Development of a novel embryonic germline gene-related prognostic model of lung adenocarcinoma

Linjun Liu [1], Ke Xu [Corresp., 2], Yubai Zhou [Corresp. 1]

[1] Department of Biotechnology, College of Life Science & Chemistry, Beijing University of Technology, Chaoyang, China, Beijing, China

[2] National Institute for Viral Disease Control and Prevention, China CDC, NHC Key Laboratory of biosafety, Beijing, China

Corresponding Authors: Ke Xu, Yubai Zhou
Email address: xuke@ivdc.chinacdc.cn, zhouyubai@bjut.edu.cn

Background: Emerging evidence implicates the correlation of embryonic germline genes with the tumor progress and patient's outcome. However, the prognostic value of these genes in lung adenocarcinoma (LUAD) has not been fully studied. Here we systematically evaluated this issue, and constructed a novel signature and a nomogram associated with embryonic germline genes for predicting the outcomes of lung adenocarcinoma.

Methods: The LUAD cohorts retrieved from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) database were used as training set and testing set, respectively. The embryonic germline genes were downloaded from the website https://venn.lodder.dev. Then, the differentially expressed embryonic germline genes (DEGGs) between the tumor and normal samples were identified by limma package. The functional enrichment and pathway analyses were also performed by clusterProfiler package. The prognostic model was constructed by the least absolute shrinkage and selection operator (LASSO)-Cox regression method. Survival and Receiver Operating Characteristic (ROC) analyses were performed to validate the model using training set and four testing GEO datasets. Finally, a prognostic nomogram based on the signature genes was constructed using multivariate regression method. **Results:** Among the identified 269 DEGGs, 249 were up-regulated and 20 were down-regulated. GO and KEGG analyses revealed that these DEGGs were mainly enriched in the process of cell proliferation and DNA damage repair. Then, 103 DEGGs with prognostic value were identified by univariate Cox regression and further filtered by LASSO method. The resulting sixteen DEGGs were included in step multivariate Cox regression and an eleven embryonic germline gene related signature (EGRS) was constructed. The model could robustly stratify the LUAD patients into high-risk and low-risk groups in both training and testing set, and low-risk patients had much better outcomes. The multi-ROC analysis also showed that the EGRS model had the best predictive efficacy compared with other common clinicopathological factors. The EGRS model also showed robust predictive ability in 4 independent external

datasets, and the area under curve (AUC) was 0.726 (GSE30219), 0.764 (GSE50081), 0.657 (GSE37745) and 0.668 (GSE72094). More importantly, the expression level of some genes in ERGS has a significant correlation with the progression of LUAD clinicopathology, suggesting these genes might play an important role in the progression of LUAD. Finally, based on EGRS genes, we built and calibrated a nomogram for conveniently evaluating patients' outcomes.

1 **Development of a novel embryonic germline gene-related prognostic model of lung**
2 **adenocarcinoma**

4 Linjun Liu[1], Ke Xu[2]*, Yubai Zhou[1]*

5 1. Department of Biotechnology, College of Life Science & Chemistry, Beijing University of
6   Technology, Chaoyang, Beijing, China

7 2. NHC Key Laboratory of biosafety, National Institute for Viral Disease Control and Prevention,
8   China CDC, Changping, Beijing, China

9 * Corresponding to:

10 Ke Xu, Ph.D., NHC Key Laboratory of biosafety, National Institute for Viral Disease Control and

11 Prevention, China CDC, Changbai 155#, District of Changping, Beijing, 102206, China.

12 E-mail: xuke@ivdc.chinacdc.cn

13 Yubai Zhou, Ph.D., Department of Biotechnology, College of Life Science & Chemistry, Beijing

14 University of Technology, Pingleyuan 100#, District of Chaoyang, Beijing 100124, China.

15 E-mail: zhouyubai@bjut.edu.cn

25

26

27

28

29

Abstract

Background: Emerging evidence implicates the correlation of embryonic germline genes with the tumor progress and patient's outcome. However, the prognostic value of these genes in lung adenocarcinoma (LUAD) has not been fully studied. Here we systematically evaluated this issue, and constructed a novel signature and a nomogram associated with embryonic germline genes for predicting the outcomes of lung adenocarcinoma. **Methods:** The LUAD cohorts retrieved from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) database were used as training set and testing set, respectively. The embryonic germline genes were downloaded from the website https://venn.lodder.dev. Then, the differentially expressed embryonic germline genes (DEGGs) between the tumor and normal samples were identified by limma package. The functional enrichment and pathway analyses were also performed by clusterProfiler package. The prognostic model was constructed by the least absolute shrinkage and selection operator (LASSO)-Cox regression method. Survival and Receiver Operating Characteristic (ROC) analyses were performed to validate the model using training set and four testing GEO datasets. Finally, a prognostic nomogram based on the signature genes was constructed using multivariate regression method. **Results:** Among the identified 269 DEGGs, 249 were up-regulated and 20 were down-regulated. GO and KEGG analyses revealed that these DEGGs were mainly enriched in the process of cell proliferation and DNA damage repair. Then, 103 DEGGs with prognostic value were identified by univariate Cox regression and further filtered by LASSO method. The resulting sixteen DEGGs were included in step multivariate Cox regression and an eleven embryonic germline gene related signature (EGRS) was constructed. The model could robustly stratify the LUAD patients into high-risk and low-risk groups in both training and testing set, and low-risk patients had much better outcomes. The multi-ROC analysis also showed that the EGRS model had the best predictive efficacy compared with other common clinicopathological factors. The EGRS model also showed robust predictive ability in 4 independent external datasets, and the area under curve (AUC) was 0.726 (GSE30219), 0.764 (GSE50081), 0.657 (GSE37745) and 0.668 (GSE72094). More importantly, the expression level of some genes in ERGS has a significant correlation with the progression of LUAD clinicopathology, suggesting these genes might play an important role in the progression of LUAD. Finally, based on EGRS genes, we built and calibrated a nomogram for conveniently evaluating patients' outcomes.

60

Introduction

As the most common pathological subtype of non-small cell lung cancer (NSCLC), lung adenocarcinoma (LUAD) largely occurs among women and non-smoking populations(Herbst et al. 2018). Benefiting from the advance in surgical technique, targeted therapy, immunotherapy and radiochemotherapy, the treatment of LUAD has undergone revolutionary changes in recent years(Yang et al. 2020). However, due to the lack of sensitive early detection methods, most patients are in advanced stages at the time of their first diagnosis. The overall 5-year survival rate has not improved significantly(Chen et al. 2014).

69   Many cancers activate genes normally associated with different developmental states,
70 including germ cell-specific genes, which suggests that some unknown mechanisms which are
71 normally limited to the development of germ cells maybe involved in the process of
72 tumorigenesis(Bruggeman et al. 2018; Erenpreisa & Cragg 2010; Kho et al. 2004; Lepourcelet et
73 al. 2005; McFarlane & Wakeman 2017; Simpson et al. 2005). With the first discovery of MAGE-1
74 in melanoma cells, a series of genes that are normally restricted to the testis, such as BAGE,
75 GAGE, HOM-MEL-40, SYCP1, and NY-ESO-1 have been found in various tumors, and these
76 genes are collectively called cancer/testis (CT) genes (Almeida et al. 2009; Boël et al. 1995; Chen
77 et al. 1997; De Backer et al. 1999; Türeci et al. 1996; Türeci et al. 1998; van der Bruggen et al.
78 1991). Although the driving role of these CT genes in tumorigenesis and development has not been
79 fully studied, there is evidence that they may play a role in the early stage of oncogenesis and the
80 maintenance of tumor phenotypes(Bruggeman et al. 2018). However, these previously identified
81 CT genes are mixed with genes derived from testicular somatic cells and are not truly germline-
82 specific genes. Currently, using bioinformatic methods, Bruggeman et.al. identified 672 so called
83 embryonic germ cell genes which are restricted to the germline and also expressed in a variety of
84 tumor tissues, including lung cancer(Bruggeman et al. 2020). However, the level of expression of
85 these genes in lung adenocarcinoma and its relationship with patient prognostics have not been
86 fully studied.
87   In this study, we systematically analyzed the expression levels of embryonic germline cell-
88 specific genes in LUAD cohort and constructed a novel prognostic signature and nomogram,
89 which can be used to predict the overall survival of LUAD patients.

90

Material and Methods

Data source and preprocessing

93 Data were collected and preprocessed as previously described(Yue et al. 2019). Briefly, as the
94 training dataset, the expression profile and clinical information of LUAD cohort were downloaded
95 from TCGA database using GDC tool and summarized into an expression and a clinical matrix,
96 respectively. The ensemble ids in the expression matrix were converted into gene symbols
97 according to the annotation file and the multiple expression data of a single gene were replaced by
98 an average value. Genes that are not expressed in all samples are also removed from the expression
99 matrix. For testing sets, we used lung adenocarcinoma as a keyword to search in the GEO database.
100 After removing the datasets which are devoid of clinical data or expression data of genes in the
101 model, four independent GEO datasets GSE30219, GSE50081, GSE37745 and GSE72094 were
102 selected and downloaded via GEOquery package(Davis & Meltzer 2007). For pan-cancer analyses,
103 the expression profiles and clinical data of 33 TCGA cancer types were downloaded from TCGA
104 database, and preprocessed according to previously mentioned procedures. Tumor cohorts with
105 normal samples < 5 were not included in the pan-cancer expression analysis. The gene expression
106 values were all converted to transcripts per million transcripts (TPM) to facilitate subsequent
107 analysis. The protein expression data of LUAD were obtained from the Clinical Proteomic Tumor
108 Analysis Consortium (CPTAC) database (https://cptac-data-portal.georgetown.edu/).

