the gene expression profiles and corresponding clinical information of patients
89   with LUAD from the TCGA database (https://portal.gdc.cancer.gov/). We obtained the gene
90   expression profiles of 59 normal samples and screened 535 LUAD samples for performance
91   status, OS, survival state, and gene expression profiles. We collected data from 201 surgically-
92   resected LUAD cases. The gene mRNA expression data were normalized using the Transcripts
93   Per Kilobase of exon model per Million mapped reads (TPM) method (Li et al. 2010). Clinical
94   information including gender, age, pathological stage and TNM stage were included in this
95   research. The Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/)
96   for microarray-based expression data was used to obtain data from patients with surgically-
97   resected LUAD, and that datasets to be considered must be containing more than 100 cases with
98   information of OS and living status. The expression data of the GSE31210 set were normalized
99   by the MAS5 algorithm and transformed with log2. The expression data of the GSE37745 set
100   were normalized by the Robust Multi-Array Average (RMA) method and transformed with log2.
101   226 cases from GSE31210 (validation set 1) and 106 cases from GSE37745 (validation set 2)
102   were screened with gene expression data and relevant clinical information. The OSluca online
103   database (http://bioinfo.henu.edu.cn/LUCA/LUCAList.jsp) contains multiple lung cancer
104   datasets(Yan et al. 2020), which can perform hazard ratio analysis on screened prognostic genes
105   related to DNA repair. TCGA data were applied as a training set. GEO data sets were applied as
106   validation sets. OS was defined as the date of the first treatment to the date of death of any cause
107   or to the time of the last follow-up. The clinical characteristics are summarized in Table 1.

Bioinformatic analysis

109   Differential expression genes (DEGs) were filtered according to the adjusted p-value <0.05,
110 |log2fold change| >1.5 in the surgically-treated LUAD and normal samples of the training set.
111   DNA repair genes were further screened from DEGs. Univariate Cox regression analysis
112   evaluated the DNA repair genes significantly related to OS according to p-value < 0.05.
113   Multivariate Cox regression analysis was used for further screening. DNA repair genes are
114   executors in the extremely complicated processes of DNA damage repairing which remedies
115   specific types of DNA damage through specific pathways. We obtained a prognostic signature of
116   6 DNA repair genes and a corresponding prognostic risk score formula based on a linear
117   combination of corresponding multivariate Cox regression coefficients. The formula for the risk
118   score was: expression level of gene 1 ×K1 + expression level of gene 2 ×K2 +···+ expression of
119   level gene n ×Kn. The risk score for every surgically-treated LUAD patient was calculated using
120   this formula and the population was categorized into high-risk and low-risk groups according to
121   the median risk score value. We used gene set enrichment analysis (GSEA) (https://www.gsea-

122   msigdb.org/gsea/msigdb) to identify pathways associated with the high-risk and the low-risk
123   groups of DNA repair gene signatures. The Kyoto Encyclopedia of Genes and Genomes (KEGG)
124   gene sets were downloaded from the MSigDB database. Enrichment false discovery rates (FDR)
125   were based on 1,000 permutations. The high-risk score vs. low-risk score and FDR < 0.07 were
126   considered to be statistically significant.

127   **Statistical analysis**
128   The gene expression differences between the surgically-treated LUAD group and the normal
129   group were compared using the unpaired two-tailed student's t-test. We evaluated the prognostic
130   value of the risk score and clinicopathological features in patients with surgically-treated LUAD
131   using univariate and multivariate Cox analysis. The Kaplan-Meier method was used to analyze
132   the correlation between the risk score and OS. The survival curves were compared using a log-
133   rank test. All statistical analyses were performed using R 3.6.3 (https://www.r-project.org/) and
134   SPSS 20.0.

135   # Results

136   **Identification of DNA repair-related prognostic genes**
137   Study was conducted in accordance of the flow chart (Fig. 1). LUAD gene expression data and
138   clinical records were downloaded from TCGA database. We screened DEGs from surgically-
139   treated LUAD and normal samples of the training dataset. In accordance with common senses
140   and definition, we collected DNA repair-related gene sets from the MSigDB database and
141   literature records (Table S1)(Pearl et al. 2015), and further selected from the DEGs to identify
142   the DNA repair-related prognostic gene signature for surgically-treated LUAD. We acquired 68
143   genes related to DNA repair (Fig. 2A). Among them, the expression levels of 4 genes in surgical
144   LUAD were lower than normal samples, while the other genes were the opposite. (Fig.2B,
145   Fig.2C). These genes were used in Univariate Cox regression analysis and resulted in 23 genes
146   significantly correlated with OS (Fig. 2D). The 23 genes were screened by multivariate Cox
147   regression analysis, resulting in a total of 6 genes, namely PLK1, FOXM1, PTTG1, CCNO,
148   HIST3H2A and BLM. A prognostic signature was constructed (Table 2) based on the 6 DNA
149   repair-related prognostic genes.

150   **The performance of DNA repair-related prognostic gene signature**
151   The risk score for surgically-treated LUAD patients was calculated based on the gene mRNA
152   level and corresponding coefficients using the formula: Risk score = (0.555713) x PLK1
153   expression value + (-0.5696) × FOXM1 expression value + (0.291471) × PTTG1 expression
154   value + (-0.30485) × CCNO expression value + (-0.17162) × HIST3H2A expression value +
155   (0.442472) × BLM expression value. The patients were divided into high-risk and low-risk
156   groups based on the median risk score. The patients' risk score distribution and OS are shown in
157   Fig. 3A and 3B. The expression profile distribution of 6 DNA repair-related genes is shown in
158   Fig. 3C. A Kaplan-Meier curve showed poor prognoses for the high-risk group (log-rank test p-
159   value <0.0001; Fig. 3D). We used univariate and multivariate Cox hazard ratio analysis to
160   further assess the performance of our signature risk score in surgically-treated LUAD patients
161   based on other common clinicopathological features. Univariate Cox regression analysis showed
162   that tumor stage and N classification were significantly correlated with survival. N classification

163   was significantly correlated with survival in multivariate Cox regression analysis. Remarkably,
164   univariate and multivariate Cox analysis showed that the risk score was more significantly
165   correlated with survival, giving it a greater prognostic value (Figs. 3E and 3F). These results
166   indicated that the 6-gene signature risk score could reliably predict the survival of surgical
167   LUAD patients.

168   Validation of the 6-gene signature for survival prediction

169   Gene expression data and clinical information of LUAD patients with surgical records and
170   survival times were selected from the GEO database to further prove the accuracy of our gene
171   signature. We obtained the GSE31210 and GSE37745 data sets. The GSE31210 expression data
172   were standardization and distributed as shown in Fig. 4A and 4B. The GSE37745 set contained
173   data from pathological samples of squamous cell lung carcinoma, lung adenocarcinoma, and
174   large-cell carcinoma. We screened surgically-treated LUAD samples; their standardization and
175   distribution is shown in Fig. 4C and 4D. The DNA repair-related signature and risk score
176   formulas were applied to verify the two validation sets and the Kaplan-Meier curve showed that
177   the high-risk group had poor OS when compared with the low-risk group (Fig. 4E p-
178   value=0.008, Fig. 4F p-value=0.003). Apart from that, multivariate Cox hazard ratio regression
179   analysis in the validation sets GSE31210 and GSE37745 had also been conducted. According to
180   the results, risk score can be considered as an independent risk factor (Fig. 4G HR=1.523, p-
181   value=0.023) in the GSE37745 set, but since only clinical stage I and II patients are included in
182   the GSE31210 set differing from the training set TCGA which contains clinical I, II, III, IV
183   stages, it could be insufficient regarding risk score as an independent risk factor after the
184   inclusion of stage staging (Fig. 4H HR=1.083, p-value =0.072). The 6 DNA repair-related genes
185   of the signature were analyzed in the OSluca online database for forest plot prognostic analysis.
186   The results showed that these 6 genes were associated with poor prognosis in most lung cancer
187   data sets (Fig.5, HR>1).The above results indicated that the signature had greater external
188   validity and reliability.

