



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

Linjun Liu¹, Ke Xu² and Yubai Zhou¹

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

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

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 sets, 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 four 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 EGRS has a significant correlation with the progression of LUAD clinicopathology, suggesting these genes might play an important role in the progression

Submitted 4 June 2021
Accepted 15 September 2021
Published 21 October 2021

Corresponding authors
Ke Xu, xuke@ivdc.chinacdc.cn
Yubai Zhou, zhouyubai@bjut.edu.cn

Academic editor
Fabrizio Bianchi

Additional Information and
Declarations can be found on
page 13

DOI 10.7717/peerj.12257

© Copyright
2021 Liu et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

of LUAD. Finally, based on EGRS genes, we built and calibrated a nomogram for conveniently evaluating patients' outcomes.

Subjects Bioinformatics, Oncology, Respiratory Medicine, Medical Genetics

Keywords Lung adenocarcinoma, Embryonic germline genes, Prognosis, Biomarkers, Bioinformatics analysis

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, Morgensztern & Boshoff, 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, Yang & Yang, 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*).

Many cancers activate genes normally associated with different developmental states, including germ cell-specific genes, which suggests that some unknown mechanisms which are normally limited to the development of germ cells maybe involved in the process of tumorigenesis (*Bruggeman et al., 2018; Erenpreisa & Cragg, 2010; Kho et al., 2004; Lepourcelet et al., 2005; McFarlane & Wakeman, 2017; Simpson et al., 2005*). With the first discovery of MAGE-1 in melanoma cells, a series of genes that are normally restricted to the testis, such as BAGE, GAGE, HOM-MEL-40, SYCP1, and NY-ESO-1 have been found in various tumors, and these genes are collectively called cancer/testis (CT) genes (*Almeida et al., 2009; Boël et al., 1995; Chen et al., 1997; De Backer et al., 1999; Türeci et al., 1996; Türeci et al., 1998; van der Bruggen et al., 1991*). Although the driving role of these CT genes in tumorigenesis and development has not been fully studied, there is evidence that they may play a role in the early stage of oncogenesis and the maintenance of tumor phenotypes (*Bruggeman et al., 2018*). However, these previously identified CT genes are mixed with genes derived from testicular somatic cells and are not truly germline-specific genes. Currently, using bioinformatic methods, *Bruggeman et al. (2020)* identified 672 so called embryonic germ cell genes which are restricted to the germline and also expressed in a variety of tumor tissues, including lung cancer. However, the level of expression of these genes in lung adenocarcinoma and its relationship with patient prognostics have not been fully studied.

In this study, we systematically analyzed the expression levels of embryonic germline cell-specific genes in LUAD cohort and constructed a novel prognostic signature and nomogram, which can be used to predict the overall survival of LUAD patients.

MATERIAL AND METHODS

Data source and preprocessing

Data were collected and preprocessed as previously described (Yue, Ma & Zhou, 2019). Briefly, as the training dataset, the expression profile and clinical information of LUAD cohort were downloaded from TCGA database using GDC tool and summarized into an expression and a clinical matrix, respectively. The ensemble ids in the expression matrix were converted into gene symbols according to the annotation file and the multiple expression data of a single gene were replaced by an average value. Genes that are not expressed in all samples are also removed from the expression matrix. For testing sets, we used lung adenocarcinoma as a keyword to search in the GEO database. After removing the datasets which are devoid of clinical data or expression data of genes in the model, four independent GEO datasets GSE30219, GSE50081, GSE37745 and GSE72094 were selected and downloaded via GEOquery package (Davis & Meltzer, 2007). For pan-cancer analyses, the expression profiles and clinical data of 33 TCGA cancer types were downloaded from TCGA database, and preprocessed according to previously mentioned procedures. Tumor cohorts with normal samples <5 were not included in the pan-cancer expression analysis. The gene expression values were all converted to transcripts per million transcripts (TPM) to facilitate subsequent analysis. The protein expression data of LUAD were obtained from the Clinical Proteomic Tumor 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, Ma & Zhou, 2019). The multiple comparisons were controlled using the false discovery rate (FDR) (Benjamini & Hochberg, 1995). The screening criteria were $|\log_2\text{Fold Chang}| > 1$ and $\text{FDR} < 0.05$.

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, Ma & Zhou, 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, Ma & Zhou, 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, Hastie & Tibshirani, 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 DEGs 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 DEGs in EGRS between normal and tumor samples were analyzed using proteomic data of LUAD cohort obtained from CPTAC database. Then, the predicted effect 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 DEGs 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 DEGs 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 used to assess the generalization power of EGRS model in four independent GEO datasets. Finally, a pan-cancer expression and prognostic analyses of 11 DEGs were performed to extend our findings to other types of cancers.

Construction and evaluation of a nomogram

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

RESULT

Identification the DEGs in LUAD

The transcriptome data of LUAD cohort which contains 497 tumor and 54 normal samples were downloaded from TCGA database, and the corresponding expression profiles of embryonic germline genes were extracted according to the data published previously (*Bruggeman et al., 2020*). The DEGs between normal and tumor samples were identified by limma package in R software (*Ritchie et al., 2015*). Totally, 269 DEGs were screened out, among which, the upregulated DEGs were 249, and the down-regulated DEGs were 20 (*Fig. 1A, Table S1*). The representative DEGs are shown by heat map (*Fig. 1B*) and box plot (*Fig. 1C*), respectively.

Functional enrichment and KEGG analysis of DEGs

The GO enrichment analysis of the DEGs revealed that the DEGs were mainly enriched in biological process of cell proliferation and DNA damage repair, such as nuclear division

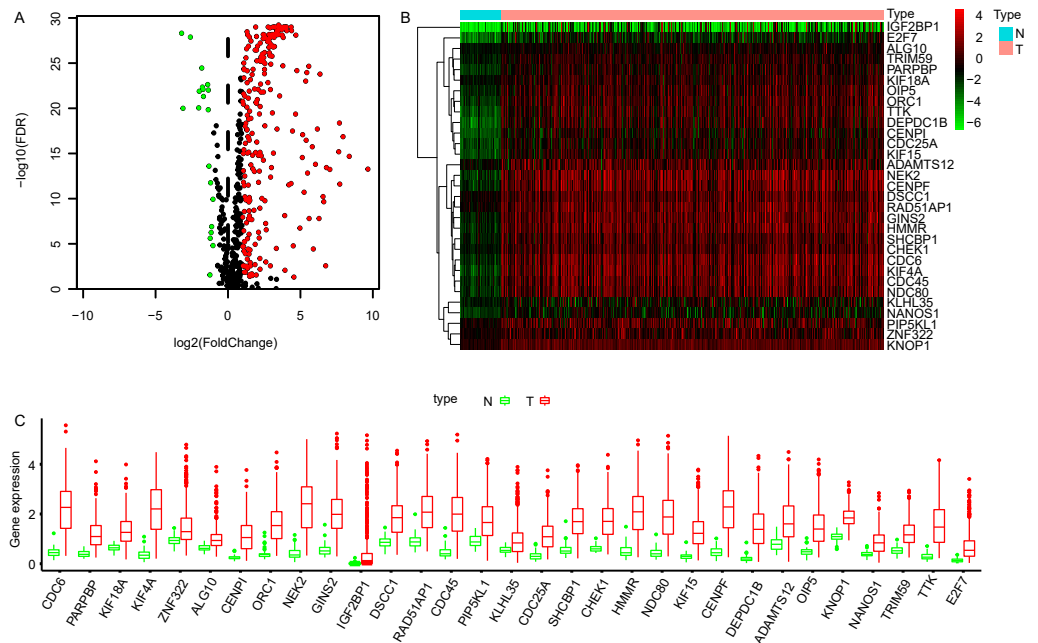


Figure 1 The differentially expressed embryonic germline genes (DEGGs) between LUAD and normal tissues. (A) The volcano plot of DEGGs. The red and green dots represent the up-regulated and down-regulated DEGGs, and the black dots represent the embryonic germline genes that do not meet the selection criteria. (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.