Identification of the differentially expressed embryonic germline genes (DEGGs)

The embryonic germline genes were downloaded from the website https://venn.lodder.dev(Bruggeman et al. 2020), and their expression data were extracted into a matrix. Then, the differentially expressed embryonic germline genes (DEGGs) between tumor and normal samples were identified using methods mentioned previously (Yue et al. 2019). The multiple comparisons were controlled using the false discovery rate (FDR) (Benjamini & Hochberg 1995). The screening criteria were |log2Fold Chang|>1 and FDR<0.05.

Functional enrichment and pathway analysis

The Gene Ontology(GO) enrichment and Kyoto Encyclopedia of Genes and Genomes(KEGG) pathway analysis were performed according to previously described methods (Yu et al. 2012; Yue et al. 2019). P. adjust (FDR) <0.05 was considered statistically significant.

Construction of EGRS in training dataset

The EGRS model was established using methods described previously(Yue et al. 2019). Briefly, the 454 LUAD samples were included in the univariate Cox regression analysis to determine the prognostic value of DEGGs. P < 0.05 was considered statistically significant. The prognostic DEGGs were further screened out using LASSO regression via glmnet package in R software for avoiding model overfitting(Friedman et al. 2010). Then, the resulting key DEGGs were included in multivariate Cox analysis, and risk score formula was constructed as follow:

$$\text{Risk score} = \sum \beta * \text{gene}$$

The β and gene represent the coefficient of multivariate Cox regression and the expression value of corresponding DEGGs in EGRS, respectively. The * represents the multiplication operator.

Validation of the EGRS in the training and testing sets

To further validate the expression pattern, the protein expression profiles of DEGGs in EGRS between normal and tumor samples were analyzed using proteomic data of LUAD cohort obtained from CPTAC database. Then, the predicted effects of the EGRS was evaluated in the training and testing sets, respectively. Samples in the training set were divided into high- and low-risk group according to the median of risk score values. The risk score distribution, survival status and the expression level of 11 DEGGs in EGRS between the high- and low-risk samples were evaluated. In addition, we divided the patients in training set into two groups according to commonly used clinicopathological factors and analyze the differentiation of the expression levels of 11 DEGGs between the two groups by student's t-test. P<0.05 was considered statistically significant. Kaplan-Meier (KM) survival analysis and receiver operating characteristic (ROC) curve were used to evaluate the predicting power of the EGRS, and the prognostic performance of other clinicopathological factors was also compared with that of EGRS model. The KM plot and ROC curve were also used to assess the generalization power of EGRS model in four independent GEO datasets. Finally, a pan-cancer expression and prognostic analyses of 11 DEGGs were performed to extend our findings to other types of cancers.

146 **Construction and evaluation of a nomogram**

147 For predicting the OS of individual LUAD patient, Nomogram was generated based on the results
148 of the multivariate Cox analysis to predict 1-, 3-, and 5-year OS. The performance of the
149 nomogram was evaluated using calibration plots.

150 **Result**

151 **Identification the DEGGs in LUAD**

152 The transcriptome data of LUAD cohort which contains 497 tumor and 54 normal samples were
153 downloaded from TCGA database, and the corresponding expression profiles of embryonic
154 germline genes were extracted according to the data published previously(Bruggeman et al. 2020).
155 The DEGGs between normal and tumor samples were identified by limma package in R
156 software(Ritchie et al. 2015). Totally, 269 DEGGs were screened out, among which, the
157 upregulated DEGGs were 249, and the down-regulated DEGGs were 20 (Fig. 1A, Supplementary
158 TableS1). The representative DEGGs are shown by heat map (Figure 1B) and box plot (Figure
159 1C), respectively.

160 **Functional enrichment and KEGG analysis of DEGGs**

161 The GO enrichment analysis of the DEGGs revealed that the DEGGs were mainly enriched in
162 biological process of cell proliferation and DNA damage repair, such as nuclear division
163 (GO:0000280), DNA replication (GO:0006260), cell cycle checkpoint (GO:0000075), double-
164 strand break repair (GO:0006302) and interstrand cross-link repair (GO:0036297) (Fig. 2A, B).
165 The KEGG analysis showed that the enriched pathways were cell cycle (hsa04110), homologous
166 recombination (hsa03440), and Fanconi anemia pathway (hsa03460) (Fig. 2C, D). The detailed
167 GO terms of the three categories, biological process (BP), cellular component (CC), molecular
168 function (MF) and KEGG results were presented in Supplementary TableS2 and Supplementary
169 TableS3, respectively.

170 **The construction of embryonic germline gene related prognostic model**

171 For EGRS construction, the univariate Cox regression analysis was conducted and the resulting
172 103 DEGGs with prognostic value (p<0.05, Fig.3A) were further screened by LASSO
173 regression. Then, the sixteen key DEGGs were selected to performed the multivariate Cox
174 regression analysis (Fig.3B, C). Finally, a prognostic signature containing eleven DEGGs
175 (RAD54L, ZNF322, CENPI, IGF2BP1, IGF2BP3, RAD51AP1, E2F7, HMMR, DNAJC5B,
176 ADAMTS12, NANOS1) was constructed and the risk score formula was presented as follow:
177 Risk score = (-0.4294) * RAD54L - 0.2411 * ZNF322 - 0.4645 * CENPI + 0.2243 * IGF2BP1 +
178 0.2095 * IGF2BP3 - 0.5093 * RAD51AP1 + 0.7979 * E2F7 + 0.7968 * HMMR - 0.3272 *
179 DNAJC5B + 0.1997 * ADAMTS12 - 0.5248 * NANOS1.

180 **The validation of the EGRS for survival prediction**

181   Due to the potential inconsistencies in gene expression between the transcriptome and the
182   proteome, we first verified the protein expression of the 11 DEGGs in EGRS in normal and tumor
183   samples through the CPTAC database. Given the limitations of available data, only 4 genes of the
184   11 DEGGs were analyzed. Consistent with the transcriptome data, the protein expression of 4
185   genes (IGF2BP1, IGF2BP3, HMMR and ADAMTS12) were up-regulated in LUAD tumor
186   samples. The log2FoldChanges of IGF2BP1, IGF2BP3, HMMR and ADAMTS12 were 0.49, 1.14,
187   0.50 and 0.87 and the p-values of the 4 genes were 2.7E-02, 9.22E-08, 1.32E-07 and 7.27E-18,
188   respectively. In addition, the pan-cancer expression analysis also revealed that 11 DEGGs in EGRS
189   showed differential expression in most types of tumors, but diverse expression trends between
190   tumor and paired adjacent normal tissues of some types of cancers were observed, suggesting that
191   the DEGGs may play different roles in the development of different types of tumors
192   (Supplementary figure S1). Then, Kaplan-Meier (KM) survival analysis indicated that the overall
193   survival between the high- and low-risk groups in training set was significantly different, and the
194   low-risk patients had significantly better outcomes than that of high-risk patients (P = 4.083e-09)
195   (Fig. 4A). Besides, our model had the best predictive power, compared with other
196   clinicopathological factors such as age, gender, TNM and stage. Its area under curve (AUC) of
197   ROC was 0.790 (Fig. 4B). The DEGGs in EGRS indicated differential expression between high-
198   and low-risk samples (Fig. 4C), and with the increase of risk score, the proportion of death cases
199   was also increasing (Fig. 4D, E). Between different pathological groups, the expression levels of
200   genes in EGRS also showed significant difference (Fig. 5). Contrary to CENPI, the expressions of
201   DNAJC5B in elderly patients (>65 years old) were higher than those in younger patients (<= 65
202   years old) (Figure 5A, B). The HMMR, RAD51AP1 were highly expressed in men than in women
203   (Figure 5 D, E). ZNF322 was related to T stage, and its expressions in T3-4 were significantly
204   lower than those in T1-2 (Figure 5G). The CENPI, DNAJC5B, HMMR, RAD51AP1 and ZNF322
205   were correlated to N stage. Except for ZNF322, other genes were highly expressed in N1-2 stage
206   (Figure 5I-L). The expressions of ADAMTS12 in M1 stage were lower than those in M0 stage
207   (Figure 5N). The expressions of DNAJC5B, HMMR, IGF2BP1, IGF2BP3, RAD51AP1, ZNF322
208   were correlated to the pathological stage. Except for ZNF322 and DNAJC5B, the expressions of
209   other genes in stage III-IV were significantly higher than those in stage I-II (Figure 5P-U). More
210   importantly, the risk scores in the advanced LUAD were significantly higher than those in the early
211   stage (Figure 5H, M and V), although its distribution had no statistical difference in some
212   pathological statuses such as age, gender and tumor metastasis (Figure 5C, F and O).
213   The univariate and multivariate Cox analyses suggested the risk score was an independent
214   prognostic factor (Fig. 6A, B). The pan-cancer prognostic analysis also indicated the 11 DEGGs
215   had prognostic value in some cancer types. Likewise, different prognostic effects based on cancer
216   type were observed in some DEGGs (the detail results were packaged into a Supplementary file).
217   In addition, the predictive power of the EGRS was further validated in four GEO testing datasets.
218   Consistent with the training set, our model also stratified the samples in the testing sets into low-
219   risk and high-risk groups, and the low-risk patients had better outcomes (Figure 6 C-F). The AUC
220   results also showed that our model had stable predictive ability. The AUC in four independent
221   GEO datasets were 0.726 (GSE30219), 0.764 (GSE50081), 0.657 (GSE37745) and 0.668

222 (GSE72094) (Fig. 6G-J).