189   Relationship between 6-gene signature risk score and clinical parameters

190   We studied the correlation between the 6-gene signature risk score and the clinical parameters
191   under the training and validation sets. The sets were divided into two subgroups that included
192   age, gender, pathological stage, TNM stage, and smoking status. The stratification results are
193   shown in Table 1. In the training set, parameters such as age, gender, TNM stage did not affect
194   risk score except for the correlation between pathological stage and risk score (Fig. 6A) (p-value
195   =0.022). In the GSE31210 set, the pathological stage and smoking status were correlated with
196   the risk score, but age and gender were not related to the risk score (Fig. 6B) (p-value=0.0004, p-
197   value=0.027, respectively). In the GSE37745 set, age, gender and pathological stage were not
198   related to the risk score (Fig. 6C).

199   The biological processes and pathways associated with the 6-gene signature

200   The population was divided into a high-risk group and a low-risk group based on the risk score
201   of DNA repair gene expression and the prognosis between the two groups was significantly
202   different. We examined the whole gene expression profile in TCGA data using GSEA analysis to

203 reveal the potential mechanism of the 6-gene signature, which indicated that the main enrichment
204 pathways were the cell cycle, oocyte meiosis, DNA replication, mismatch repair, homologous
205 recombination and nucleotide excision repair pathways. Our results suggested that the poor
206 prognosis of patients with surgically-resected LUAD was related to the DNA damage repair
207 pathway (Fig. 7).

208 Discussion

209 Lung cancer has the highest mortality rate of malignant tumors and adenocarcinoma is the most
210 common histological subtype (Network 2014). Video-assisted thoracoscopic surgical lobectomy
211 for early-stage NSCLC and robotic lobectomy for advanced NSCLC is the primary surgical
212 treatment for LUAD (Cerfolio et al. 2018; Chen et al. 2013; Veronesi et al. 2018).
213 Advancements in imaging techniques, diagnostic methods, surgical skills, and anesthesia
214 management have led to an increase in the number of LUAD patients who were evaluated for
215 and underwent surgery (Ucvet et al. 2020). Surgical treatment of patients with LUAD can extend
216 survival time but there may be limited benefits to the partial surgical treatment of LUAD patients
217 because of local recurrence and distant metastasis. Therefore, effective prognostic and
218 therapeutic strategies to improve the survival rate of surgical LUAD are being actively explored.
219 The development of high-throughput sequencing technology has improved the identification of
220 potential biomarkers for tumor progression and prognosis. For example, Xu et al. (2020)
221 confirmed that flamingo subfamily Cadherin EGF LAG seven-pass G-type receptor 2 (CELSR2)
222 is significantly overexpressed in hepatocellular carcinoma and may serve as a novel prognostic
223 biomarker. Feng et al. (2020) used the Cox proportional hazards model to confirm that MHC
224 class I chain-related B (MICB) expression was significantly associated with clinical parameters
225 in colorectal cancer (CRC) and high MICB expression was an independent protective factor for
226 OS. However, the predictive power of such biomarkers is limited in context-specific cancers
227 because it is influenced by many factors, including transcription modification, translation
228 modification, and temporal and spatial heterogeneity. Thus, a signature based on multi-gene
229 cooperation can improve the predictive ability of each gene, resulting in a more reliable and
230 robust predictive ability versus a single biomarker (Ashley 2015; Jiang et al. 2020; Wolfe et al.
231 2020; Zhu et al. 2016). DNA damage repair genes are needed to maintain cell homeostasis and
232 are extensively involved in cell replication, chromatin modification, DNA repair, checkpoint
233 signaling, reactive oxygen species, metabolism, senescence, and apoptosis cellular mechanisms
234 (Brinkman et al. 2020; Csiszar et al. 2019; Gonzalo & Coll-Bonfill 2019; Lee et al. 2018; Royce
235 et al. 2019; Ungvari et al. 2019). Genetic mutations can effectively induce tumorigenesis and
236 progression (Pearl et al. 2015). Targeted inhibition of the function of specific genes in DNA
237 damage repair response has potential for clinical applications and the development of targeted
238 drugs (Li et al. 2019; Nickoloff et al. 2017; O'Connor 2015; Tarantini et al. 2019). We used
239 DNA repair genes to construct and validate a prognostic signature that successfully predicts the
240 survival time of surgically-treated LUAD patients.
241 We filtered the data downloaded from the TCGA database to obtain gene expression profiles and
242 clinical data of patients with surgically-treated LUAD. The standardized DNA repair gene DEGs
243 were screened for OS-related univariate and multivariate Cox regression analysis and the DNA
244 repair-related prognosis signature and risk score calculation formulas were obtained. We divided
245 the surgically-treated LUAD patients into high- and low-risk groups according to the calculated
246 formula. Kaplan-Meier survival analysis showed that the high-risk group had a poor prognosis

247   and Cox analysis showed that the risk score could be used as an independent prognostic factor,
248   verifying the predictive effect of the signature. The GSE31210 and GSE377745 sets downloaded
249   from GEO were used to perform survival analysis and the results were consistent with those
250   from the Kaplan-Meier and Cox analyses. The DNA repair-related gene signature was shown to
251   be highly reliable. GSEA analysis suggested that the poor prognosis of patients with surgically-
252   resected LUAD was related to the DNA damage repair pathway, according to p-value <0.05 and
253   FDR <0.07.
254   We obtained 6 DNA repair genes making up the prognostic signature. Some of the 6 genes
255   (PLK1, FOXM1, PTTG1, CCNO, HIST3H2A and BLM) were associated with tumorigenesis.
256   PLK1 is a member of the Polo-like kinases (PLKs) family of serine/threonine protein kinases and
257   plays an essential role in many stages of mitosis. The expression of PLK1 is cell cycle-dependent
258   and its expression peaks in the M phase of a normal cell. PLK1 is highly expressed in many
259   human cancers and its overexpression is associated with a poor prognosis. Previous studies have
260   shown that the PLK1 protein interacts with multiple tumor suppressors to cause tumorigenesis
261   and that targeting the protein can improve the sensitivity of tumors to chemoradiotherapy. The
262   target PLK1 inhibitor has shown considerable promise in clinical studies and is being tested in
263   clinical trials (Gutteridge et al. 2016; Liu et al. 2017). FOXM1 is a proliferation specific
264   transcriptional modulator and is involved in the cell cycle, mitotic spindle integrity,
265   angiogenesis, metastasis, apoptosis, and DNA damage repair pathway. This protein dysfunction
266   contributes to tumorigenesis (Nandi et al. 2018). A meta-analysis reported that elevated FOXM1-
267   protein expression was significantly associated with poor survival in most solid tumors. This
268   suggests that FOXM1 is a potential biomarker for predicting prognoses in solid human tumors
269   (Li et al. 2017). PTTG1 has been identified as a critical signature gene associated with tumor
270   metastasis. It is overexpressed in a variety of endocrine-related tumors especially those of the
271   pituitary, breast, and testis, as well as in nonendocrine-related cancers involving the pulmonary
272   and gastrointestinal systems (Vlotides et al. 2007). Studies have shown that PTTG1 is highly
273   expressed in breast cancer and PTTG1 may increase breast cancer cell growth through the
274   nuclear exclusion of p27 (Xiea & Wangb 2016). BLM is a member of the RecQ helicases, and
275   can unwind forked dsDNA and anneal ssDNA, and play critical roles in DNA repair,
276   recombination, replication, and transcription. The mutations of BLM cause Bloom syndrome
277   (Croteau et al. 2014). The small molecule inhibitor of BLM, ML216, exhibits cell-based activity
278   and can induce sister chromatid exchanges, enhance the toxicity of aphidicolin, and exert
279   antiproliferative activity in cells expressing BLM (Nguyen et al. 2013). CCNO and HIST3H2A
280   were less prevalent in the high-risk group and there are few reports about the two genes. CCNO
281   is part of the cyclin family and is an essential regulator of endogenous apoptosis and participates
282   in DNA damage repair (Krokan & Bjoras 2013). HIST3H2A is a replication-dependent histone
283   H2A gene located in the HIST3 cluster on chromosome 1 in the human genome sequence. We
284   integrated the 6 DNA repair genes into a signature and found that patients in the high-risk group
285   were associated with worse survival, which confirms the predictive value of the DNA repair

