Identification of key genes associated with progression and prognosis for Lung Squamous Cell Carcinoma

Xiaohan Ma [Equal first author, 1, 2] , **Huijun Ren** [Equal first author, 1, 2] , **Ruoyu Peng** [1, 2] , **Yi Li** [1, 2] , **Liang Ming** [Corresp. 1, 2]

1 Department of Clinical Laboratory, the First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China

2 Key Clinical Laboratory of Henan Province, Zhengzhou, Henan, China

Corresponding Author: Liang Ming
Email address: mingliang_2015@sina.com

Background. Lung squamous cell carcinoma (LUSC) is a major subtype of lung cancer with limited therapeutic options and poor clinical prognosis.

Methods. Three datasets (GSE19188, GSE33532, and GSE33479) were obtained from the gene expression omnibus (GEO) database. Differentially expressed genes (DEGs) between LUSC and normal tissues were identified by GEO2R, and functional analysis was employed using DAVID online tool. Protein-protein interaction (PPI) and hub genes were identified via Search Tool for the Retrieval of Interacting Genes (STRING) and Cytoscape software. Hub genes were further validated in TCGA database. Subsequently, survival analysis was performed using the Kapla-Meier curve and Cox progression analysis. Based on univariate and multivariate Cox progression analysis , a gene signature was established to predict overall survival. Receiver operating characteristic (ROC) curve was used to evaluate the prognostic value of the model.

Results. A total of 116 up-regulated genes and 84 down-regulated genes were identified. These DEGs were mainly enriched in the two pathways: cell cycle and p53 signaling way. According to the degree of protein nodes in the PPI network, ten hub genes were identified. The mRNA expression levels of the ten hub genes in LUSC were also significantly up-regulated in TCGA database. Furthermore, a novel seven-gene signature (FLRT3, PPP2R2C, MMP3, MMP12, CAPN8, FILIP1 and SPP1) from the DEGs was constructed and acted as a significant and independent prognostic signature for LUSC.

Conclusions. The ten hub genes might be tightly correlated with LUSC progression. The seven-gene signature might be an independent biomarker with a significant predictive value in LUSC overall survival.

1 Identification of key genes associated with progression and

2 prognosis for Lung Squamous Cell Carcinoma

4   Xiaohan Ma[*1,2], Huijun Ren[*1,2], Ruoyu Peng[1,2],Yi Li[1,2], Liang Ming[1,2]

5   [1] Department of Clinical Laboratory, the First Affiliated Hospital of Zhengzhou University, Zhengzhou,
6   Henan, China

7   [2] Key Clinical Laboratory of Henan Province, Zhengzhou, Henan, China

8   * These authors are co-first authors on this work.

10
11
12
13 Corresponding Author: Liang Ming
14 No. 1 Jianshe Road, Zhengzhou, Henan, 450052, China
15 Email address: mingliang_2015@sina.com

Abstract

16 **Background.** Lung squamous cell carcinoma (LUSC) is a major subtype of lung cancer with

17 limited therapeutic options and poor clinical prognosis.

18 **Methods.** Three datasets (GSE19188, GSE33532, and GSE33479) were obtained from the gene

19 expression omnibus (GEO) database. Differentially expressed genes (DEGs) between LUSC and

20 normal tissues were identified by GEO2R, and functional analysis was employed using DAVID

21 online tool. Protein-protein interaction (PPI) and hub genes were identified via Search Tool for

22 the Retrieval of Interacting Genes (STRING) and Cytoscape software. Hub genes were further

23 validated in TCGA database. Subsequently, survival analysis was performed using the Kapla-

24 Meier curve and Cox progression analysis. Based on univariate and multivariate Cox progression

25 analysis, a gene signature was established to predict overall survival. Receiver operating

26 characteristic (ROC) curve was used to evaluate the prognostic value of the model.

27 **Results.** A total of 116 up-regulated genes and 84 down-regulated genes were identified. These

28 DEGs were mainly enriched in the two pathways: cell cycle and p53 signaling way. According to

29 the degree of protein nodes in the PPI network, ten hub genes were identified. The mRNA

30 expression levels of the ten hub genes in LUSC were also significantly up-regulated in TCGA

31 database. Furthermore, a novel seven-gene signature (FLRT3, PPP2R2C, MMP3, MMP12,

32 CAPN8, FILIP1 and SPP1) from the DEGs was constructed and acted as a significant and

33 independent prognostic signature for LUSC.

34 **Conclusions.** The ten hub genes might be tightly correlated with LUSC progression. The seven-

35 gene signature might be an independent biomarker with a significant predictive value in LUSC

36 overall survival.