Construction and evaluation of a nomogram

224 In order to more straightforwardly predict the prognosis of LUAD patients, a nomogram was
225 constructed (Fig. 7A). According to the expression level of 11 genes in an individual patient, a
226 numeric value was calculated to directly predict the 1-, 3-year and 5-year prognostic survival rate.
227 The calibration plots showed that the predictive values of OS probabilities in 1-, 3-, and 5-year fit
228 well with the observation values, suggesting that the nomogram is suitable for predicting the
229 prognostic survival rate of LUAD patients (Figs. 7B, 7C, 7D).

230

Discussion

232 Although the data published recently on lung cancer in the United States are encouraging,
233 lung cancer is still a malignant tumor with high morbidity and mortality in the world(Siegel et al.
234 2020; Sung et al. 2021). In China, the incidence of lung cancer ranks the first and the second among
235 men and women, respectively, and it is still the malignant disease with the highest mortality
236 rate(Zhang et al. 2020b). The lung adenocarcinoma is the most common pathological subtype of
237 non-small cells lung cancer (NSCLC), which accounts for about 40% of all lung cancer
238 cases(Kleczko et al. 2019). Benefiting from carrying a higher proportion of actionable mutations,
239 lung adenocarcinoma has more treatment options than other subtypes of lung cancer, however, its
240 five-year overall survival rate remains unsatisfactory(Zhang et al. 2019b). Accumulating evidence
241 indicates that patients with early stage of disease have a better prognosis, so finding more effective
242 early diagnostic biomarkers or therapeutic targets is important for developing more effective
243 treatment regimens, prolonging patient's survival time and improving their quality of life.

244 The various types of human cells are derived from a single fertilized egg cell through
245 differentiation mechanisms that we do not fully understand(McKenna & Gagnon 2019). In this
246 process, specific types of genes in differentiating cells are expressed in a well-organized
247 spatiotemporal sequence under the cues of various internal and external factors, and finally
248 generate tissue cells with distinct morphology, function and gene expression profiles(Trott &
249 Martinez Arias 2013). Germline cell-specific genes are a class of genes that are expressed only in
250 germ cells. Accumulating evidence shows that tumor cells and germ cells have similar behaviors,
251 so it is speculated that tumor cells may adopt certain similar regulatory mechanisms in the process
252 of tumorigenesis and development(Bruggeman et al. 2018; Simpson et al. 2005). Indeed, follow-
253 up studies have detected multiple germ cell-specific genes in different tumor tissues(Erenpreisa &
254 Cragg 2010; McFarlane & Wakeman 2017; Whitehurst 2014). The products of these genes are
255 regarded as ideal immunotherapy targets due to their restricted range of expression and have
256 attracted more and more attention.

257 Currently, Bruggeman et al. identified 672 true embryonic germ cell-specific genes by
258 removing contaminated somatic genes(Bruggeman et al. 2020). However, the expression status
259 and their prognostic value of these genes have not been fully studied in lung adenocarcinoma. To
260 take these issues, we systematically studied the expression levels of the above-mentioned 672

261   genes in the TCGA LUAD cohort, evaluated their correlation with the prognosis of LUAD
262   patients, and finally constructed and validated a prognostic model consisting of 11 DEGGs. Our
263   model can well stratify the patients into high- and low-risk groups and robustly predict the
264   outcomes of patients in both training and testing sets. Among the 11 DEGGs, ZNF322 (also known
265   as ZNF322A, ZNF388 or ZNF489) is a member of the zinc finger transcription factor family, and
266   may act as a positive regulator in MAPK signaling pathway(Li et al. 2004). Recent studies have
267   shown that the overexpression of ZNF322 is related to the oncogenesis of many tumors including
268   lung cancer through a variety of mechanisms, such as directly suppressing the expression of c-
269   Myc to promote cell stemness or up-regulating the transcription of α-adducin (ADD1) and cyclin
270   D1 (CCND1) to promote tumor growth and metastasis(Jen et al. 2017; Jen et al. 2019; Liao et al.
271   2017; Williams et al. 2014). DNA binding protein RAD51AP1 is a component of DNA damage
272   repair system, which plays an important role in RAD51-mediated homologous
273   recombination(Pires et al. 2017). Further studies indicates that overexpression of RAD51AP1
274   promotes the proliferation of lung cancer cells and correlates to poor prognosis(Chudasama et al.
275   2018; Li et al. 2018b; Wu et al. 2019). HMMR coding a 724 amino acids protein, which may
276   associate with the motility of cancer cells(Sankaran et al. 2012). Quite a few of studies had shown
277   that the highly expression of HMMR is closely related to the prognosis of LUAD patients, and it
278   may promote the progression of lung adenocarcinoma by regulating the metabolic state of tumor
279   cells(He & Zuo 2019; Jiao et al. 2020; Li et al. 2020a; Li et al. 2020c; Liu et al. 2019; Man et al.
280   2014; Shen et al. 2019; Stevens et al. 2017; Zhang et al. 2019a; Zhou et al. 2015). The ADAMTS12
281   is a novel anti-tumor metalloprotease, and the expression is epigenetically silenced in tumor
282   cells(Li et al. 2018a; Moncada-Pazos et al. 2009). A bulk of studies suggest the overexpression of
283   IGF2BP1 and IGF2BP3 facilitate the progress of lung cancer, and relate to the prognosis of LUAD
284   patients(Beljan Perak et al. 2012; Gong et al. 2016; Guo et al. 2021; Huang et al. 2019; Kato et al.
285   2007; Li & Zhan 2020; Li et al. 2020b; Mizutani et al. 2016; Müller et al. 2019; Ohdaira et al.
286   2012; Wang et al. 2020; Xueqing et al. 2020; Zhang et al. 2020a). E2F7 plays an important
287   physiological role in embryonic development and cell cycle regulation (Park et al. 2015). Recent
288   researches show the various non-coding RNA can influence the progress of lung cancer by
289   regulating the expression of E2F7(Liang et al. 2018; Liu et al. 2020; Wang et al. 2021; Yuan et al.
290   2021). NANOS1 was up-regulated in lung cancer, and promoted the progress of tumor
291   progress(Bonnomet et al. 2008; De Keuckelaere et al. 2018). RAD54L is also a potential
292   prognostic biomarker of NSCLC(Tu et al. 2021; Zheng et al. 2021). Literature retrieval showed
293   that the relation of LUAD and DEGGs DNAJC5B, CENPI has not been reported so far. In addition,
294   pan-cancer expression and prognosis analyses also indicated that 11 DEGGs showed differential
295   expression and prognostic value in a variety of tumors, suggesting that these DEGGs may also
296   play a role in the development of tumors other than lung adenocarcinoma, which needed to be
297   validated by further functional experiments.
298   Although we constructed and validated an 11-gene prognostic model, the present study still
299   has several limitations: First, the number of samples in training set is not large enough; secondly,
300   due to tumor heterogeneity and individual differences among LUAD patients, the generalizing
301   abilities of the model need further improvement. Therefore, further model optimization and

302 validation using LUAD cohort with larger samples are needed. In addition, the results of
303 bioinformatics analysis did not carry out related experimental verifications. Therefore, whether the
304 genes in the model play a functional role in the progress of LUAD remains to be resolved. These
305 issues mentioned above will be the main content of our follow-up studies.
306

Conclusion

308   In summary, we used LUAD transcriptome data to identify DEGGs. From the DEGGs, an 11-
309 gene signature and a prognostic nomogram were constructed and validated for predicting the
310 outcomes of LUAD patients. Further studies on these genes will provide a new insight into the
311 potential relationship between tumor microenvironment and LUAD prognosis.
312