286   gene signature in surgically-treated LUAD. The development of drugs against these target genes
287   may be used as a potential treatment in surgically-treated LUAD.
288   We filtered out patients with combined radiotherapy, chemotherapy, neoadjuvant therapy and
289   postoperative adjuvant therapy within the pre-set framework. In TCGA, due to the small number
290   of LUAD cases with surgical treatment combined with radiotherapy, neoadjuvant treatment or
291   adjuvant treatment, we did not conduct further analysis. 69 cases were found with chemotherapy
292   and the survival rate of low-risk group was higher, but no significant difference was obtained
293   (p=0.17) due to the small size of the sample, follow-up needs to expand the case for verification.
294   Tumor markers, as a class of substances synthesized and released by tumor cells, or increased as
295   the body reacting to tumor cells, are mainly applied for auxiliary diagnosis, prognostic judgment,
296   curative effect observation and guiding follow-up treatment in current clinical practice. Although
297   being convenient, highly-efficient, repeatably-detected and of low price, they are still limited by
298   its relatively poor specificity and sensitivity. As specified for patients with LUAD after surgery,
299   our multi-gene signature is supposed to have better specificity and the efficacy of radiotherapy
300   can be predicted based on the expression of DNA repair-related genes for those with
301   postoperative radiotherapy combined. Disadvantages mainly include the need for clinical sample
302   sequencing which could be difficult to obtain and high sequencing cost.
303   We use data mining and bioinformatics analysis on the TCGA, GEO and OSluca databases to
304   construct a DNA repair-related gene prognosis signature. Since the number of enrolled cases is
305   relatively small, more cases should be recruited to expand the study. In addition, we need more
306   clinical practice to further verify our conclusions.

307   Conclusions

308   We used data from surgically-treated LUAD to identify DNA repair-related DEGs and
309   constructed and validated a 6-gene signature for predicting the outcomes of surgically-treated
310   LUAD. Further study of these DNA repair-related genes showed that the prognosis-risk
311   signature of DNA repair-related genes can be used as potential predictive markers and
312   therapeutic targets for the surgical treatment of LUAD.

313   Acknowledgements

314   Thanks to all the researchers and staff working for The Cancer Genome Atlas database and Gene
315   Expression Omnibus Database.

316

317   **References**
318   Ashley EA. 2015. The precision medicine initiative: a new national effort. *Jama* 313:2119-2120.
319   10.1001/jama.2015.3595
320   Brinkman JA, Liu Y, and Kron SJ. 2020. Small-molecule drug repurposing to target DNA
321   damage repair and response pathways. *Seminars in Cancer Biology*.
322   10.1016/j.semcancer.2020.02.013
323   Cerfolio RJ, Ghanim AF, Dylewski M, Veronesi G, Spaggiari L, and Park BJ. 2018. The long-
324   term survival of robotic lobectomy for non-small cell lung cancer: A multi-institutional
325   study. *J Thorac Cardiovasc Surg* 155:778-786. 10.1016/j.jtcvs.2017.09.016
326   Chatterjee N, and Walker GC. 2017. Mechanisms of DNA damage, repair, and mutagenesis.
327   *Environ Mol Mutagen* 58:235-263. 10.1002/em.22087

328 Chen FF, Zhang D, Wang YL, and Xiong B. 2013. Video-assisted thoracoscopic surgery
329   lobectomy versus open lobectomy in patients with clinical stage non-small cell lung
330   cancer: a meta-analysis. *Eur J Surg Oncol* 39:957-963. 10.1016/j.ejso.2013.06.016
331 Croteau DL, Popuri V, Opresko PL, and Bohr VA. 2014. Human RecQ helicases in DNA repair,
332   recombination, and replication. *Annu Rev Biochem* 83:519-552. 10.1146/annurev-
333   biochem-060713-035428
334 Csiszar A, Balasubramanian P, Tarantini S, Yabluchanskiy A, Zhang XA, Springo Z, Benbrook
335   D, Sonntag WE, and Ungvari Z. 2019. Chemically induced carcinogenesis in rodent
336   models of aging: assessing organismal resilience to genotoxic stressors in geroscience
337   research. *Geroscience* 41:209-227. 10.1007/s11357-019-00064-4
338 De Miguel-Perez D, Bayarri-Lara CI, Ortega FG, Russo A, Moyano Rodriguez MJ, Alvarez-
339   Cubero MJ, Maza Serrano E, Lorente JA, Rolfo C, and Serrano MJ. 2019. Post-Surgery
340   Circulating Tumor Cells and AXL Overexpression as New Poor Prognostic Biomarkers in
341   Resected Lung Adenocarcinoma. *Cancers (Basel)* 11. 10.3390/cancers11111750
342 Denisenko TV, Budkevich IN, and Zhivotovsky B. 2018. Cell death-based treatment of lung
343   adenocarcinoma. *Cell Death Dis* 9:117. 10.1038/s41419-017-0063-y
344 Gavande NS, VanderVere-Carozza PS, Hinshaw HD, Jalal SI, Sears CR, Pawelczak KS, and
345   Turchi JJ. 2016. DNA repair targeted therapy: The past or future of cancer treatment?
346   *Pharmacol Ther* 160:65-83. 10.1016/j.pharmthera.2016.02.003
347 Gonzalo S, and Coll-Bonfill N. 2019. Genomic instability and innate immune responses to self-
348   DNA in progeria. *Geroscience* 41:255-266. 10.1007/s11357-019-00082-2
349 Gutteridge REA, Ndiaye MA, Liu X, and Ahmad N. 2016. Plk1 Inhibitors in Cancer Therapy:
350   From Laboratory to Clinics. *Molecular Cancer Therapeutics* 15:1427-1435.
351   10.1158/1535-7163.Mct-15-0897
352 Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ, Wu Y-L, and Paz-Ares L. 2017.
353   Lung cancer: current therapies and new targeted treatments. *The Lancet* 389:299-311.
354   10.1016/s0140-6736(16)30958-8
355 Jeggo PA, Pearl LH, and Carr AM. 2016. DNA repair, genome stability and cancer: a historical
356   perspective. *Nat Rev Cancer* 16:35-42. 10.1038/nrc.2015.4
357 Jiang W, Zhu D, Wang C, and Zhu Y. 2020. An immune relevant signature for predicting
358   prognoses and immunotherapeutic responses in patients with muscle‐invasive bladder
359   cancer (MIBC). *Cancer Medicine* 9:2774-2790. 10.1002/cam4.2942
360 Krokan HE, and Bjoras M. 2013. Base excision repair. *Cold Spring Harb Perspect Biol*
361   5:a012583. 10.1101/cshperspect.a012583
362 Lee HJ, Feliers D, Barnes JL, Oh S, Choudhury GG, Diaz V, Galvan V, Strong R, Nelson J,
363   Salmon A, Kevil CG, and Kasinath BS. 2018. Hydrogen sulfide ameliorates aging-
364   associated changes in the kidney. *Geroscience* 40:163-176. 10.1007/s11357-018-0018-
365   y
366 Li A, Yi M, Qin S, Chu Q, Luo S, and Wu K. 2019. Prospects for combining immune checkpoint
367   blockade with PARP inhibition. *J Hematol Oncol* 12:98. 10.1186/s13045-019-0784-8
368 Li B, Ruotti V, Stewart RM, Thomson JA, and Dewey CN. 2010. RNA-Seq gene expression
369   estimation with read mapping uncertainty. *Bioinformatics* 26:493-500.
370   10.1093/bioinformatics/btp692
371 Li L, Wu D, Yu Q, Li L, and Wu P. 2017. Prognostic value of FOXM1 in solid tumors: a
372   systematic review and meta-analysis. *Oncotarget* 8:32298-32308.
373   10.18632/oncotarget.15764
374 Lima ZS, Ghadamzadeh M, Arashloo FT, Amjad G, Ebadi MR, and Younesi L. 2019. Recent
375   advances of therapeutic targets based on the molecular signature in breast cancer:
376   genetic mutations and implications for current treatment paradigms. *J Hematol Oncol*
377   12:38. 10.1186/s13045-019-0725-6