Full-size [DOI: 10.7717/peerj.12257/fig-1](https://doi.org/10.7717/peerj.12257/fig-1)

(GO:0000280), DNA replication (GO:0006260), cell cycle checkpoint (GO:0000075), double-strand break repair (GO:0006302) and interstrand cross-link repair (GO:0036297) (Figs. 2A, 2B). The KEGG analysis showed that the enriched pathways were cell cycle (hsa04110), homologous recombination (hsa03440), and Fanconi anemia pathway (hsa03460) (Figs. 2C, 2D). The detailed GO terms of the three categories, biological process (BP), cellular component (CC), molecular function (MF) and KEGG results were presented in Tables S2 and S3, respectively.

The construction of embryonic germline gene related prognostic model

For EGRS construction, the univariate Cox regression analysis was conducted and the resulting 103 DEGGs with prognostic value ($p < 0.05$, Fig. 3A) were further screened by LASSO regression. Then, the sixteen key DEGGs were selected to performed the multivariate Cox regression analysis (Figs. 3B, 3C). Finally, a prognostic signature containing eleven DEGGs (RAD54L, ZNF322, CENPI, IGF2BP1, IGF2BP3, RAD51AP1, E2F7, HMMR, DNAJC5B, ADAMTS12, NANOS1) was constructed and the risk score formula was presented as follow:

$$\text{Risk score} = (-0.4294) * \text{RAD54L} - 0.2411 * \text{ZNF322} - 0.4645 * \text{CENPI} + 0.2243 *$$

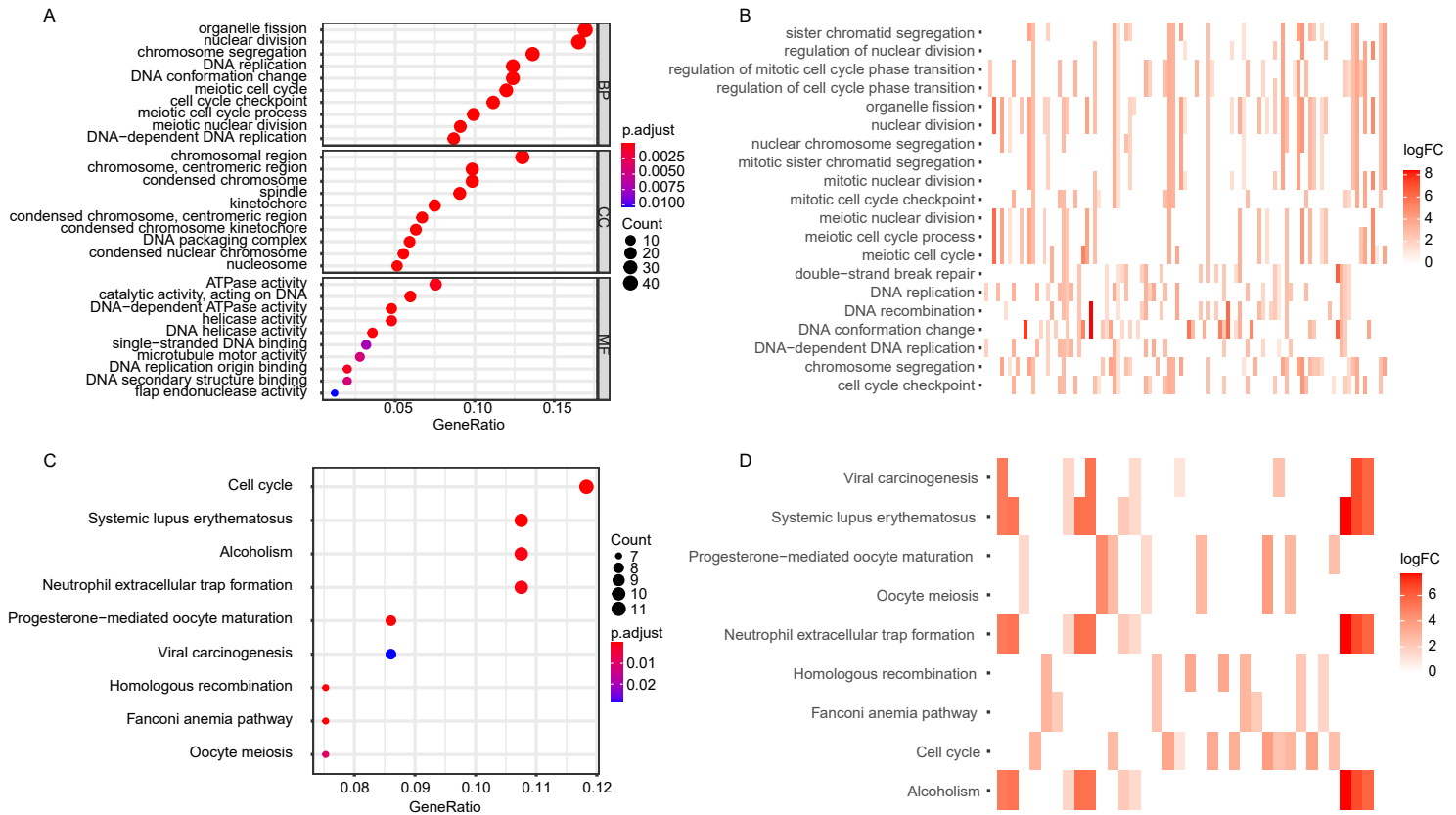


Figure 2 The GO and KEGG enrichment analysis of DEGGs. (A, C) The bubble plot of GO and KEGG pathway analysis of DEGGs, respectively. The color scale represents large (blue) to small (red) p.adjust values. The size of the black dots represents the number of DEGGs enriched in the corresponding terms or 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.

Full-size DOI: 10.7717/peerj.12257/fig-2

$$\text{IGF2BP1} + 0.2095 * \text{IGF2BP3} - 0.5093 * \text{RAD51AP1} + 0.7979 * \text{E2F7} + 0.7968 * \\ \text{HMMR} - 0.3272 * \text{DNAJC5B} + 0.1997 * \text{ADAMTS12} - 0.5248 * \text{NANOS1}.$$

The validation of the EGRS for survival prediction

Due to the potential inconsistencies in gene expression between the transcriptome and the proteome, we first verified the protein expression of the 11 DEGGs in EGRS in normal and tumor samples through the CPTAC database. Given the limitations of available data, only 4 genes of the 11 DEGGs were analyzed. Consistent with the transcriptome data, the protein expression of 4 genes (IGF2BP1, IGF2BP3, HMMR and ADAMTS12) were up-regulated in LUAD tumor samples. The log₂FoldChanges of IGF2BP1, IGF2BP3, HMMR and ADAMTS12 were 0.49, 1.14, 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, respectively. In addition, the pan-cancer expression analysis also revealed that 11 DEGGs in EGRS showed differential expression in most types

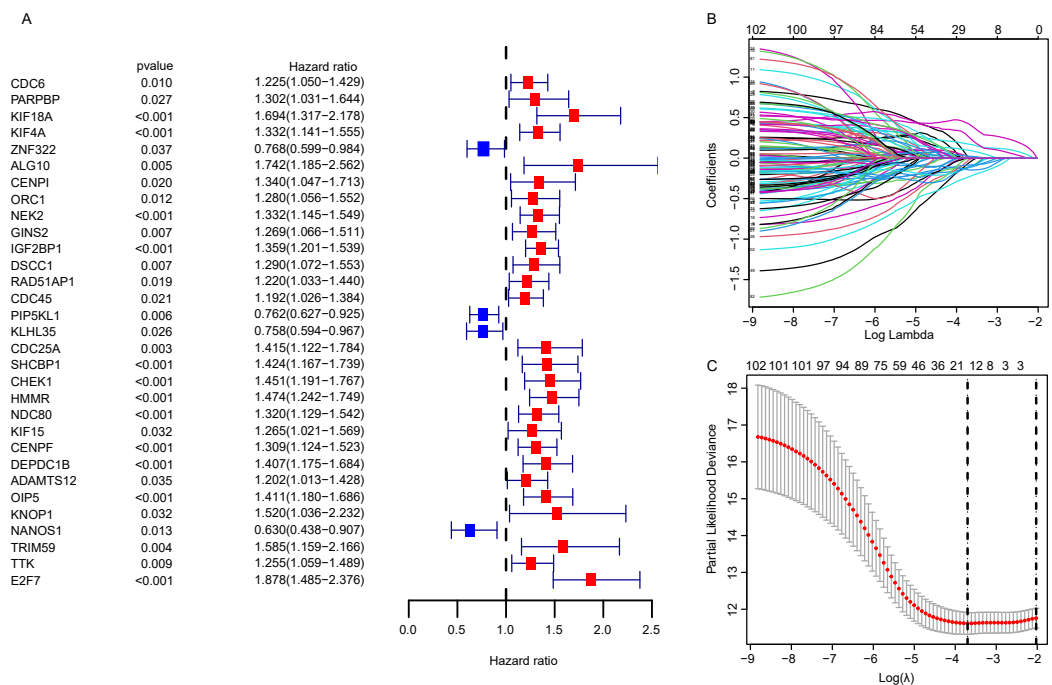


Figure 3 The identification of the DEGGs with prognostic value. (A) The forest plot of the representative prognostic DEGGs screened out by univariate Cox proportional hazards regression. The red rectangles represent hazard factors, and blue rectangles represent protective factors. (B) Lasso co-efficient 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.