Acknowledgments

314 Not applicable
315

Reference

317 Almeida LG, Sakabe NJ, deOliveira AR, Silva MC, Mundstein AS, Cohen T, Chen YT, Chua R, Gurung S, Gnjatic
318   S, Jungbluth AA, Caballero OL, Bairoch A, Kiesler E, White SL, Simpson AJ, Old LJ, Camargo AA, and
319   Vasconcelos AT. 2009. CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis
320   antigens. Nucleic Acids Res 37:D816-819. 10.1093/nar/gkn673
321 Beljan Perak R, Durdov MG, Capkun V, Ivcevic V, Pavlovic A, Soljic V, and Peric M. 2012. IMP3 can predict
322   aggressive behaviour of lung adenocarcinoma. Diagn Pathol 7:165. 10.1186/1746-1596-7-165
323 Benjamini Y, and Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to
324   multiple testing. Journal of the Royal statistical society: series B 57:289-300.
325 Boël P, Wildmann C, Sensi ML, Brasseur R, Renauld JC, Coulie P, Boon T, and van der Bruggen P. 1995. BAGE: a
326   new gene encoding an antigen recognized on human melanomas by cytolytic T lymphocytes. Immunity
327   2:167-175. 10.1016/s1074-7613(95)80053-0
328 Bonnomet A, Polette M, Strumane K, Gilles C, Dalstein V, Kileztky C, Berx G, van Roy F, Birembaut P, and
329   Nawrocki-Raby B. 2008. The E-cadherin-repressed hNanos1 gene induces tumor cell invasion by
330   upregulating MT1-MMP expression. Oncogene 27:3692-3699. 10.1038/sj.onc.1211035
331 Bruggeman JW, Irie N, Lodder P, van Pelt AMM, Koster J, and Hamer G. 2020. Tumors Widely Express Hundreds
332   of Embryonic Germline Genes. Cancers (Basel) 12. 10.3390/cancers12123812
333 Bruggeman JW, Koster J, Lodder P, Repping S, and Hamer G. 2018. Massive expression of germ cell-specific genes
334   is a hallmark of cancer and a potential target for novel treatment development. Oncogene 37:5694-5700.
335   10.1038/s41388-018-0357-2
336 Chen YT, Scanlan MJ, Sahin U, Türeci O, Gure AO, Tsang S, Williamson B, Stockert E, Pfreundschuh M, and Old
337   LJ. 1997. A testicular antigen aberrantly expressed in human cancers detected by autologous antibody
338   screening. Proc Natl Acad Sci U S A 94:1914-1918. 10.1073/pnas.94.5.1914
339 Chen Z, Fillmore CM, Hammerman PS, Kim CF, and Wong KK. 2014. Non-small-cell lung cancers: a heterogeneous
340   set of diseases. Nat Rev Cancer 14:535-546. 10.1038/nrc3775

341   Chudasama D, Bo V, Hall M, Anikin V, Jeyaneethi J, Gregory J, Pados G, Tucker A, Harvey A, Pink R, and Karteris
342   E. 2018. Identification of cancer biomarkers of prognostic value using specific gene regulatory networks
343   (GRN): a novel role of RAD51AP1 for ovarian and lung cancers. Carcinogenesis 39:407-417.
344   10.1093/carcin/bgx122
345   Davis S, and Meltzer PS. 2007. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor.
346   Bioinformatics 23:1846-1847. 10.1093/bioinformatics/btm254
347   De Backer O, Arden KC, Boretti M, Vantomme V, De Smet C, Czekay S, Viars CS, De Plaen E, Brasseur F, Chomez
348   P, Van den Eynde B, Boon T, and van der Bruggen P. 1999. Characterization of the GAGE genes that are
349   expressed in various human cancers and in normal testis. Cancer Res 59:3157-3165.
350   De Keuckelaere E, Hulpiau P, Saeys Y, Berx G, and van Roy F. 2018. Nanos genes and their role in development and
351   beyond. Cell Mol Life Sci 75:1929-1946. 10.1007/s00018-018-2766-3
352   Erenpreisa J, and Cragg MS. 2010. MOS, aneuploidy and the ploidy cycle of cancer cells. Oncogene 29:5447-5451.
353   10.1038/onc.2010.310
354   Friedman J, Hastie T, and Tibshirani R. 2010. Regularization Paths for Generalized Linear Models via Coordinate
355   Descent. J Stat Softw 33:1-22.
356   Gong F, Ren P, Zhang Y, Jiang J, and Zhang H. 2016. MicroRNAs-491-5p suppresses cell proliferation and invasion
357   by inhibiting IGF2BP1 in non-small cell lung cancer. Am J Transl Res 8:485-495.
358   Guo W, Huai Q, Wan H, Guo L, Song P, Gao S, and He J. 2021. Prognostic Impact of IGF2BP3 Expression in Patients
359   with Surgically Resected Lung Adenocarcinoma. DNA Cell Biol 40:316-331. 10.1089/dna.2020.6136
360   He R, and Zuo S. 2019. A Robust 8-Gene Prognostic Signature for Early-Stage Non-small Cell Lung Cancer. Front
361   Oncol 9:693. 10.3389/fonc.2019.00693
362   Herbst RS, Morgensztern D, and Boshoff C. 2018. The biology and management of non-small cell lung cancer. Nature
363   553:446-454. 10.1038/nature25183
364   Huang H, Wang D, Guo W, Zhuang X, and He Y. 2019. Correlated low IGF2BP1 and FOXM1 expression predicts a
365   good prognosis in lung adenocarcinoma. Pathol Res Pract 215:152433. 10.1016/j.prp.2019.152433
366   Jen J, Lin LL, Lo FY, Chen HT, Liao SY, Tang YA, Su WC, Salgia R, Hsu CL, Huang HC, Juan HF, and Wang YC.
367   2017. Oncoprotein ZNF322A transcriptionally deregulates alpha-adducin, cyclin D1 and p53 to promote
368   tumor growth and metastasis in lung cancer. Oncogene 36:5219. 10.1038/onc.2017.203
369   Jen J, Liu CY, Chen YT, Wu LT, Shieh YC, Lai WW, and Wang YC. 2019. Oncogenic zinc finger protein ZNF322A
370   promotes stem cell-like properties in lung cancer through transcriptional suppression of c-Myc expression.
371   Cell Death Differ 26:1283-1298. 10.1038/s41418-018-0204-6
372   Jiao Z, Yu A, He X, Xuan Y, Zhang H, Wang G, Shi M, and Wang T. 2020. Bioinformatics analysis to determine the
373   prognostic value and prospective pathway signaling of miR-126 in non-small cell lung cancer. Ann Transl
374   Med 8:1639. 10.21037/atm-20-7520
375   Kato T, Hayama S, Yamabuki T, Ishikawa N, Miyamoto M, Ito T, Tsuchiya E, Kondo S, Nakamura Y, and Daigo Y.
376   2007. Increased expression of insulin-like growth factor-II messenger RNA-binding protein 1 is associated
377   with tumor progression in patients with lung cancer. Clin Cancer Res 13:434-442. 10.1158/1078-0432.Ccr-
378   06-1297
379   Kho AT, Zhao Q, Cai Z, Butte AJ, Kim JY, Pomeroy SL, Rowitch DH, and Kohane IS. 2004. Conserved mechanisms
380   across development and tumorigenesis revealed by a mouse development perspective of human cancers.
381   Genes Dev 18:629-640. 10.1101/gad.1182504

382 Kleczko EK, Kwak JW, Schenk EL, and Nemenoff RA. 2019. Targeting the Complement Pathway as a Therapeutic
383   Strategy in Lung Cancer. Front Immunol 10:954. 10.3389/fimmu.2019.00954

384 Lepourcelet M, Tou L, Cai L, Sawada J, Lazar AJ, Glickman JN, Williamson JA, Everett AD, Redston M, Fox EA,
385   Nakatani Y, and Shivdasani RA. 2005. Insights into developmental mechanisms and cancers in the
386   mammalian intestine derived from serial analysis of gene expression and study of the hepatoma-derived
387   growth factor (HDGF). Development 132:415-427. 10.1242/dev.01579

388 Li J, Liao Y, Huang J, Sun Y, Chen H, Chen C, Li S, and Yang Z. 2018a. Epigenetic silencing of ADAMTS5 is
389   associated with increased invasiveness and poor survival in patients with colorectal cancer. J Cancer Res Clin
390   Oncol 144:215-227. 10.1007/s00432-017-2545-9

391 Li N, and Zhan X. 2020. Identification of pathology-specific regulators of m(6)A RNA modification to optimize lung
392   cancer management in the context of predictive, preventive, and personalized medicine. Epma j 11:485-504.
393   10.1007/s13167-020-00220-3

394 Li S, Xuan Y, Gao B, Sun X, Miao S, Lu T, Wang Y, and Jiao W. 2018b. Identification of an eight-gene prognostic
395   signature for lung adenocarcinoma. Cancer Manag Res 10:3383-3392. 10.2147/cmar.S173941

396 Li W, Pan T, Jiang W, and Zhao H. 2020a. HCG18/miR-34a-5p/HMMR axis accelerates the progression of lung
397   adenocarcinoma. Biomed Pharmacother 129:110217. 10.1016/j.biopha.2020.110217

398 Li Y, Gu J, Xu F, Zhu Q, Chen Y, Ge D, and Lu C. 2020b. Molecular characterization, biological function, tumor
399   microenvironment association and clinical significance of m6A regulators in lung adenocarcinoma. Brief
400   Bioinform. 10.1093/bib/bbaa225

401 Li Y, Wang Y, Zhang C, Yuan W, Wang J, Zhu C, Chen L, Huang W, Zeng W, Wu X, and Liu M. 2004. ZNF322, a
402   novel human C2H2 Kruppel-like zinc-finger protein, regulates transcriptional activation in MAPK signaling
403   pathways. Biochem Biophys Res Commun 325:1383-1392. 10.1016/j.bbrc.2004.10.183

404 Li Z, Qi F, and Li F. 2020c. Establishment of a Gene Signature to Predict Prognosis for Patients with Lung
405   Adenocarcinoma. Int J Mol Sci 21. 10.3390/ijms21228479

406 Liang R, Xiao G, Wang M, Li X, Li Y, Hui Z, Sun X, Qin S, Zhang B, Du N, Liu D, and Ren H. 2018. SNHG6
407   functions as a competing endogenous RNA to regulate E2F7 expression by sponging miR-26a-5p in lung
408   adenocarcinoma. Biomed Pharmacother 107:1434-1446. 10.1016/j.biopha.2018.08.099

409 Liao SY, Chiang CW, Hsu CH, Chen YT, Jen J, Juan HF, Lai WW, and Wang YC. 2017. CK1δ/GSK3β/FBXW7α
410   axis promotes degradation of the ZNF322A oncoprotein to suppress lung cancer progression. Oncogene
411   36:5722-5733. 10.1038/onc.2017.168