378 Liu Z, Sun Q, and Wang X. 2017. PLK1, A Potential Target for Cancer Therapy. *Translational*
379   *Oncology* 10:22-32. 10.1016/j.tranon.2016.10.003
380 Mansuet-Lupo A, Bobbio A, Blons H, Becht E, Ouakrim H, Didelot A, Charpentier M-C, Bain S,
381   Marmey B, Bonjour P, Biton J, Cremer I, Dieu-Nosjean M-C, Sautès-Fridman C,
382   Régnard J-F, Laurent-Puig P, Alifano M, and Damotte D. 2014. The New Histologic
383   Classification of Lung Primary Adenocarcinoma Subtypes Is a Reliable Prognostic
384   Marker and Identifies Tumors With Different Mutation Status. *Chest* 146:633-643.
385   10.1378/chest.13-2499
386 Mei L, Zhang J, He K, and Zhang J. 2019. Ataxia telangiectasia and Rad3-related inhibitors and
387   cancer therapy: where we stand. *J Hematol Oncol* 12:43. 10.1186/s13045-019-0733-6
388 Nandi D, Cheema PS, Jaiswal N, and Nag A. 2018. FoxM1: Repurposing an oncogene as a
389   biomarker. *Seminars in Cancer Biology* 52:74-84. 10.1016/j.semcancer.2017.08.009
390 Network TCGAR. 2014. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*
391   511:543-550. 10.1038/nature13385
392 Nguyen GH, Dexheimer TS, Rosenthal AS, Chu WK, Singh DK, Mosedale G, Bachrati CZ,
393   Schultz L, Sakurai M, Savitsky P, Abu M, McHugh PJ, Bohr VA, Harris CC, Jadhav A,
394   Gileadi O, Maloney DJ, Simeonov A, and Hickson ID. 2013. A small molecule inhibitor of
395   the BLM helicase modulates chromosome stability in human cells. *Chem Biol* 20:55-62.
396   10.1016/j.chembiol.2012.10.016
397 Nickoloff JA, Jones D, Lee SH, Williamson EA, and Hromas R. 2017. Drugging the Cancers
398   Addicted to DNA Repair. *J Natl Cancer Inst* 109. 10.1093/jnci/djx059
399 O'Connor MJ. 2015. Targeting the DNA Damage Response in Cancer. *Mol Cell* 60:547-560.
400   10.1016/j.molcel.2015.10.040
401 Oskarsdottir GN, Bjornsson J, Jonsson S, Isaksson HJ, and Gudbjartsson T. 2016. Primary
402   adenocarcinoma of the lung - histological subtypes and outcome after surgery, using the
403   IASLC/ATS/ERS classification of lung adenocarcinoma. *Apmis* 124:384-392.
404   10.1111/apm.12522
405 Pearl LH, Schierz AC, Ward SE, Al-Lazikani B, and Pearl FM. 2015. Therapeutic opportunities
406   within the DNA damage response. *Nat Rev Cancer* 15:166-180. 10.1038/nrc3891
407 Ray Chaudhuri A, and Nussenzweig A. 2017. The multifaceted roles of PARP1 in DNA repair
408   and chromatin remodelling. *Nat Rev Mol Cell Biol* 18:610-621. 10.1038/nrm.2017.53
409 Royce GH, Brown-Borg HM, and Deepa SS. 2019. The potential role of necroptosis in
410   inflammaging and aging. *Geroscience* 41:795-811. 10.1007/s11357-019-00131-w
411 Siegel RL, Miller KD, and Jemal A. 2020. Cancer statistics, 2020. *CA Cancer J Clin* 70:7-30.
412   10.3322/caac.21590
413 Tarantini S, Yabluchanskiy A, Csipo T, Fulop G, Kiss T, Balasubramanian P, DelFavero J, Ahire
414   C, Ungvari A, Nyúl-Tóth Á, Farkas E, Benyo Z, Tóth A, Csiszar A, and Ungvari Z. 2019.
415   Treatment with the poly(ADP-ribose) polymerase inhibitor PJ-34 improves
416   cerebromicrovascular endothelial function, neurovascular coupling responses and
417   cognitive performance in aged mice, supporting the NAD+ depletion hypothesis of
418   neurovascular aging. *Geroscience* 41:533-542. 10.1007/s11357-019-00101-2
419 Tessoulin B, Moreau-Aubry A, Descamps G, Gomez-Bougie P, Maïga S, Gaignard A, Chiron D,
420   Ménoret E, Le Gouill S, Moreau P, Amiot M, and Pellat-Deceunynck C. 2018. Whole-
421   exon sequencing of human myeloma cell lines shows mutations related to myeloma
422   patients at relapse with major hits in the DNA regulation and repair pathways. *J Hematol*
423   *Oncol* 11:137. 10.1186/s13045-018-0679-0
424 Ucvet A, Gursoy S, and Yazgan S. 2020. Changes in the Surgical Treatment Strategies for
425   Nonsmall Cell Lung Cancer in the Past 20 Years: A Single-Center Experience. *Turk*
426   *Thorac J* 21:8-13. 10.5152/TurkThoracJ.2019.180124
427 Ungvari Z, Tarantini S, Nyúl-Tóth Á, Kiss T, Yabluchanskiy A, Csipo T, Balasubramanian P,
428   Lipecz A, Benyo Z, and Csiszar A. 2019. Nrf2 dysfunction and impaired cellular