Full-size [DOI: 10.7717/peerj.12257/fig-3](https://doi.org/10.7717/peerj.12257/fig-3)

of tumors, but diverse expression trends between tumor and paired adjacent normal tissues of some types of cancers were observed, suggesting that the DEGGs may play different roles in the development of different types of tumors (Fig. S1). Then, Kaplan–Meier (KM) survival analysis indicated that the overall survival between the high- and low-risk groups in training set was significantly different, and the low-risk patients had significantly better outcomes than that of high-risk patients ($P = 4.083e-09$) (Fig. 4A). Besides, our model had the best predictive power, compared with other clinicopathological factors such as age, gender, TNM and stage. Its area under curve (AUC) of ROC was 0.790 (Fig. 4B). The DEGGs in EGRS indicated differential expression between high- and low-risk samples (Fig. 4C), and with the increase of risk score, the proportion of death cases was also increasing (Figs. 4D, 4E). Between different pathological groups, the expression levels of genes in EGRS also showed significant difference (Fig. 5). Contrary to CENPI, the expressions of DNAJC5B in elderly patients (>65 years old) were higher than those in younger patients (≤ 65 years old) (Figs. 5A, 5B). The HMMR, RAD51AP1 were highly expressed in men than in women (Figs. 5D, 5E). ZNF322 was related to T stage, and its expressions in T3-4 were significantly lower than those in T1-2 (Fig. 5G). The CENPI, DNAJC5B, HMMR, RAD51AP1 and ZNF322 were correlated to N stage. Except for ZNF322, other genes were highly expressed in N1-2 stage (Figs. 5I–5L). The expressions of ADAMTS12 in M1 stage

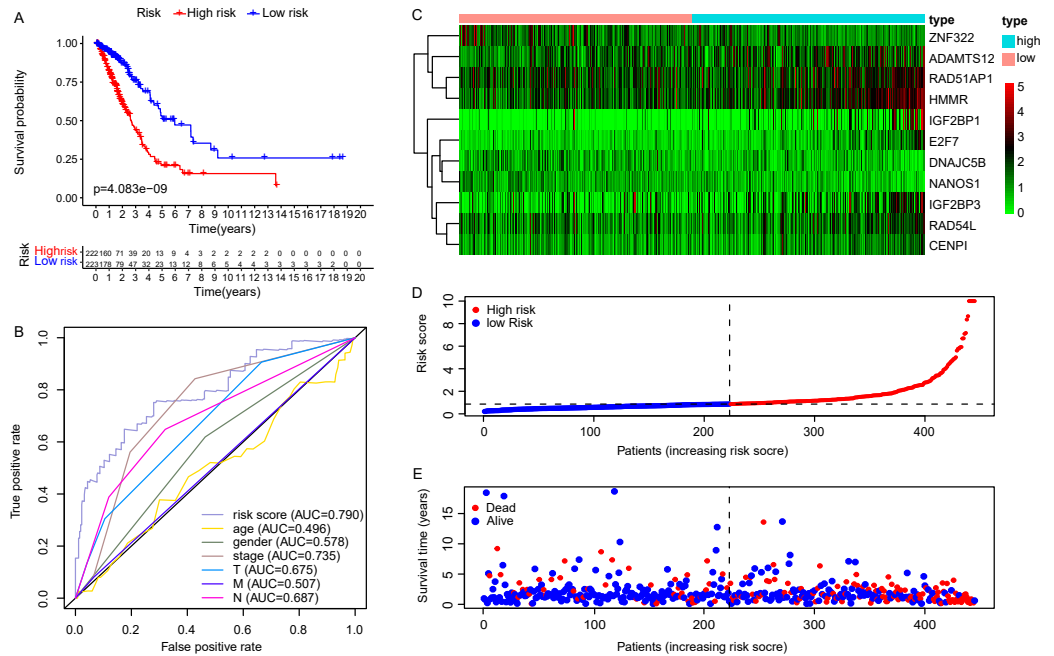


Figure 4 The construction of embryonic germline gene related prognostic model. (A) Survival curve of low- and high-risk groups of training set stratified by EGRS. (B) The multi-ROC curve of EGRS and other clinicopathological factors for survival prediction in training set. $P < 0.05$ was considered as 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 (blue) 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 blue 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.

Full-size [DOI: 10.7717/peerj.12257/fig-4](https://doi.org/10.7717/peerj.12257/fig-4)

were lower than those in M0 stage (Fig. 5N). The expressions of DNAJC5B, HMMR, IGF2BP1, IGF2BP3, RAD51AP1, ZNF322 were correlated to the pathological stage. Except for ZNF322 and DNAJC5B, the expressions of other genes in stage III-IV were significantly higher than those in stage I-II (Figs. 5P–5U). More importantly, the risk scores in the advanced LUAD were significantly higher than those in the early stage (Fig. 5H, 5M and 5V), although its distribution had no statistical difference in some pathological statuses such as age, gender and tumor metastasis (Figs. 5C, 5F and 5O).

The univariate and multivariate Cox analyses suggested the risk score was an independent prognostic factor (Figs. 6A, 6B). The pan-cancer prognostic analysis also indicated the 11 DEGGs had prognostic value in some cancer types. Likewise, different prognostic effects based on cancer type were observed in some DEGGs (the detail results were packaged into a Supplementary file). In addition, the predictive power of the EGRS was further validated in four GEO testing datasets. Consistent with the training set, our model also stratified the samples in the testing sets into low-risk and high-risk groups, and the low-risk patients had better outcomes (Figs. 6C–6F). The AUC results also showed that our model had stable