412 Liu C, Li Y, Wei M, Zhao L, Yu Y, and Li G. 2019. Identification of a novel glycolysis-related gene signature that
413   can predict the survival of patients with lung adenocarcinoma. Cell Cycle 18:568-579.
414   10.1080/15384101.2019.1578146

415 Liu P, Wang M, Tang W, Li G, and Gong N. 2020. Circ_SATB2 Attenuates the Anti-Tumor Role of Celastrol in Non-
416   Small-Cell Lung Carcinoma Through Targeting miR-33a-5p/E2F7 Axis. Onco Targets Ther 13:11899-
417   11912. 10.2147/ott.S279434

418 Man Y, Cao J, Jin S, Xu G, Pan B, Shang L, Che D, Yu Q, and Yu Y. 2014. Newly identified biomarkers for detecting
419   circulating tumor cells in lung adenocarcinoma. Tohoku J Exp Med 234:29-40. 10.1620/tjem.234.29

420 McFarlane RJ, and Wakeman JA. 2017. Meiosis-like Functions in Oncogenesis: A New View of Cancer. Cancer Res
421   77:5712-5716. 10.1158/0008-5472.CAN-17-1535

422 McKenna A, and Gagnon JA. 2019. Recording development with single cell dynamic lineage tracing. Development

423   146. 10.1242/dev.169730

424   Mizutani R, Imamachi N, Suzuki Y, Yoshida H, Tochigi N, Oonishi T, Suzuki Y, and Akimitsu N. 2016. Oncofetal
425   protein IGF2BP3 facilitates the activity of proto-oncogene protein eIF4E through the destabilization of
426   EIF4E-BP2 mRNA. Oncogene 35:3495-3502. 10.1038/onc.2015.410

427   Moncada-Pazos A, Obaya AJ, Fraga MF, Viloria CG, Capellá G, Gausachs M, Esteller M, López-Otín C, and Cal S.
428   2009. The ADAMTS12 metalloprotease gene is epigenetically silenced in tumor cells and transcriptionally
429   activated in the stroma during progression of colon cancer. J Cell Sci 122:2906-2913. 10.1242/jcs.050468

430   Müller S, Glaß M, Singh AK, Haase J, Bley N, Fuchs T, Lederer M, Dahl A, Huang H, Chen J, Posern G, and
431   Hüttelmaier S. 2019. IGF2BP1 promotes SRF-dependent transcription in cancer in a m6A- and miRNA-
432   dependent manner. Nucleic Acids Res 47:375-390. 10.1093/nar/gky1012

433   Ohdaira H, Sekiguchi M, Miyata K, and Yoshida K. 2012. MicroRNA-494 suppresses cell proliferation and induces
434   senescence in A549 lung cancer cells. Cell Prolif 45:32-38. 10.1111/j.1365-2184.2011.00798.x

435   Park SA, Platt J, Lee JW, López-Giráldez F, Herbst RS, and Koo JS. 2015. E2F8 as a Novel Therapeutic Target for
436   Lung Cancer. J Natl Cancer Inst 107. 10.1093/jnci/djv151

437   Pires E, Sung P, and Wiese C. 2017. Role of RAD51AP1 in homologous recombination DNA repair and
438   carcinogenesis. DNA Repair (Amst) 59:76-81. 10.1016/j.dnarep.2017.09.008

439   Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK. 2015. limma powers differential expression
440   analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47. 10.1093/nar/gkv007

441   Sankaran D, Pakala SB, Nair VS, Sirigiri DN, Cyanam D, Ha NH, Li DQ, Santhoshkumar TR, Pillai MR, and Kumar
442   R. 2012. Mechanism of MTA1 protein overexpression-linked invasion: MTA1 regulation of hyaluronan-
443   mediated motility receptor (HMMR) expression and function. J Biol Chem 287:5483-5491.
444   10.1074/jbc.M111.324632

445   Shen S, Kong J, Qiu Y, Yang X, Wang W, and Yan L. 2019. Identification of core genes and outcomes in
446   hepatocellular carcinoma by bioinformatics analysis. J Cell Biochem 120:10069-10081. 10.1002/jcb.28290

447   Siegel RL, Miller KD, and Jemal A. 2020. Cancer statistics, 2020. CA Cancer J Clin 70:7-30. 10.3322/caac.21590

448   Simpson AJ, Caballero OL, Jungbluth A, Chen YT, and Old LJ. 2005. Cancer/testis antigens, gametogenesis and
449   cancer. Nat Rev Cancer 5:615-625. 10.1038/nrc1669

450   Stevens LE, Cheung WKC, Adua SJ, Arnal-Estapé A, Zhao M, Liu Z, Brewer K, Herbst RS, and Nguyen DX. 2017.
451   Extracellular Matrix Receptor Expression in Subtypes of Lung Adenocarcinoma Potentiates Outgrowth of
452   Micrometastases. Cancer Res 77:1905-1917. 10.1158/0008-5472.Can-16-1978

453   Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, and Bray F. 2021. Global cancer statistics
454   2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA
455   Cancer J Clin. 10.3322/caac.21660

456   Trott J, and Martinez Arias A. 2013. Single cell lineage analysis of mouse embryonic stem cells at the exit from
457   pluripotency. Biol Open 2:1049-1056. 10.1242/bio.20135934

458   Tu Z, Chen X, Tian T, Chen G, and Huang M. 2021. Prognostic significance of epigenetic regulatory gene expression
459   in patients with non-small-cell lung cancer. Aging (Albany NY) 13:7397-7415. 10.18632/aging.202600

460   Türeci O, Sahin U, Schobert I, Koslowski M, Scmitt H, Schild HJ, Stenner F, Seitz G, Rammensee HG, and
461   Pfreundschuh M. 1996. The SSX-2 gene, which is involved in the t(X;18) translocation of synovial sarcomas,
462   codes for the human tumor antigen HOM-MEL-40. Cancer Res 56:4766-4772.

463   Türeci O, Sahin U, Zwick C, Koslowski M, Seitz G, and Pfreundschuh M. 1998. Identification of a meiosis-specific

464   protein as a member of the class of cancer/testis antigens. Proc Natl Acad Sci U S A 95:5211-5216.
465   10.1073/pnas.95.9.5211

466 van der Bruggen P, Traversari C, Chomez P, Lurquin C, De Plaen E, Van den Eynde B, Knuth A, and Boon T. 1991.
467   A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. Science
468   254:1643-1647. 10.1126/science.1840703

469 Wang Y, He R, and Ma L. 2020. Characterization of lncRNA-Associated ceRNA Network to Reveal Potential
470   Prognostic Biomarkers in Lung Adenocarcinoma. Front Bioeng Biotechnol 8:266. 10.3389/fbioe.2020.00266

471 Wang Y, Wo Y, Lu T, Sun X, Liu A, Dong Y, Du W, Su W, Huang Z, and Jiao W. 2021. Circ-AASDH functions as
472   the progression of early stage lung adenocarcinoma by targeting miR-140-3p to activate E2F7 expression.
473   Transl Lung Cancer Res 10:57-70. 10.21037/tlcr-20-1062

474 Whitehurst AW. 2014. Cause and consequence of cancer/testis antigen activation in cancer. Annu Rev Pharmacol
475   Toxicol 54:251-272. 10.1146/annurev-pharmtox-011112-140326

476 Williams KA, Lee M, Hu Y, Andreas J, Patel SJ, Zhang S, Chines P, Elkahloun A, Chandrasekharappa S, Gutkind
477   JS, Molinolo AA, and Crawford NP. 2014. A systems genetics approach identifies CXCL14, ITGAX, and
478   LPCAT2 as novel aggressive prostate cancer susceptibility genes. PLoS Genet 10:e1004809.
479   10.1371/journal.pgen.1004809

480 Wu Y, Wang H, Qiao L, Jin X, Dong H, and Wang Y. 2019. Silencing of RAD51AP1 suppresses epithelial-
481   mesenchymal transition and metastasis in non-small cell lung cancer. Thorac Cancer 10:1748-1763.
482   10.1111/1759-7714.13124

483 Xueqing H, Jun Z, Yueqiang J, Xin L, Liya H, Yuanyuan F, Yuting Z, Hao Z, Hua W, Jian L, and Tiejun Y. 2020.
484   IGF2BP3 May Contributes to Lung Tumorigenesis by Regulating the Alternative Splicing of PKM. Front
485   Bioeng Biotechnol 8:679. 10.3389/fbioe.2020.00679

486 Yang CY, Yang JC, and Yang PC. 2020. Precision Management of Advanced Non-Small Cell Lung Cancer. Annu
487   Rev Med 71:117-136. 10.1146/annurev-med-051718-013524

488 Yu G, Wang LG, Han Y, and He QY. 2012. clusterProfiler: an R package for comparing biological themes among
489   gene clusters. Omics 16:284-287. 10.1089/omi.2011.0118

490 Yuan Y, Zhou X, Kang Y, Kuang H, Peng Q, Zhang B, Liu X, and Zhang M. 2021. Circ-CCS is identified as a cancer-
491   promoting circRNA in lung cancer partly by regulating the miR-383/E2F7 axis. Life Sci 267:118955.
492   10.1016/j.lfs.2020.118955

493 Yue C, Ma H, and Zhou Y. 2019. Identification of prognostic gene signature associated with microenvironment of
494   lung adenocarcinoma. PeerJ 7:e8128. 10.7717/peerj.8128

495 Zhang J, Luo W, Chi X, Zhang L, Ren Q, Wang H, and Zhang W. 2020a. IGF2BP1 silencing inhibits proliferation
496   and induces apoptosis of high glucose-induced non-small cell lung cancer cells by regulating Netrin-1. Arch
497   Biochem Biophys 693:108581. 10.1016/j.abb.2020.108581

498 Zhang L, Zhang Z, and Yu Z. 2019a. Identification of a novel glycolysis-related gene signature for predicting
499   metastasis and survival in patients with lung adenocarcinoma. J Transl Med 17:423. 10.1186/s12967-019-
500   02173-2

501 Zhang S, Sun K, Zheng R, Zeng H, Wang S, Chen R, Wei W, and He J. 2020b. Cancer incidence and mortality in
502   China, 2015. Journal of the National Cancer Center. 10.1016/j.jncc.2020.12.001

503 Zhang Y, Wang LF, Gao JH, Li L, Jiang P, Lv X, Yu LX, Yang J, Li RT, and Liu BR. 2019b. Clinical significance of
504   epithelial-mesenchymal transition-related molecules in lung adenocarcinoma. Curr Oncol 26:e121-e127.