429   resilience to oxidative stressors in the aged vasculature: from increased cellular
430   senescence to the pathogenesis of age-related vascular diseases. *Geroscience* 41:727-
431   738. 10.1007/s11357-019-00107-w
432 Veronesi G, Park B, Cerfolio R, Dylewski M, Toker A, Fontaine JP, Hanna WC, Morenghi E,
433   Novellis P, Velez-Cubian FO, Amaral MH, Dieci E, Alloisio M, and Toloza EM. 2018.
434   Robotic resection of Stage III lung cancer: an international retrospective study†.
435   *European Journal of Cardio-Thoracic Surgery* 54:912-919. 10.1093/ejcts/ezy166
436 Vlotides G, Eigler T, and Melmed S. 2007. Pituitary tumor-transforming gene: physiology and
437   implications for tumorigenesis. *Endocr Rev* 28:165-186. 10.1210/er.2006-0042
438 Westaway DD, Toon CW, Farzin M, Sioson L, Watson N, Brady PW, Marshman D, Mathur MM,
439   and Gill AJ. 2013. The International Association for the Study of Lung Cancer/American
440   Thoracic Society/European Respiratory Society grading system has limited prognostic
441   significance in advanced resected pulmonary adenocarcinoma. *Pathology* 45:553-558.
442   10.1097/PAT.0b013e32836532ae
443 Wolfe AR, Wald P, Webb A, Sebastian N, Walston S, Robb R, Chen W, Vedaie M, Dillhoff M,
444   Frankel WL, Kwon W, Jang JY, and Williams TM. 2020. A microRNA-based signature
445   predicts local-regional failure and overall survival after pancreatic cancer resection.
446   *Oncotarget* 11:913-923. 10.18632/oncotarget.27496
447 Xiea Y, and Wangb R. 2016. Pttg1 Promotes Growth of Breast Cancer through P27 Nuclear
448   Exclusion. *Cell Physiol Biochem* 38:393-400. 10.1159/000438660
449 Yan Z, Wang Q, Lu Z, Sun X, Song P, Dang Y, Xie L, Zhang L, Li Y, Zhu W, Xie T, Ma J, Zhang
450   Y, and Guo X. 2020. OSluca: An Interactive Web Server to Evaluate Prognostic
451   Biomarkers for Lung Cancer. *Frontiers in Genetics* 11. 10.3389/fgene.2020.00420
452 Zhu X, Tian X, Yu C, Shen C, Yan T, Hong J, Wang Z, Fang JY, and Chen H. 2016. A long non-
453   coding RNA signature to improve prognosis prediction of gastric cancer. *Mol Cancer*
454   15:60. 10.1186/s12943-016-0544-0
455

Clinic pathological characteristics of surgical LUAD patients from the training(TCGA) and validation sets(GSE31210, GSE37745)

Characteristics	TCGA (n = 260)		GSE31210 (n = 226)		GSE37745 (n = 106)	
	Number of cases %		Number of cases %		Number of cases %	
Platform			GPL570		GPL570	
Noncancerous samples	59		0		0	
Surgically-resected LUAD samples	201		226		106	
Age (years)						
≥65	99	49.3	62	27.4	52	49.1
<65	97	48.3	164	72.6	54	50.9
Gender						
Male	91	45.3	105	46.5	46	43.4
Female	110	54.7	121	53.5	60	56.6
Stage						
Stage 2-4	95	47.3	58	25.7	36	34.0
Stage 1	106	52.7	168	74.3	70	66.0
T						
T3-4	30	14.9				
T1-2	169	84.1				
N						
N1-3	66	32.8				
N0	128	63.7				
M						
M1-Mx	57	28.4				
M0	144	71.6				
Smoking						
Yes			111	49.1		
No			115	50.9		

2 **Notes: Abbreviations: LUAD, lung adenocarcinoma; T, tumor size; N, lymph node status; M,**
3 **metastasis.**

The information of 6 prognostic mRNAs importantly associated with overall survival in patients with surgical LUAD

mRNA	Ensemble ID	Location	Coefficient	HR	P-value
PLK1	ENSG00000166851	Chr16:23,677,656-23,690,367	0.555713	1.743183	0.043687
FOXM1	ENSG00000111206	Chr12: 2,857,681-2,877,155	-0.5696	0.565754	0.024525
PTTG1	ENSG00000164611	Chr5: 160,421,855-160,428,739	0.291471	1.338395	0.064214
CCNO	ENSG00000152669	Chr5: 55,231,152-55,233,608	-0.30485	0.737234	0.004893
HIST3H2A	ENSG00000181218	Chr1: 228,456,979-228,457,873	-0.17162	0.842298	0.030497
BLM	ENSG00000197299	Chr15: 90,717,346-90,816,166	0.442472	1.556551	0.119493

2 **Notes: Abbreviations: LUAD, lung adenocarcinoma; HR, hazard ratio; ID, identity**

Figure 1

Flow diagram of methods for developing and verifying the 6 DNA repair-related prognostic genes signature.
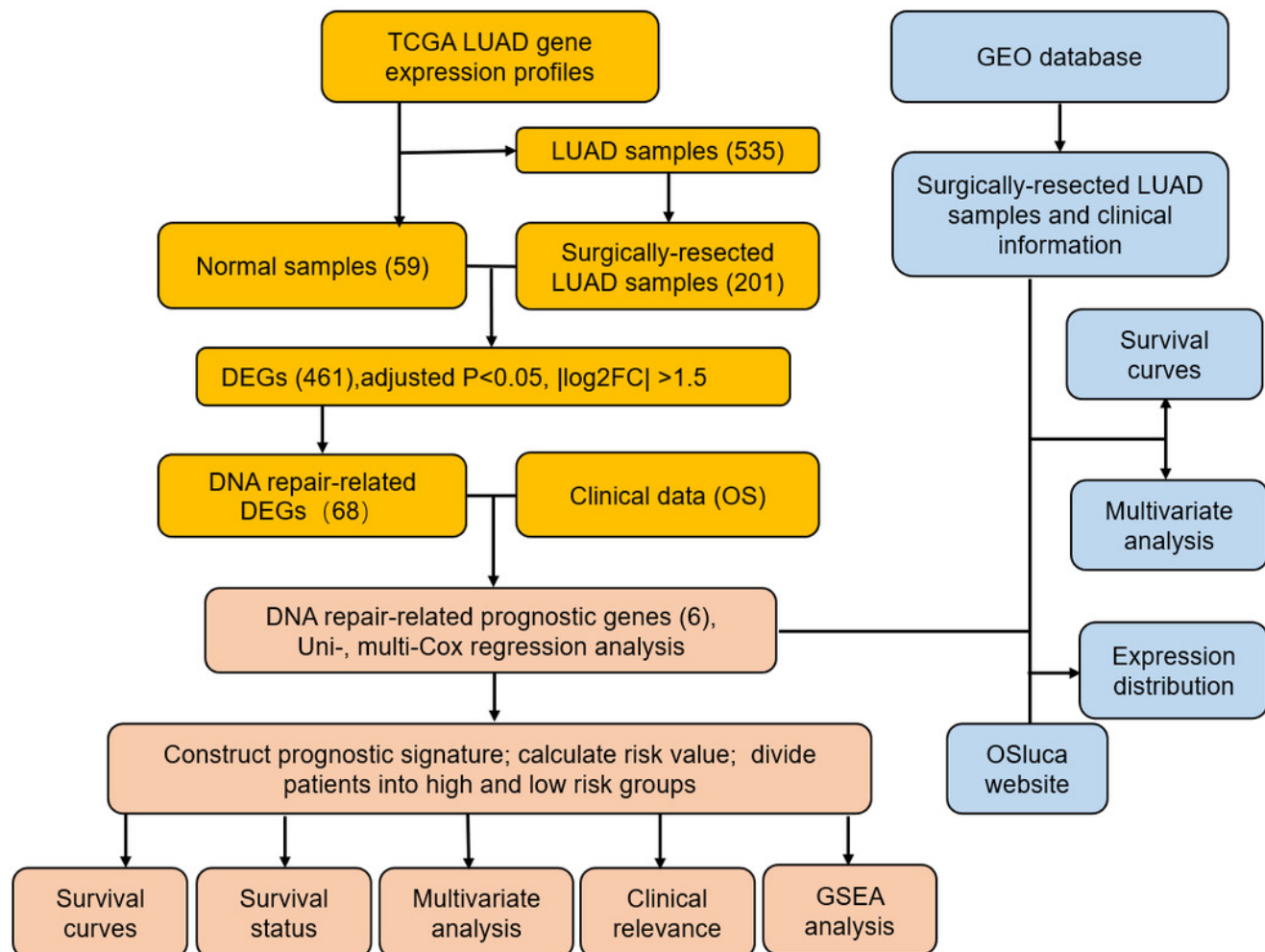
Figure 2