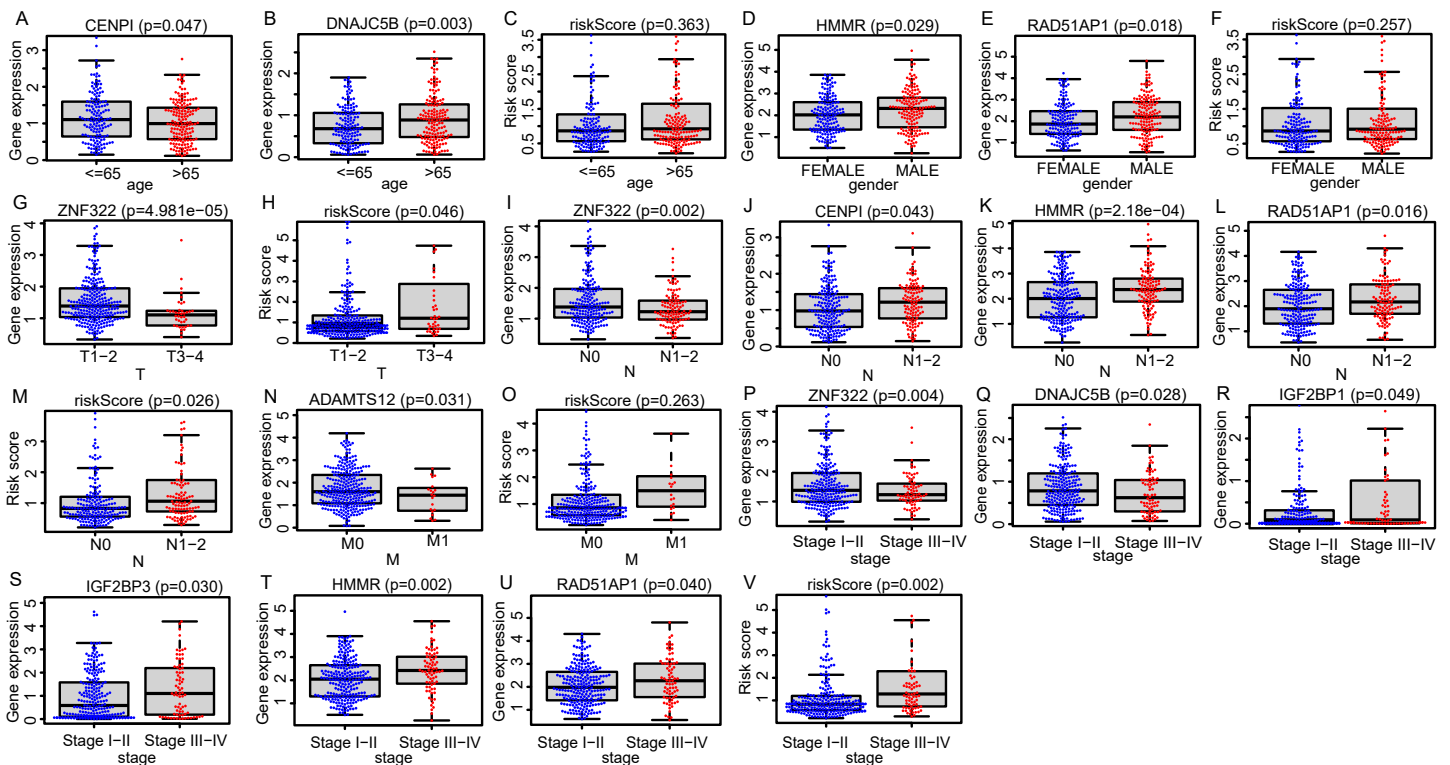


Figure 5 The distributions of EGRS genes' expressions or risk score to pathological status in training set. (A–C) Age, (D–F) gender, (G, H) T, (I–M) N, (N, O) M and (P–V) stage. EGRS, embryonic germline gene related signature.

Full-size [DOI: 10.7717/peerj.12257/fig-5](https://doi.org/10.7717/peerj.12257/fig-5)

predictive ability. The AUC in four independent GEO datasets were 0.726 (GSE30219), 0.764 (GSE50081), 0.657 (GSE37745) and 0.668 (GSE72094) (Figs. 6G–6J).

Construction and evaluation of a nomogram

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

DISCUSSION

Although the data published recently on lung cancer in the United States are encouraging, lung cancer is still a malignant tumor with high morbidity and mortality in the world (Siegel, Miller & Jemal, 2020; Sung et al., 2021). In China, the incidence of lung cancer ranks the first and the second among men and women, respectively, and it is still the malignant disease with the highest mortality rate (Zhang et al., 2020b). The lung adenocarcinoma is the most common pathological subtype of non-small cells lung cancer (NSCLC), which accounts for

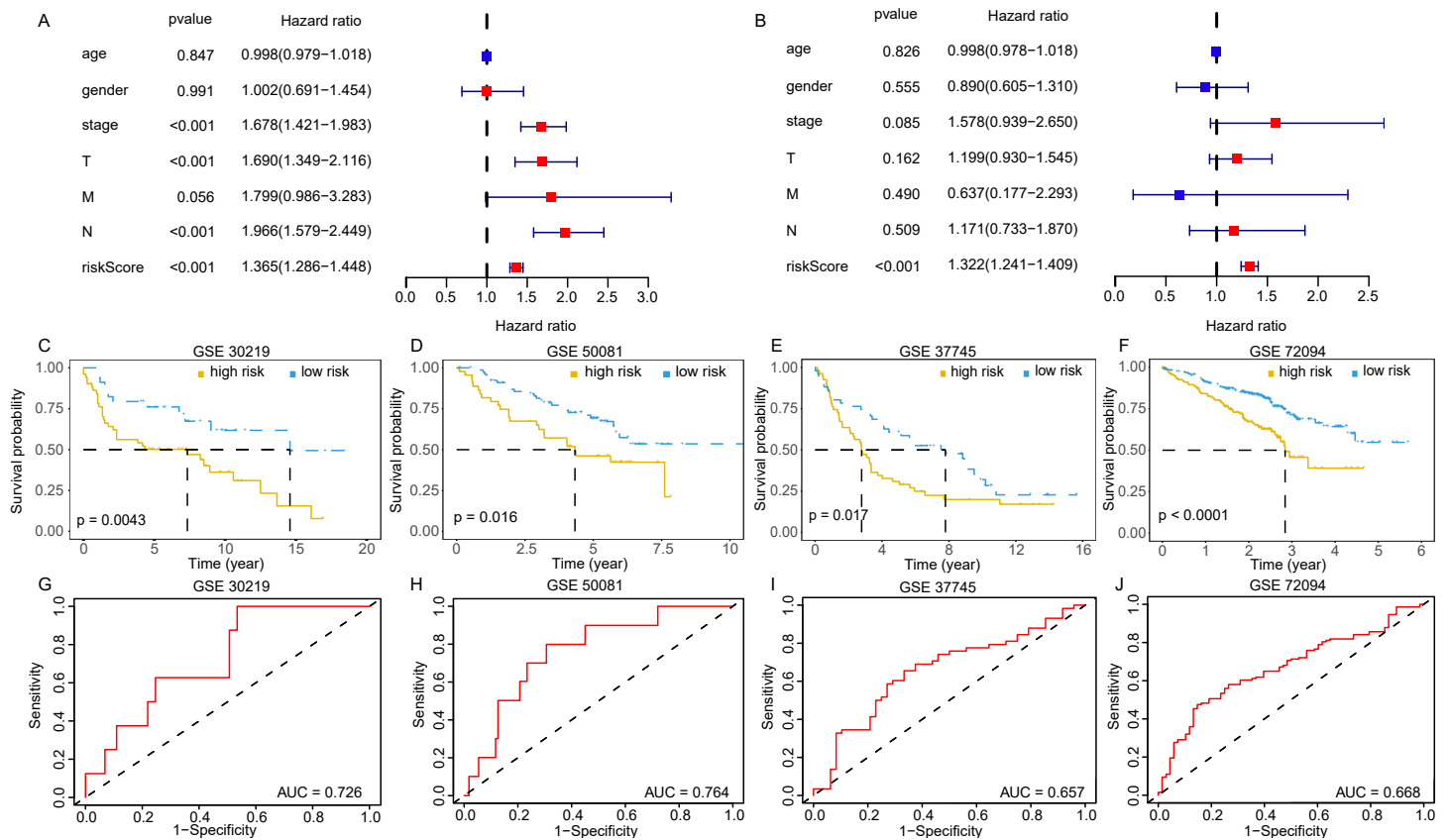


Figure 6 The validation of embryonic germline gene related prognostic model. (A) The prognostic effect analyses of EGRS and commonly used prognostic factors using univariate Cox regression model. (B) The independent prognostic effect analyses of EGRS and commonly used prognostic factors using multivariate 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. EGRS, embryonic germline gene related signature; ROC, Receiver Operating Characteristic; AUC, Area Under Curve.

Full-size DOI: 10.7717/peerj.12257/fig-6

about 40% of all lung cancer cases (Kleczo *et al.*, 2019). Benefiting from carrying a higher proportion of actionable mutations, lung adenocarcinoma has more treatment options than other subtypes of lung cancer, however, its five-year overall survival rate remains unsatisfactory (Zhang *et al.*, 2019b). Accumulating evidence indicates that patients with early stage of disease have a better prognosis, so finding more effective early diagnostic biomarkers or therapeutic targets is important for developing more effective treatment regimens, prolonging patient's survival time and improving their quality of life.