505   10.3747/co.26.4471

506   Zheng S, Yao L, Li F, Huang L, Yu Y, Lin Z, Li H, Xia J, Lanuti M, and Zhou H. 2021. Homologous recombination

507   repair rathway and RAD54L in early-stage lung adenocarcinoma. PeerJ 9:e10680. 10.7717/peerj.10680

508   Zhou H, Wang L, Huang J, Jiang M, Zhang X, Zhang L, Wang Y, Jiang Z, and Zhang Z. 2015. High EGFR_1 Inside-

509   Out Activated Inflammation-Induced Motility through SLC2A1-CCNB2-HMMR-KIF11-NUSAP1-PRC1-

510   UBE2C. J Cancer 6:519-524. 10.7150/jca.11404

Figure 1

The differentially expressed embryonic germline genes (DEGGs) between LUAD and normal tissues.

(A) The volcano plot of DEGGs. The red and green dots represent the up-regulated and down-regulated DEGGs, and the black dots represent the embryonic germline genes that do not meet the selection criteria. (B) The heatmap of the representative DEGGs. N and T represent the normal and LUAD samples, respectively. The color of scale bar (from green to red) indicates the expression status of DEGGs (from lowly expression to highly expression). (C) The boxplot of the representative DEGGs. P<0.05 was considered as statistical significance. LUAD: Lung adenocarcinoma; FDR: False Discovery Rate.
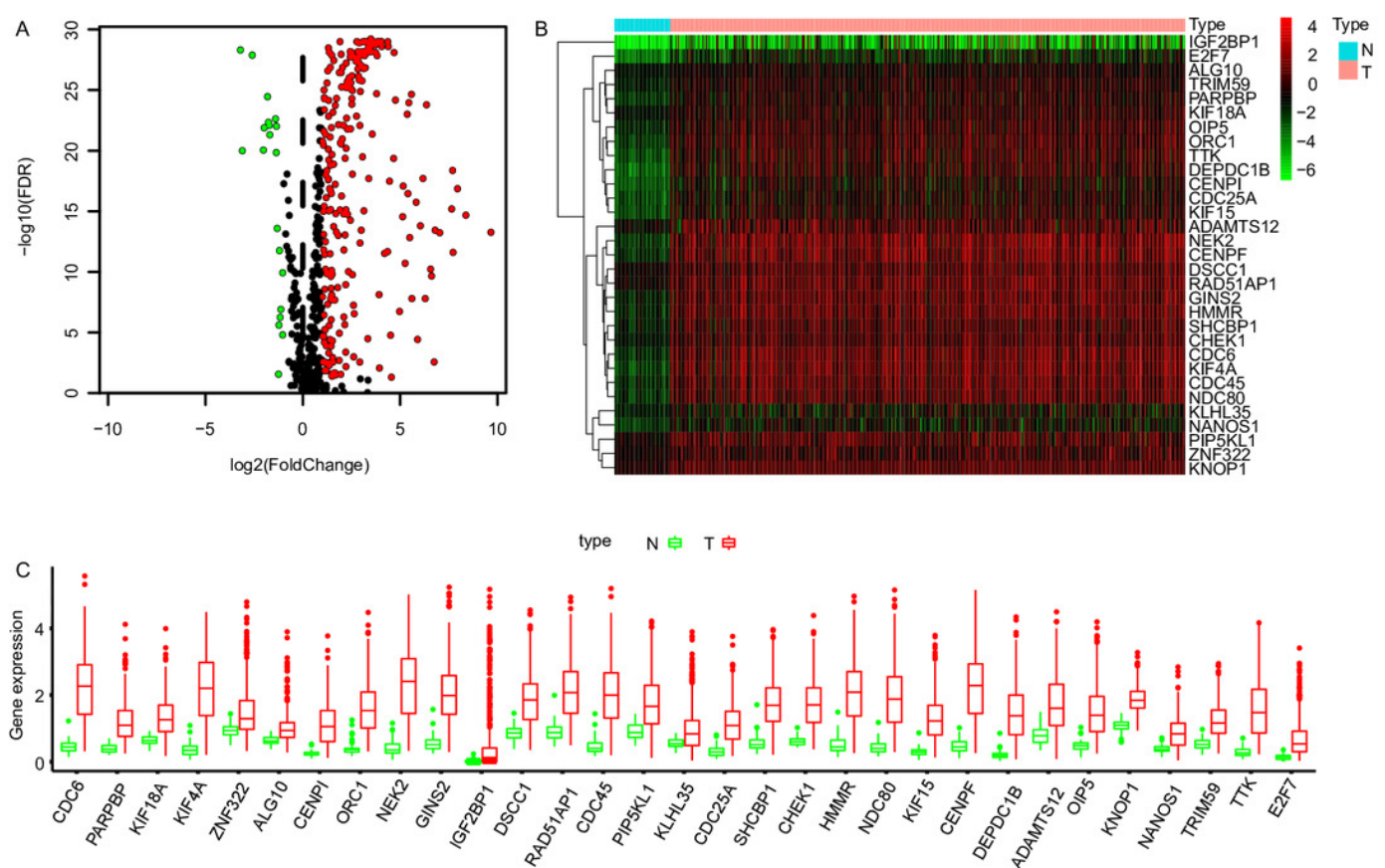
Figure 2

The GO and KEGG enrichment analysis of DEGGs.

(A, C): The bubble plot of GO and KEGG pathway analysis of DEGGs, respectively. The color scale represents large (blue) to small (red) p.adjust values. The size of the black dots represents the number of DEGGs enriched in the corresponding terms or pathways. (B, D): The heat plot of GO and KEGG pathway analysis of DEGGs , respectively. The color depth represents the expression status of DEGGs, and the length of the bar represents the number of DEGGs enriched in the corresponding terms or pathways. BP: Biological Process; CC: Cellular Component and MF: Molecular Function; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; logFC: log2FoldChange; DEGGs: differentially expressed embryonic germline genes.
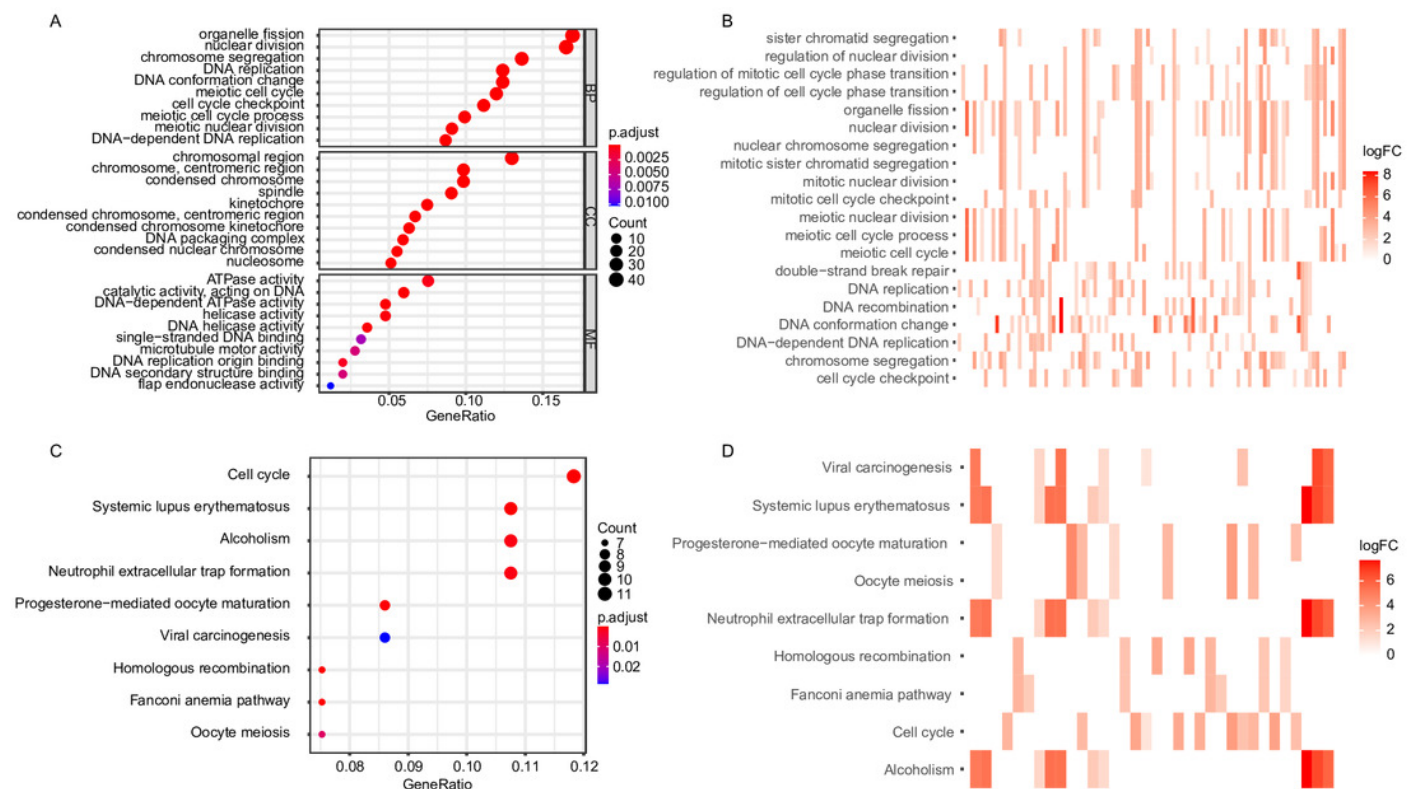
Figure 3