Identification of DNA repair-related prognostic genes.

(A) Volcano plot of DNA repair-related DEGs. |log2(Fold Change)| > 1.5 and FDR < 0.05 were set as screening criteria. The green, red and black dots represented the down-, up-regulated DNA repair-related DEGs and genes that were not satisfied the screening criteria, respectively. (B) The heatmap of all 68 DNA repair-related DEGs distributions. (C) The Boxplot of all 68 DNA repair-related DEGs distributions. (D) The forest plot of 23 prognostic DNA repair-related DEGs screened out by univariate Cox regression. DEGs, Differentially expressed genes; N, Normal samples; T, Surgical lung adenocarcinoma.
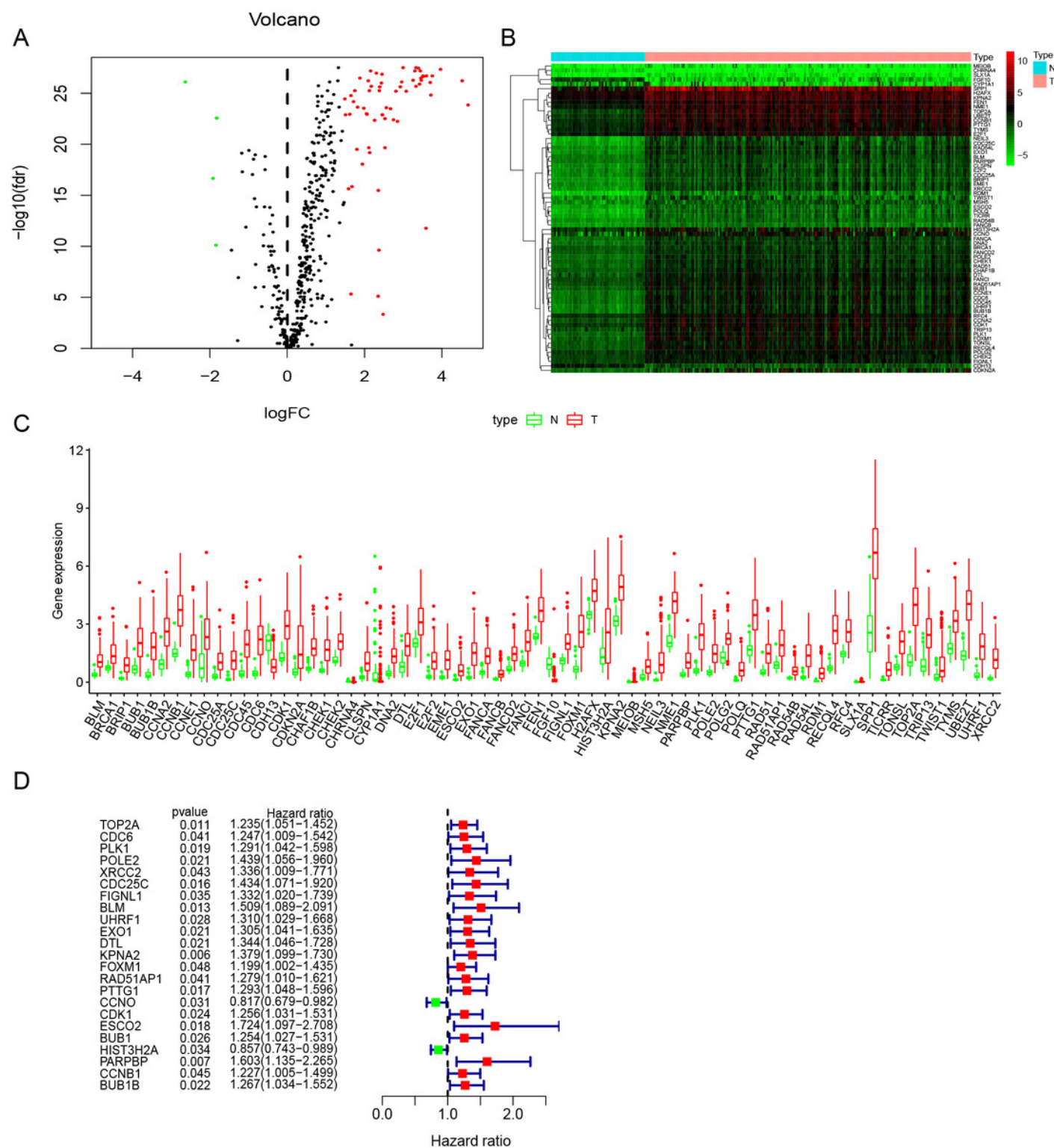
Figure 3

The performance of DNA repair-related prognostic gene signature.