The various types of human cells are derived from a single fertilized egg cell through differentiation mechanisms that we do not fully understand (McKenna & Gagnon, 2019). In this process, specific types of genes in differentiating cells are expressed in a well-organized spatiotemporal sequence under the cues of various internal and external factors, and finally generate tissue cells with distinct morphology, function and gene expression profiles (Trott & Martinez Arias, 2013). Germline cell-specific genes are a class of genes that are expressed only in germ cells. Accumulating evidence shows that tumor cells and germ cells have

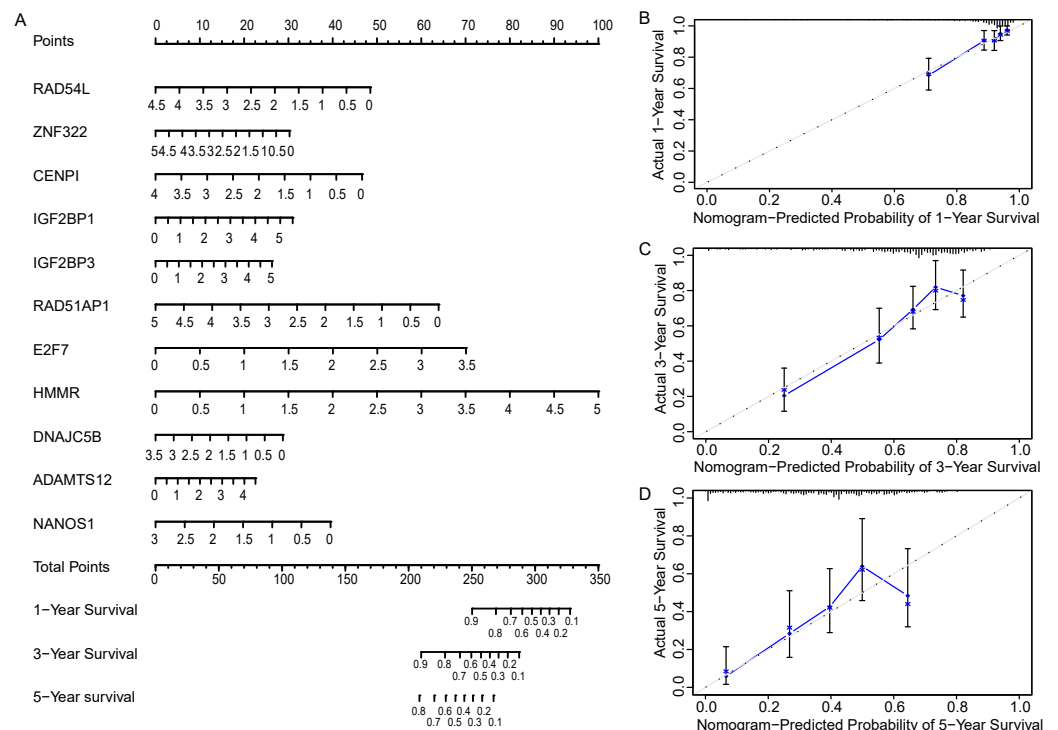


Figure 7 Prognostic nomogram and calibration plots of EGRS. (A) The nomogram of EGRS. (B, C and D) calibration plots of nomogram predicting 1-year, 3-year and 5-year survival, respectively.

Full-size DOI: [10.7717/peerj.12257/fig-7](https://doi.org/10.7717/peerj.12257/fig-7)

similar behaviors, so it is speculated that tumor cells may adopt certain similar regulatory mechanisms in the process of tumorigenesis and development (*Bruggeman et al., 2018*; *Simpson et al., 2005*). Indeed, follow-up studies have detected multiple germ cell-specific genes in different tumor tissues (*Erenpreisa & Cragg, 2010*; *McFarlane & Wakeman, 2017*; *Whitehurst, 2014*). The products of these genes are regarded as ideal immunotherapy targets due to their restricted range of expression and have attracted more and more attention.

Currently, *Bruggeman et al. (2020)* identified 672 true embryonic germ cell-specific genes by removing contaminated somatic genes. However, the expression status and their prognostic value of these genes have not been fully studied in lung adenocarcinoma. To take these issues, we systematically studied the expression levels of the above-mentioned 672 genes in the TCGA LUAD cohort, evaluated their correlation with the prognosis of LUAD patients, and finally constructed and validated a prognostic model consisting of 11 DEGGs. Our model can well stratify the patients into high- and low-risk groups and robustly predict the outcomes of patients in both training and testing sets. Among the 11 DEGGs, ZNF322 (also known as ZNF322A, ZNF388 or ZNF489) is a member of the zinc finger transcription factor family, and may act as a positive regulator in MAPK signaling pathway (*Li et al., 2004*). Recent studies have shown that the overexpression of ZNF322 is related to the oncogenesis of many tumors including lung cancer through a variety of mechanisms, such as directly suppressing the expression of c-Myc to promote

cell stemness or up-regulating the transcription of α -adducin (ADD1) and cyclin D1 (CCND1) to promote tumor growth and metastasis (Jen et al., 2016; Jen et al., 2019; Liao et al., 2017; Williams et al., 2014). DNA binding protein RAD51AP1 is a component of DNA damage repair system, which plays an important role in RAD51-mediated homologous recombination (Pires, Sung & Wiese, 2017). Further studies indicates that overexpression of RAD51AP1 promotes the proliferation of lung cancer cells and correlates to poor prognosis (Chudasama et al., 2018; Li et al., 2018b; Wu et al., 2019). HMMR coding a 724 amino acids protein, which may associate with the motility of cancer cells (Sankaran et al., 2012). Quite a few of studies had shown that the highly expression of HMMR is closely related to the prognosis of LUAD patients, and it may promote the progression of lung adenocarcinoma by regulating the metabolic state of tumor cells (He & Zuo, 2019; Jiao et al., 2020; Li et al., 2020a; Li, Qi & Li, 2020c; Liu et al., 2019; Man et al., 2014; Shen et al., 2019; Stevens et al., 2017; Zhang, Zhang & Yu, 2019a; Zhou et al., 2015). The ADAMTS12 is a novel anti-tumor metalloprotease, and the expression is epigenetically silenced in tumor cells (Li et al., 2018a; Moncada-Pazos et al., 2009). A bulk of studies suggest the overexpression of IGF2BP1 and IGF2BP3 facilitate the progress of lung cancer, and relate to the prognosis of LUAD patients (Beljan Perak et al., 2012; Gong et al., 2016; Guo et al., 2021; Huang et al., 2019; Kato et al., 2007; Li & Zhan, 2020; Li et al., 2020b; Mizutani et al., 2016; Müller et al., 2019; Ohdaira et al., 2012; Wang, He & Ma, 2020; Xueqing et al., 2020; Zhang et al., 2020a). E2F7 plays an important physiological role in embryonic development and cell cycle regulation (Park et al., 2015). Recent researches show the various non-coding RNA can influence the progress of lung cancer by regulating the expression of E2F7 (Liang et al., 2018; Liu et al., 2020; Wang et al., 2021; Yuan et al., 2021). NANOS1 was up-regulated in lung cancer, and promoted the progress of tumor progress (Bonnomet et al., 2008; De Keuckelaere et al., 2018). RAD54L is also a potential prognostic biomarker of NSCLC (Tu et al., 2021; Zheng et al., 2021). Literature retrieval showed that the relation of LUAD and DEGGs DNAJC5B, CENPI has not been reported so far. In addition, pan-cancer expression and prognosis analyses also indicated that 11 DEGGs showed differential expression and prognostic value in a variety of tumors, suggesting that these DEGGs may also play a role in the development of tumors other than lung adenocarcinoma, which needed to be validated by further functional experiments.

Although we constructed and validated an 11-gene prognostic model, the present study still has several limitations: first, the number of samples in training set is not large enough; secondly, due to tumor heterogeneity and individual differences among LUAD patients, the generalizing abilities of the model need further improvement. Therefore, further model optimization and validation using LUAD cohort with larger samples are needed. In addition, the results of bioinformatics analysis did not carry out related experimental verifications. Therefore, whether the genes in the model play a functional role in the progress of LUAD remains to be resolved. These issues mentioned above will be the main content of our follow-up studies.

CONCLUSION

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

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Major National Research and Development Projects (2012ZX10001005-007, 2018ZX10731101-002-007, 2018ZX10731101-001-017); the National High Technology Research Program of China (2012AA02A404); and the State Key Laboratory for Infectious Disease Prevention and Control (2011SKLID103). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Major National Research and Development Projects: 2012ZX10001005-007, 2018ZX10731101-002-007, 2018ZX10731101-001-017.

National High Technology Research Program of China: 2012AA02A404.

State Key Laboratory for Infectious Disease Prevention and Control: 2011SKLID103.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Linjun Liu performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Ke Xu conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Yubai Zhou conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data are available at TCGA and GEO.