The identification of the DEGGs with prognostic value.

(A): The forest plot of the representative prognostic DEGGs screened out by univariate Cox proportional hazards regression. The red rectangles represent hazard factors, and green rectangles represent protective factors. (B): Lasso co-efficient profiles of prognostic DEGGs by optimal lambda. (C): The partial likelihood deviance plot presented the minimum number corresponds to the covariates used for multivariate Cox analysis. DEGGs: differentially expressed embryonic germline genes; Lasso: least absolute shrinkage and selection operator.
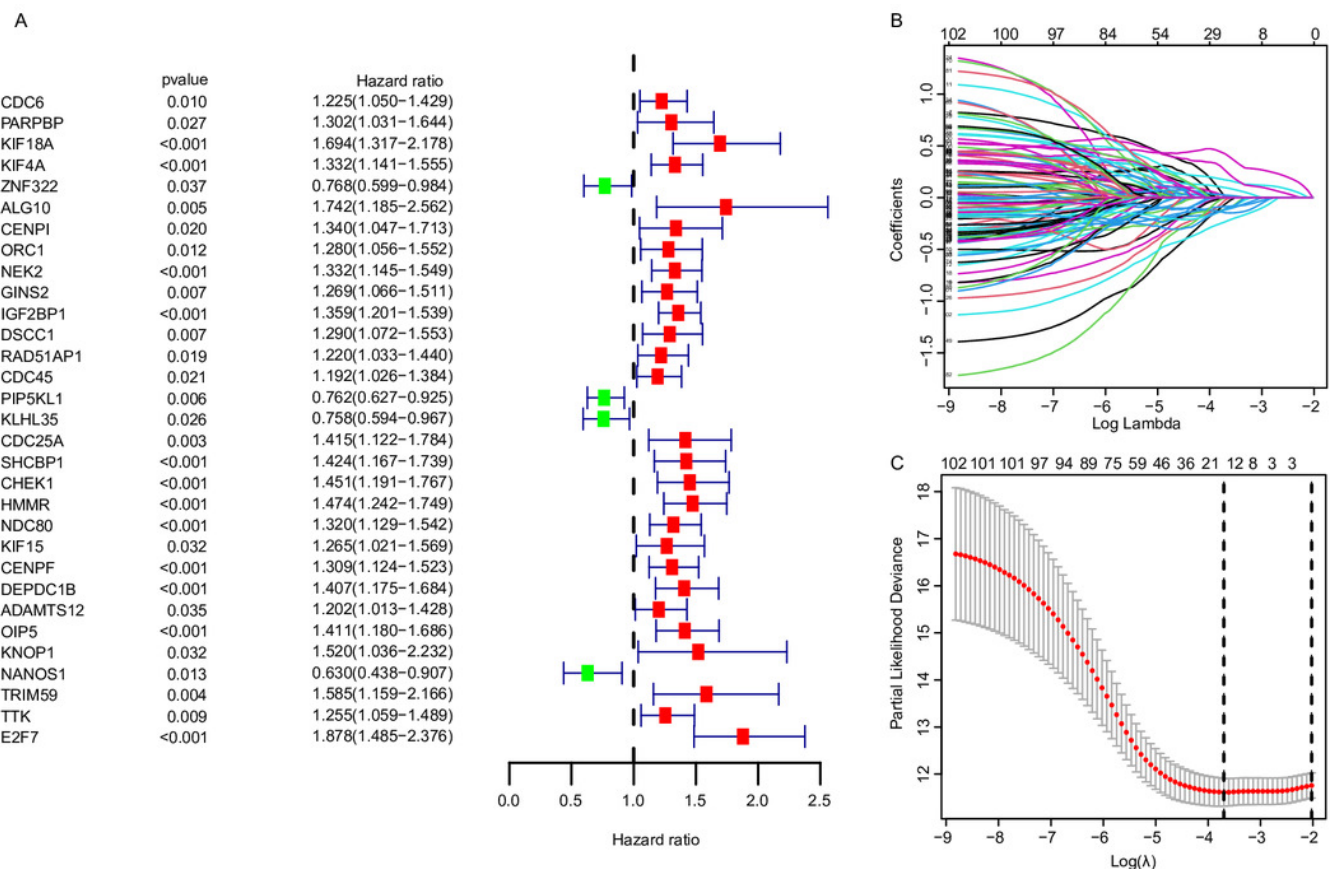
Figure 4

The construction of embryonic germline gene related prognostic model.

(A): Survival curve of low- and high-risk groups of training set stratified by EGRS. (B): The multi-ROC curve of EGRS and other clinicopathological factors for survival prediction in training set. P < 0.05 was considered as statistical significance. (C): The expression status of 11 risk DEGGs in the high-risk (cyan) and low-risk (light red) groups of LUAD patients in training set. (D): The risk score distribution of low-risk (green) and high-risk (red) groups of LUAD patients in training set. (E): The scatter plot of survival status of LUAD patients in training set. The red dots represent the dead patients, and the green dots represent the alive patients. LUAD: Lung adenocarcinoma; DEGGs: differentially expressed embryonic germline genes; EGRS: embryonic germline gene related signature ; ROC: Receiver Operating Characteristic; AUC: Area Under Curve.
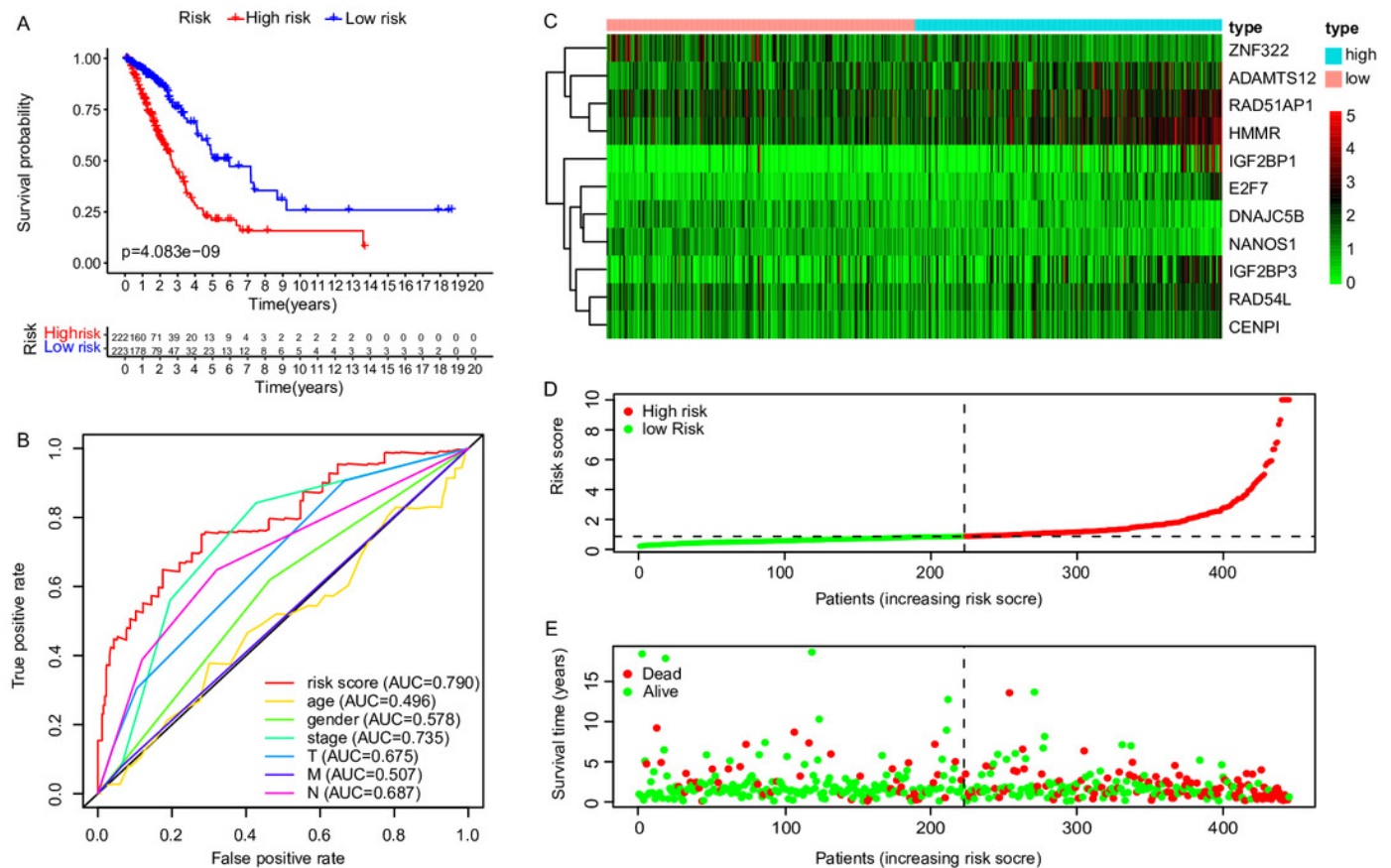
Figure 5

The distributions of ERGS genes' expressions or risk score to pathological status in training set.