(A) 6 prognostic DNA repair-related DEGs screened out by multivariate Cox regression to construct signature and calculate risk score, the distribution of risk score of each patient. (B) The distribution of survival status of each patient. (C) The heatmap of 6 DNA repair-related signature DEGs distributions in risk score groups. (D) Kaplan-Meier survival curves for overall survival in the DNA repair-related risk score groups. (E) Univariate Cox hazard ratio analysis shown the signature risk score and other clinicopathological features related to overall survival of surgical LUAD. (F) Multivariate Cox hazard ratio analysis shown the signature risk score and other clinicopathological features related to overall survival of surgical LUAD. High: high risk; Low: low risk; gender:0 (female); 1 (male); stage: 1, 2, 3, 4; T:1, 2, 3, 4;N:0, 1, 2, 3, M:0, 1.
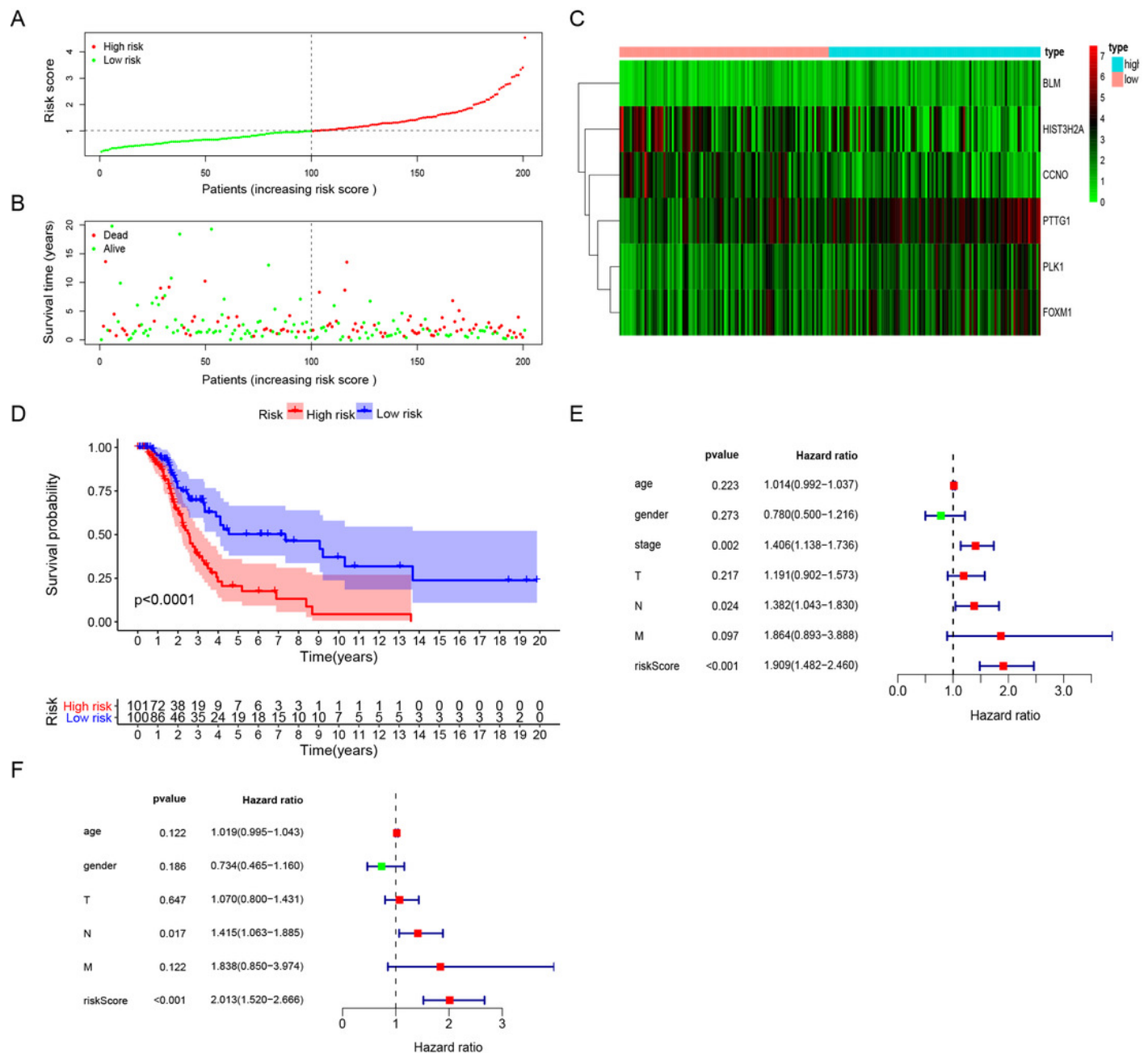
Figure 4

Development and validation of DNA repair-related signature for survival prediction.

(A-B) The GSE31210 expression data were standardization and distribution. (C-D) The GSE37745 expression data were standardization and distribution. (E) Kaplan-Meier survival curves of overall survival in the DNA repair-related risk score groups in GSE31210 set. (F) Kaplan-Meier survival curves of overall survival in the DNA repair-related risk score groups in GSE37745 set. (G) Multivariate Cox hazard ratio analysis shown the signature risk score and other clinicopathological features related to overall survival of GSE31210. (H) Multivariate Cox hazard ratio analysis shown the signature risk score and other clinicopathological features related to overall survival of GSE37745. gender:0 (female); 1 (male); stage: 1, 2, 3, 4.

A GSE31210

B GSE31210

C GSE37745

D GSE37745

E

F

G

H
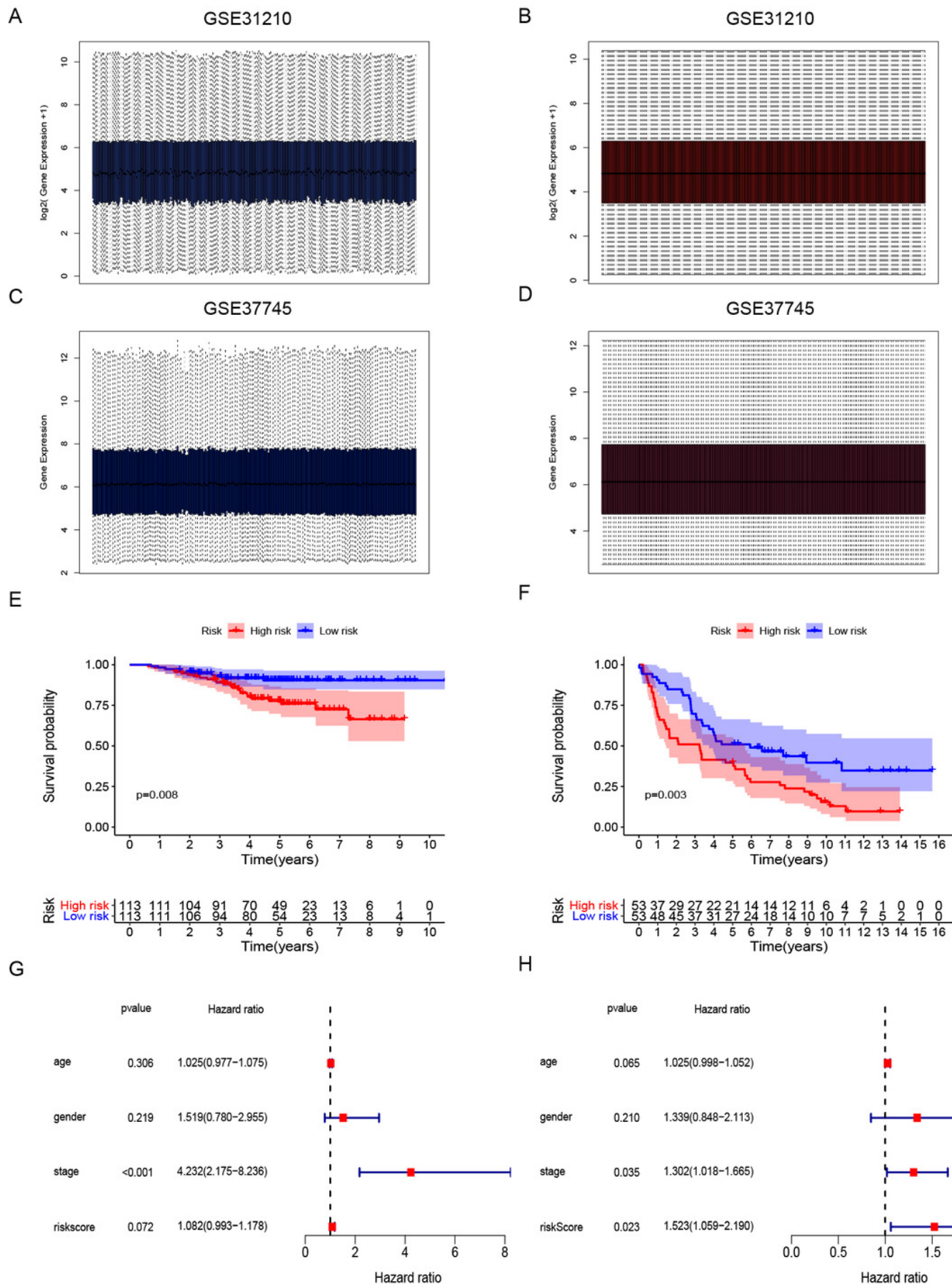
Figure 5

Forest plot of 6 DNA repair-related gene in Osluca online database.