TCGA: TCGA-LUAD, TCGA-LUSC, TCGA-BRCA, TCGA-GBM, TCGA-OV, TCGA-UCEC, TCGA-KIRC, TCGA-HNSC, TCGA-LGG, TCGA-THCA, TCGA-PRAD, TCGA-SKCM, TCGA-COAD, TCGA-STAD, TCGA-BLCA, TCGA-LIHC, TCGA-CESC, TCGA-KIRP, TCGA-SARC, TCGA-LAML, TCGA-PAAD, TCGA-ESCA, TCGA-PCPG, TCGA-READ, TCGA-TGCT, TCGA-THYM, TCGA-KICH, TCGA-ACC, TCGA-MESO, TCGA-UVM, TCGA-DLBC, TCGA-UCS and TCGA-CHOL.

GEO: [GSE30219](#), [GSE50081](#), [GSE37745](#) and [GSE72094](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.12257#supplemental-information>.

REFERENCES

- Almeida LG, Sakabe NJ, De Oliveira AR, Silva MC, Mundstein AS, Cohen T, Chen YT, Chua R, Gurung S, Gnjjatic S, Jungbluth AA, Caballero OL, Bairoch A, Kiesler E, White SL, Simpson AJ, Old LJ, Camargo AA, Vasconcelos AT. 2009. CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Research* 37:D816–D819 DOI 10.1093/nar/gkn673.
- Beljan Perak R, Durdov MG, Capkun V, Ivcevic V, Pavlovic A, Soljic V, Peric M. 2012. IMP3 can predict aggressive behaviour of lung adenocarcinoma. *Diagnostic Pathology* 7:165 DOI 10.1186/1746-1596-7-165.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B* 57:289–300.
- Boël P, Wildmann C, Sensi ML, Brasseur R, Renauld JC, Coulie P, Boon T, Van der Bruggen P. 1995. BAGE: a new gene encoding an antigen recognized on human melanomas by cytolytic T lymphocytes. *Immunity* 2:167–175 DOI 10.1016/s1074-7613(95)80053-0.
- Bonnomet A, Polette M, Strumane K, Gilles C, Dalstein V, Kileztky C, Berx G, Van Roy F, Birembaut P, Nawrocki-Raby B. 2008. The E-cadherin-repressed hNanos1 gene induces tumor cell invasion by upregulating MT1-MMP expression. *Oncogene* 27:3692–3699 DOI 10.1038/sj.onc.1211035.
- Bruggeman JW, Irie N, Lodder P, Van Pelt AMM, Koster J, Hamer G. 2020. Tumors widely express hundreds of embryonic germline genes. *Cancers (Basel)* 12(12):3812 DOI 10.3390/cancers12123812.
- Bruggeman JW, Koster J, Lodder P, Repping S, Hamer G. 2018. Massive expression of germ cell-specific genes is a hallmark of cancer and a potential target for novel treatment development. *Oncogene* 37:5694–5700 DOI 10.1038/s41388-018-0357-2.
- Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. 2014. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nature Reviews Cancer* 14:535–546 DOI 10.1038/nrc3775.

- Chen YT, Scanlan MJ, Sahin U, Türeci O, Gure AO, Tsang S, Williamson B, Stockert E, Pfreundschuh M, Old LJ. 1997.** A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening. *Proceedings of the National Academy of Sciences of the United States of America* **94**:1914–1918 DOI [10.1073/pnas.94.5.1914](https://doi.org/10.1073/pnas.94.5.1914).
- Chudasama D, Bo V, Hall M, Anikin V, Jeyaneethi J, Gregory J, Pados G, Tucker A, Harvey A, Pink R, Karteris E. 2018.** Identification of cancer biomarkers of prognostic value using specific gene regulatory networks (GRN): a novel role of RAD51AP1 for ovarian and lung cancers. *Carcinogenesis* **39**:407–417 DOI [10.1093/carcin/bgx122](https://doi.org/10.1093/carcin/bgx122).
- Davis S, Meltzer PS. 2007.** GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**:1846–1847 DOI [10.1093/bioinformatics/btm254](https://doi.org/10.1093/bioinformatics/btm254).
- De Backer O, Arden KC, Boretti M, Vantomme V, De Smet C, Czekay S, Viars CS, De Plaen E, Brasseur F, Chomez P, Vanden Eynde B, Boon T, Vander Bruggen P. 1999.** Characterization of the GAGE genes that are expressed in various human cancers and in normal testis. *Cancer Research* **59**:3157–3165.
- De Keuckelaere E, Hulpiau P, Saeys Y, Berx G, van Roy F. 2018.** Nanos genes and their role in development and beyond. *Cellular and Molecular Life Sciences* **75**:1929–1946 DOI [10.1007/s00018-018-2766-3](https://doi.org/10.1007/s00018-018-2766-3).
- Erenpreisa J, Cragg MS. 2010.** MOS, aneuploidy and the ploidy cycle of cancer cells. *Oncogene* **29**:5447–5451 DOI [10.1038/onc.2010.310](https://doi.org/10.1038/onc.2010.310).
- Friedman J, Hastie T, Tibshirani R. 2010.** Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **33**:1–22.
- Gong F, Ren P, Zhang Y, Jiang J, Zhang H. 2016.** MicroRNAs-491-5p suppresses cell proliferation and invasion by inhibiting IGF2BP1 in non-small cell lung cancer. *American Journal of Translational Research* **8**:485–495.
- Guo W, Huai Q, Wan H, Guo L, Song P, Gao S, He J. 2021.** Prognostic impact of IGF2BP3 expression in patients with surgically resected lung adenocarcinoma. *DNA and Cell Biology* **40**:316–331 DOI [10.1089/dna.2020.6136](https://doi.org/10.1089/dna.2020.6136).
- He R, Zuo S. 2019.** A Robust 8-gene prognostic signature for early-stage non-small cell lung cancer. *Frontiers in Oncology* **9**:693 DOI [10.3389/fonc.2019.00693](https://doi.org/10.3389/fonc.2019.00693).
- Herbst RS, Morgensztern D, Boshoff C. 2018.** The biology and management of non-small cell lung cancer. *Nature* **553**:446–454 DOI [10.1038/nature25183](https://doi.org/10.1038/nature25183).
- Huang H, Wang D, Guo W, Zhuang X, He Y. 2019.** Correlated low IGF2BP1 and FOXM1 expression predicts a good prognosis in lung adenocarcinoma. *Pathology - Research and Practice* **215**:152433 DOI [10.1016/j.prp.2019.152433](https://doi.org/10.1016/j.prp.2019.152433).
- Jen J, Lin LL, Lo FY, Chen HT, Liao SY, Tang YA, Su WC, Salgia R, Hsu CL, Huang HC, Juan HF, Wang YC. 2016.** Oncoprotein ZNF322A transcriptionally deregulates alpha-adducin, cyclin D1 and p53 to promote tumor growth and metastasis in lung cancer. *Oncogene* **36**:5219 DOI [10.1038/onc.2017.203](https://doi.org/10.1038/onc.2017.203).
- Jen J, Liu CY, Chen YT, Wu LT, Shieh YC, Lai WW, Wang YC. 2019.** Oncogenic zinc finger protein ZNF322A promotes stem cell-like properties in lung cancer through