(A-C): age, (D-F): gender, (G, H): T, (I-M): N, (N, O): M and (P-V): stage. EGRS: embryonic germline gene related signature.
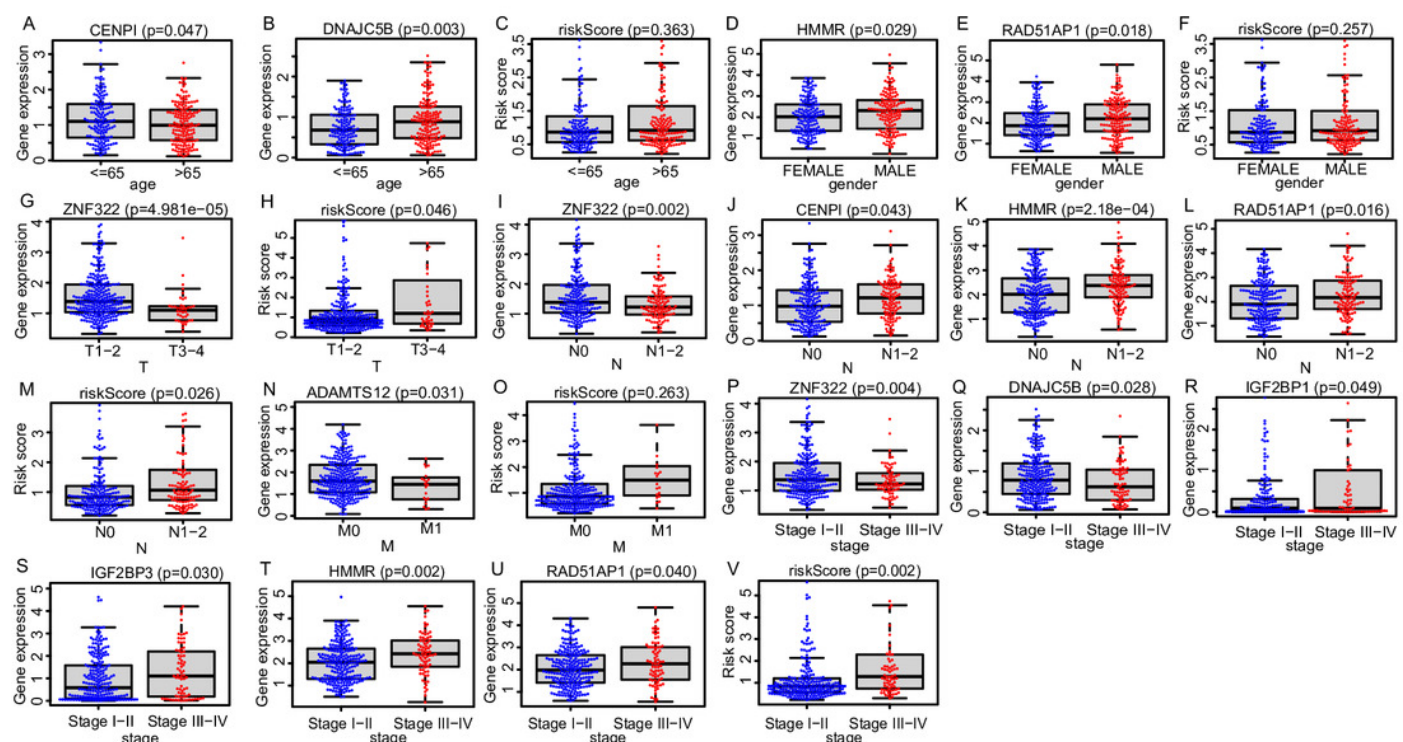
Figure 6

The validation of embryonic germline gene related prognostic model.

(A): The prognostic effect analyses of EGRS and commonly used prognostic factors using univariate Cox regression model. (B): The independent prognostic effect analyses of EGRS and commonly used prognostic factors using multivariate Cox regression model. Kaplan-Meier survival curves showing overall survival outcomes of GSE30219 (C), GSE50081 (D), GSE37745 (E) and GSE72094 (F) according to high-risk and low-risk patients stratified by EGRS. The ROC analysis of GSE30219 (G), GSE50081 (H), GSE37745 (I) and GSE72094 (J) for survival prediction by the EGRS. ROC: Receiver Operating Characteristic; AUC: Area Under Curve.
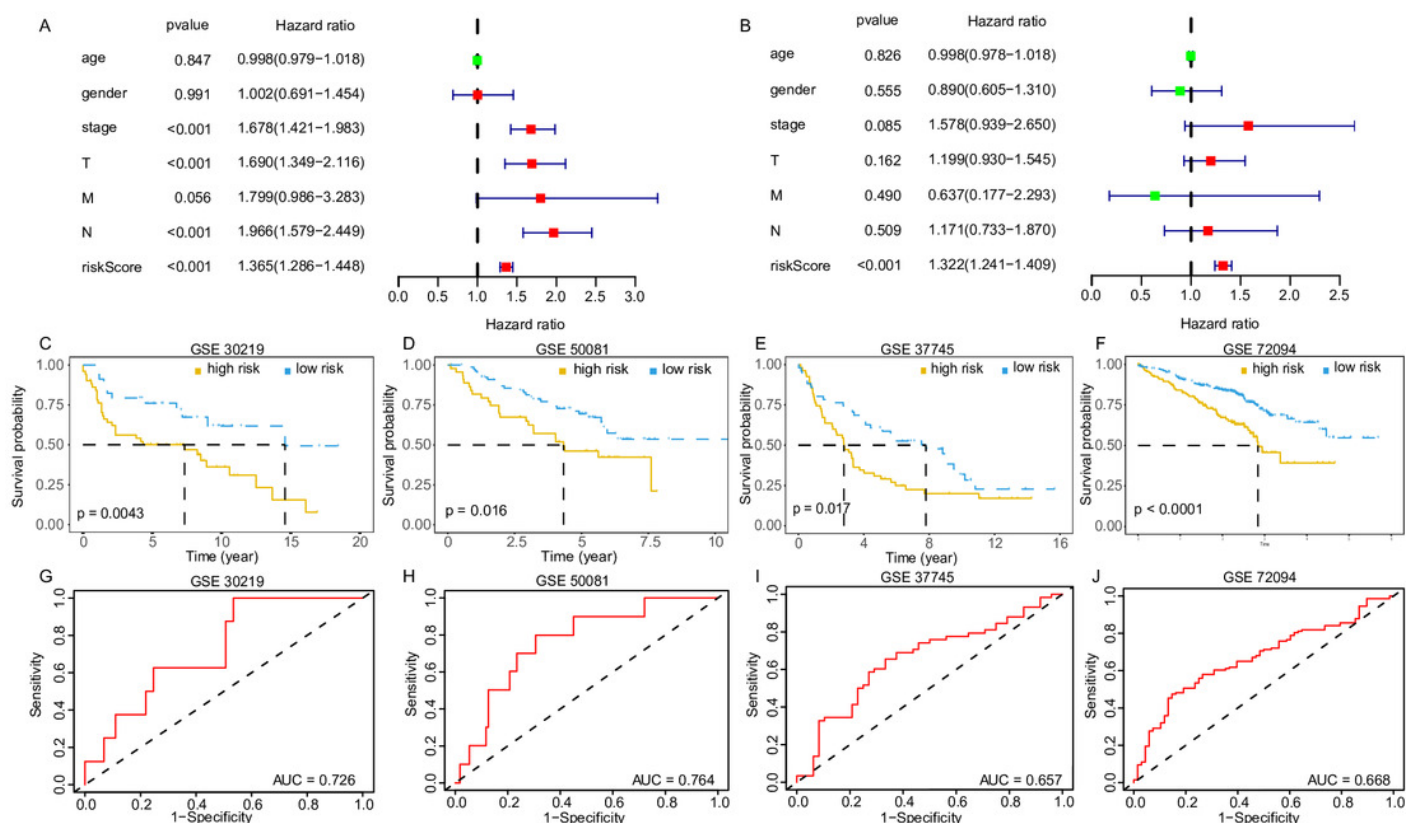
Figure 7

Prognostic nomogram and calibration plots of EGRS.

(A): The nomogram of EGRS. (B, C and D): calibration plots of nomogram predicting 1-year, 3-year and 5-year survival, respectively.