The hazard ratio of each DNA repair-related gene of the signature in the multiple lung cancer datasets (A) PLK1, (B) FOXM1, (C) PTTG1, (D) CCNO, (E) HIST3H2A, (F) BLM.
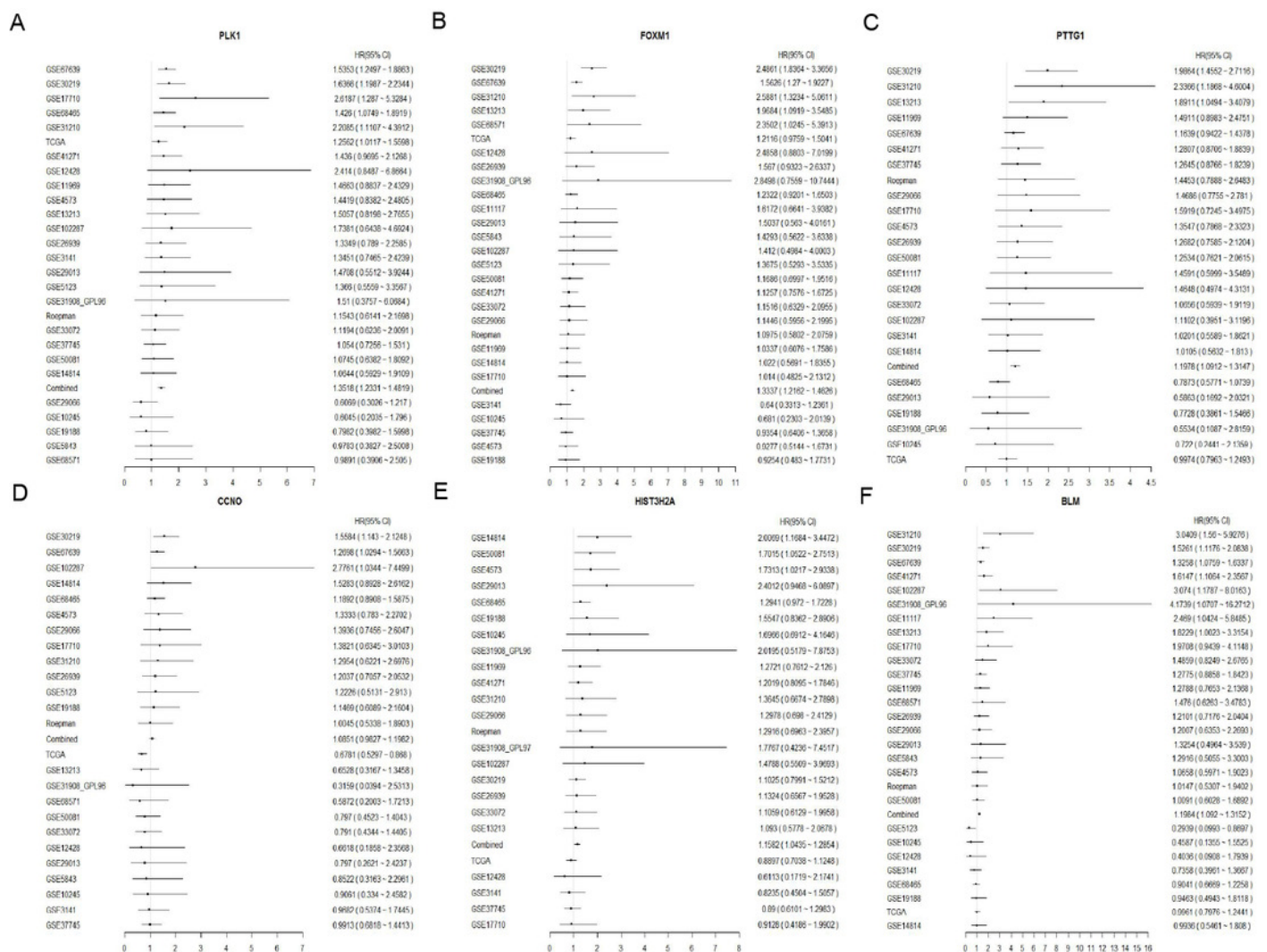
Figure 6

Relationship between 6 DNA repair-related gene signature risk score and clinicopathological features.

(A) The relationship between risk score distribution and clinical parameters stratification. Including age, gender, tumor pathologic stage, T stage, N stage and M stage in TCGA surgical LUAD. (B) The relationship between risk score distribution and clinical parameters stratification. Including age, gender, tumor pathologic stage and smoking in GSE31210 surgical LUAD. (C) The relationship between risk score distribution and clinical parameters stratification. Including age, gender, and tumor pathologic in GSE37745 surgical LUAD. *Represent for p-value<0.05; *Represent for p-value<0.01; NS represent no statistical difference.
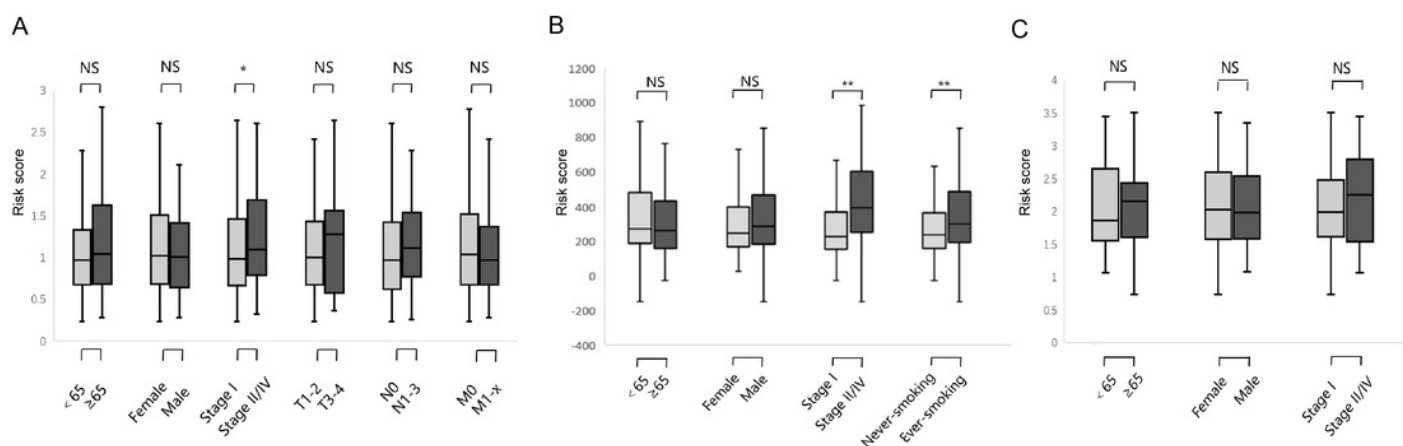
Figure 7

6 gene sets significantly enriched in the surgical LUAD phenotype using GSEA.

Including (A) cell cycle, (B) oocyte meiosis, (C) DNA replication, (D) homologous recombination, (E) mismatch repair and (F) nucleotide excision repair pathway.