- transcriptional suppression of c-Myc expression. *Cell Death & Differentiation* 26:1283–1298 DOI 10.1038/s41418-018-0204-6.
- Jiao Z, Yu A, He X, Xuan Y, Zhang H, Wang G, Shi M, Wang T. 2020.** Bioinformatics analysis to determine the prognostic value and prospective pathway signaling of miR-126 in non-small cell lung cancer. *Annals of Translational Medicine* 8(24):1639 DOI 10.21037/atm-20-7520.
- Kato T, Hayama S, Yamabuki T, Ishikawa N, Miyamoto M, Ito T, Tsuchiya E, Kondo S, Nakamura Y, Daigo Y. 2007.** Increased expression of insulin-like growth factor-II messenger RNA-binding protein 1 is associated with tumor progression in patients with lung cancer. *Clinical Cancer Research* 13:434–442 DOI 10.1158/1078-0432.Ccr-06-1297.
- Kho AT, Zhao Q, Cai Z, Butte AJ, Kim JY, Pomeroy SL, Rowitch DH, Kohane IS. 2004.** Conserved mechanisms across development and tumorigenesis revealed by a mouse development perspective of human cancers. *Genes and Development* 18:629–640 DOI 10.1101/gad.1182504.
- Kleczyko EK, Kwak JW, Schenk EL, Nemenoff RA. 2019.** Targeting the complement pathway as a therapeutic strategy in lung cancer. *Frontiers in Immunology* 10:954 DOI 10.3389/fimmu.2019.00954.
- Lepourcelet M, Tou L, Cai L, Sawada J, Lazar AJ, Glickman JN, Williamson JA, Everett AD, Redston M, Fox EA, Nakatani Y, Shivdasani RA. 2005.** Insights into developmental mechanisms and cancers in the mammalian intestine derived from serial analysis of gene expression and study of the hepatoma-derived growth factor (HDGF). *Development* 132:415–427 DOI 10.1242/dev.01579.
- Li Y, Gu J, Xu F, Zhu Q, Chen Y, Ge D, Lu C. 2020b.** Molecular characterization, biological function, tumor microenvironment association and clinical significance of m6A regulators in lung adenocarcinoma. *Briefings in Bioinformatics* DOI 10.1093/bib/bbaa225.
- Li J, Liao Y, Huang J, Sun Y, Chen H, Chen C, Li S, Yang Z. 2018a.** Epigenetic silencing of ADAMTS5 is associated with increased invasiveness and poor survival in patients with colorectal cancer. *Journal of Cancer Research and Clinical Oncology* 144:215–227 DOI 10.1007/s00432-017-2545-9.
- Li W, Pan T, Jiang W, Zhao H. 2020a.** HCG18/miR-34a-5p/HMMR axis accelerates the progression of lung adenocarcinoma. *Biomedicine & Pharmacotherapy* 129:110217 DOI 10.1016/j.biopha.2020.110217.
- Li Z, Qi F, Li F. 2020c.** Establishment of a gene signature to predict prognosis for patients with lung adenocarcinoma. *International Journal of Molecular Sciences* 21:8479–8498 DOI 10.3390/ijms21228479.
- Li Y, Wang Y, Zhang C, Yuan W, Wang J, Zhu C, Chen L, Huang W, Zeng W, Wu X, Liu M. 2004.** ZNF322, a novel human C2H2 Kruppel-like zinc-finger protein, regulates transcriptional activation in MAPK signaling pathways. *Biochemical and Biophysical Research Communications* 325:1383–1392 DOI 10.1016/j.bbrc.2004.10.183.

- Li S, Xuan Y, Gao B, Sun X, Miao S, Lu T, Wang Y, Jiao W. 2018b.** Identification of an eight-gene prognostic signature for lung adenocarcinoma. *Cancer Management and Research* **10**:3383–3392 DOI [10.2147/cmar.S173941](https://doi.org/10.2147/cmar.S173941).
- Li N, Zhan X. 2020.** Identification of pathology-specific regulators of m(6)A RNA modification to optimize lung cancer management in the context of predictive, preventive, and personalized medicine. *Epma Journal* **11**:485–504 DOI [10.1007/s13167-020-00220-3](https://doi.org/10.1007/s13167-020-00220-3).
- Liang R, Xiao G, Wang M, Li X, Li Y, Hui Z, Sun X, Qin S, Zhang B, Du N, Liu D, Ren H. 2018.** SNHG6 functions as a competing endogenous RNA to regulate E2F7 expression by sponging miR-26a-5p in lung adenocarcinoma. *Biomedicine & Pharmacotherapy* **107**:1434–1446 DOI [10.1016/j.biopha.2018.08.099](https://doi.org/10.1016/j.biopha.2018.08.099).
- Liao SY, Chiang CW, Hsu CH, Chen YT, Jen J, Juan HF, Lai WW, Wang YC. 2017.** CK1 δ /GSK3 β /FBXW7 α axis promotes degradation of the ZNF322A oncoprotein to suppress lung cancer progression. *Oncogene* **36**:5722–5733 DOI [10.1038/onc.2017.168](https://doi.org/10.1038/onc.2017.168).
- Liu C, Li Y, Wei M, Zhao L, Yu Y, Li G. 2019.** Identification of a novel glycolysis-related gene signature that can predict the survival of patients with lung adenocarcinoma. *Cell Cycle* **18**:568–579 DOI [10.1080/15384101.2019.1578146](https://doi.org/10.1080/15384101.2019.1578146).
- Liu P, Wang M, Tang W, Li G, Gong N. 2020.** Circ_SATB2 attenuates the anti-tumor role of celastrol in non-small-cell lung carcinoma through targeting miR-33a-5p/E2F7 axis. *OncoTargets and Therapy* **13**:11899–11912 DOI [10.2147/ott.S279434](https://doi.org/10.2147/ott.S279434).
- Man Y, Cao J, Jin S, Xu G, Pan B, Shang L, Che D, Yu Q, Yu Y. 2014.** Newly identified biomarkers for detecting circulating tumor cells in lung adenocarcinoma. *The Tohoku Journal of Experimental Medicine* **234**:29–40 DOI [10.1620/tjem.234.29](https://doi.org/10.1620/tjem.234.29).
- McFarlane RJ, Wakeman JA. 2017.** Meiosis-like functions in oncogenesis: a new view of cancer. *Cancer Research* **77**:5712–5716 DOI [10.1158/0008-5472.CAN-17-1535](https://doi.org/10.1158/0008-5472.CAN-17-1535).
- McKenna A, Gagnon JA. 2019.** Recording development with single cell dynamic lineage tracing. *Development* **146**:169730–169739 DOI [10.1242/dev.169730](https://doi.org/10.1242/dev.169730).
- Mizutani R, Imamachi N, Suzuki Y, Yoshida H, Tochigi N, Oonishi T, Suzuki Y, Akimitsu N. 2016.** Oncofetal protein IGF2BP3 facilitates the activity of proto-oncogene protein eIF4E through the destabilization of EIF4E-BP2 mRNA. *Oncogene* **35**:3495–3502 DOI [10.1038/onc.2015.410](https://doi.org/10.1038/onc.2015.410).
- Moncada-Pazos A, Obaya AJ, Fraga MF, Vilorio CG, Capellá G, Gausachs M, Esteller M, López-Otín C, Cal S. 2009.** The ADAMTS12 metalloprotease gene is epigenetically silenced in tumor cells and transcriptionally activated in the stroma during progression of colon cancer. *Journal of Cell Science* **122**:2906–2913 DOI [10.1242/jcs.050468](https://doi.org/10.1242/jcs.050468).
- Müller S, Glaß M, Singh AK, Haase J, Bley N, Fuchs T, Lederer M, Dahl A, Huang H, Chen J, Posern G, Hüttelmaier S. 2019.** IGF2BP1 promotes SRF-dependent transcription in cancer in a m6A- and miRNA-dependent manner. *Nucleic Acids Research* **47**:375–390 DOI [10.1093/nar/gky1012](https://doi.org/10.1093/nar/gky1012).

- Ohdaira H, Sekiguchi M, Miyata K, Yoshida K. 2012.** MicroRNA-494 suppresses cell proliferation and induces senescence in A549 lung cancer cells. *Cell Proliferation* 45:32–38 DOI [10.1111/j.1365-2184.2011.00798.x](https://doi.org/10.1111/j.1365-2184.2011.00798.x).
- Park SA, Platt J, Lee JW, López-Giráldez F, Herbst RS, Koo JS. 2015.** E2F8 as a novel therapeutic target for lung cancer. *Journal of National Cancer Institute* 107:151–166 DOI [10.1093/jnci/djv151](https://doi.org/10.1093/jnci/djv151).
- Pires E, Sung P, Wiese C. 2017.** Role of RAD51AP1 in homologous recombination DNA repair and carcinogenesis. *DNA Repair* 59:76–81 DOI [10.1016/j.dnarep.2017.09.008](https://doi.org/10.1016/j.dnarep.2017.09.008).
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015.** limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43:e47 DOI [10.1093/nar/gkv007](https://doi.org/10.1093/nar/gkv007).
- Sankaran D, Pakala SB, Nair VS, Sirigiri DN, Cyanam D, Ha NH, Li DQ, Santhoshkumar TR, Pillai MR, Kumar R. 2012.** Mechanism of MTA1 protein overexpression-linked invasion: MTA1 regulation of hyaluronan-mediated motility receptor (HMMR) expression and function. *Journal of Biological Chemistry* 287:5483–5491 DOI [10.1074/jbc.M111.324632](https://doi.org/10.1074/jbc.M111.324632).
- Shen S, Kong J, Qiu Y, Yang X, Wang W, Yan L. 2019.** Identification of core genes and outcomes in hepatocellular carcinoma by bioinformatics analysis. *Journal of Cellular Biochemistry* 120:10069–10081 DOI [10.1002/jcb.28290](https://doi.org/10.1002/jcb.28290).
- Siegel RL, Miller KD, Jemal A. 2020.** Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians* 70:7–30 DOI [10.3322/caac.21590](https://doi.org/10.3322/caac.21590).
- Simpson AJ, Caballero OL, Jungbluth A, Chen YT, Old LJ. 2005.** Cancer/testis antigens, gametogenesis and cancer. *Nature Reviews Cancer* 5:615–625 DOI [10.1038/nrc1669](https://doi.org/10.1038/nrc1669).
- Stevens LE, Cheung WKC, Adua SJ, Arnal-Estapé A, Zhao M, Liu Z, Brewer K, Herbst RS, Nguyen DX. 2017.** Extracellular Matrix Receptor Expression in Subtypes of Lung Adenocarcinoma Potentiates Outgrowth of Micrometastases. *Cancer Research* 77:1905–1917 DOI [10.1158/0008-5472.Can-16-1978](https://doi.org/10.1158/0008-5472.Can-16-1978).
- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. 2021.** Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* 71:209–249 DOI [10.3322/caac.21660](https://doi.org/10.3322/caac.21660).
- Trott J, Martinez Arias A. 2013.** Single cell lineage analysis of mouse embryonic stem cells at the exit from pluripotency. *Biology Open* 2:1049–1056 DOI [10.1242/bio.20135934](https://doi.org/10.1242/bio.20135934).
- Tu Z, Chen X, Tian T, Chen G, Huang M. 2021.** Prognostic significance of epigenetic regulatory gene expression in patients with non-small-cell lung cancer. *Aging (Albany NY)* 13:7397–7415.
- Türeci O, Sahin U, Schobert I, Koslowski M, Scmitt H, Schild HJ, Stenner F, Seitz G, Rammensee HG, Pfreundschuh M. 1996.** The SSX-2 gene, which is involved in the t(X;18) translocation of synovial sarcomas, codes for the human tumor antigen HOM-MEL-40. *Cancer Research* 56:4766–4772.
- Türeci O, Sahin U, Zwick C, Koslowski M, Seitz G, Pfreundschuh M. 1998.** Identification of a meiosis-specific protein as a member of the class of cancer/testis antigens.

- Proceedings of the National Academy of Sciences of the United States of America* 95:5211–5216 DOI 10.1073/pnas.95.9.5211.
- Van der Bruggen P, Traversari C, Chomez P, Lurquin C, DePlaen E, Vanden Eynde B, Knuth A, Boon T. 1991.** A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science* 254:1643–1647 DOI 10.1126/science.1840703.
- Wang Y, He R, Ma L. 2020.** Characterization of lncRNA-associated ceRNA network to reveal potential prognostic biomarkers in lung adenocarcinoma. *Frontiers in Bioengineering and Biotechnology* 8:266 DOI 10.3389/fbioe.2020.00266.
- Wang Y, Wo Y, Lu T, Sun X, Liu A, Dong Y, Du W, Su W, Huang Z, Jiao W. 2021.** Circ-AASDH functions as the progression of early stage lung adenocarcinoma by targeting miR-140-3p to activate E2F7 expression. *Translational Lung Cancer Research* 10:57–70 DOI 10.21037/tlcr-20-1062.
- Whitehurst AW. 2014.** Cause and consequence of cancer/testis antigen activation in cancer. *Annual Review of Pharmacology and Toxicology* 54:251–272 DOI 10.1146/annurev-pharmtox-011112-140326.
- Williams KA, Lee M, Hu Y, Andreas J, Patel SJ, Zhang S, Chines P, Elkahlon A, Chandrasekharappa S, Gutkind JS, Molinolo AA, Crawford NP. 2014.** A systems genetics approach identifies CXCL14, ITGAX, and LPCAT2 as novel aggressive prostate cancer susceptibility genes. *PLoS Genetics* 10:e1004809 DOI 10.1371/journal.pgen.1004809.
- Wu Y, Wang H, Qiao L, Jin X, Dong H, Wang Y. 2019.** Silencing of RAD51AP1 suppresses epithelial-mesenchymal transition and metastasis in non-small cell lung cancer. *Thorac Cancer* 10:1748–1763 DOI 10.1111/1759-7714.13124.
- Xueqing H, Jun Z, Yueqiang J, Xin L, Liya H, Yuanyuan F, Yuting Z, Hao Z, Hua W, Jian L, Tiejun Y. 2020.** IGF2BP3 may contribute to lung tumorigenesis by regulating the alternative splicing of PKM. *Frontiers in Bioengineering and Biotechnology* 8:679 DOI 10.3389/fbioe.2020.00679.
- Yang CY, Yang JC, Yang PC. 2020.** Precision management of advanced non-small cell lung cancer. *Annual Review of Medicine* 71:117–136 DOI 10.1146/annurev-med-051718-013524.
- Yu G, Wang LG, Han Y, He QY. 2012.** clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 16:284–287 DOI 10.1089/omi.2011.0118.
- Yuan Y, Zhou X, Kang Y, Kuang H, Peng Q, Zhang B, Liu X, Zhang M. 2021.** Circ-CCS is identified as a cancer-promoting circRNA in lung cancer partly by regulating the miR-383/E2F7 axis. *Life Sciences* 267:118955 DOI 10.1016/j.lfs.2020.118955.
- Yue C, Ma H, Zhou Y. 2019.** Identification of prognostic gene signature associated with microenvironment of lung adenocarcinoma. *PeerJ* 7:e8128 DOI 10.7717/peerj.8128.
- Zhang J, Luo W, Chi X, Zhang L, Ren Q, Wang H, Zhang W. 2020a.** IGF2BP1 silencing inhibits proliferation and induces apoptosis of high glucose-induced non-small cell lung cancer cells by regulating Netrin-1. *Archives of Biochemistry and Biophysics* 693:108581 DOI 10.1016/j.abb.2020.108581.

- Zhang S, Sun K, Zheng R, Zeng H, Wang S, Chen R, Wei W, He J. 2020b.** Cancer incidence and mortality in China, 2015. *Journal of the National Cancer Center* 1:2–11 DOI [10.1016/j.jncc.2020.12.001](https://doi.org/10.1016/j.jncc.2020.12.001).
- Zhang L, Zhang Z, Yu Z. 2019a.** Identification of a novel glycolysis-related gene signature for predicting metastasis and survival in patients with lung adenocarcinoma. *Journal of Translational Medicine* 17:423 DOI [10.1186/s12967-019-02173-2](https://doi.org/10.1186/s12967-019-02173-2).
- Zhang Y, Wang LF, Gao JH, Li L, Jiang P, Lv X, Yu LX, Yang J, Li RT, Liu BR. 2019b.** Clinical significance of epithelial-mesenchymal transition-related molecules in lung adenocarcinoma. *Current Oncology* 26:e121–e127 DOI [10.3747/co.26.4471](https://doi.org/10.3747/co.26.4471).
- Zheng S, Yao L, Li F, Huang L, Yu Y, Lin Z, Li H, Xia J, Lanuti M, Zhou H. 2021.** Homologous recombination repair pathway and RAD54L in early-stage lung adenocarcinoma. *PeerJ* 9:e10680 DOI [10.7717/peerj.10680](https://doi.org/10.7717/peerj.10680).
- Zhou H, Wang L, Huang J, Jiang M, Zhang X, Zhang L, Wang Y, Jiang Z, Zhang Z. 2015.** High EGFR_1 inside-out activated inflammation-induced motility through SLC2A1-CCNB2-HMMR-KIF11-NUSAP1-PRC1-UBE2C. *Journal of Cancer* 6:519–524 DOI [10.7150/jca.11404](https://doi.org/10.7150/jca.11404).