37

Introduction

40 Lung cancer, as a highly malignant cancer, is still a common cause for healthy issues

41 worldwide. With approximately 1.8 million deaths in 2018 (*Bray et al., 2018*), lung cancer has

42 been ranked in top 2 in China and possessed the highest mortality rate (*Sun et al., 2018*). Among

43 all the lung cancer types, non-small cell lung cancer (NSCLC) accounts for over 85% of total

44 cases. Lung squamous cell carcinoma (LUSC) is the second most frequent subtype of NSCLC,

45 accounting for about 40% of NSCLC (*Chen et al., 2014*), and its therapy and prognosis are still

46 facing huge challenges. Nowadays, surgery and adjuvant chemotherapy are the standard

47 treatment for stage I and II of NSCLC, however, molecular analysis is the key to select a first-
48 line therapy for advanced cancer (*Kuo et al., 2019*). Large numbers of gene mutations been
49 reported to be served as specific biomarkers for diagnosis, treatment and prognosis for LUAD
50 (lung adenocarcinoma) (*Calvayrac et al., 2017*), such as bevacizumab against VEGF. Although
51 complex genomic alterations were found in LUSC, which were different from LUAD, there were
52 no available and specific targeted agents for LUSC until now (*Hirsch et al., 2017*). Despite great
53 progress made in combined therapies, the prognosis of LUSC is still dismal. Hence, there is an
54 imperative and urgent need for identify key molecules for therapy and prognosis of LUSC.

55   Recently, high throughput sequencing and microarray technologies have been widely used
56 to investigate relationship between diverse diseases and key molecules, including genes,
57 miRNAs, long non-coding RNAs (lncRNAs), and circRNAs. Yeo et al (*Yeo et al., 2017*) proved
58 that programmed cell death 1 was over-expressed in LUSC and could be useful for the prediction
59 of poor prognosis. Hou et al identified 95 up-regulated and 749 down-regulated lncRNAs in
60 response to cisplatin chemo, which indicate that dysregulated lncRNAs are related with therapy
61 and are prognostic markers in LUSC (*Hou et al., 2014*). Several other key molecules were also
62 found involved in the development, diagnosis, and prognosis of LUSC, such as
63 circRNA_103827, circRNA_000122 (*Xu et al., 2018*), peroxiredoxin 4 (*Hwang et al., 2015*),
64 AURKA, BIRC5, LINC00094 (*Li et al., 2017*), and so on. The dysregulation of these mentioned
65 molecules is associated with the progression and prognosis of LUSC, however, limited samples
66 and significant variability among different projects reduce the credibility of these obtained
67 results.

68   In order to search for promising key genes associated with the progression and prognosis of
69 LUSC, differentially expressed genes (DEGs) in LUSC were identified using three microarray
70 datasets from the Gene Expression Omnibus (GEO) database. Subsequently, ten hub genes were
71 identified by Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG)
72 enrichment analysis and protein-protein interaction (PPI) network, which were involved in cell
73 cycle and p53 pathway in LUSC. The expression of the ten hub genes was validated in TCGA
74 database. In addition, a novel seven-gene signature was established to predict effectively overall
75 survival in LUSC.

76

77 **Materials & Methods**

78 **Data collection**

79   The mRNA expression profiles and corresponding clinical information were acquired from

80 the GEO database (https://www.ncbi.nlm.nih.gov/geo/) and the Cancer Genome Atlas (TCGA)

81 database (https://portal.gdc.cancer.gov). With searching for "lung squamous cell carcinoma", a

82 total of 4627 series about LUSC were searched from the GEO database. After a careful review,

83 three gene expression profiles (GSE19188, GSE33532, and GSE33479) were collected. The

84 former two databases were both based on GPL570 platform ([HG-U133_Plus_2] Affymetrix

85 Human Genome U133 Plus 2.0 Array), and GSE33479 was based on GPL6480 platform

86 (Agilent-014850 Whole Human Genome Microarray 4x44K G4112F). 57 LUSC tissue samples

87 and 112 normal lung tissue samples were collected from the three GEO datasets. 551 samples

88 were collected from TCGA database, containing 502 LUSC tissues samples and 49 normal tissue

89 samples.

90 **Data pre-processing and identification of DEGs**

91   In this study, the GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/), an interactive web

92 tool used to compare two groups of samples according to GEO series, was applied to detect the

93 DEGs between 57 LUSC samples and 112 normal samples. The adjusted P-value < 0.05 and

94 |log2 (fold change)| ≥ 2.0 were regarded as the cutoff criteria to select the DEGs. Then the DEGs

95 obtained from different GSE datasets were further identified using the Venn diagram web tool

96 (bioinformatics.psb.ugent.be/ webtools/Venn/).

97 **Functional analysis and PPI network construction of DEGs**

98   Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of

99 DEGs were performed using the Database for Annotation, Visualization, and Integrated

100 Discovery (DAVID) (https://david.ncifcrf.gov/). The adjusted *P*-value < 0.05 was considered as

101 statistically significant. In order to evaluate protein-protein interactions of DEGs, Cytoscape

102 software and the Search Tool for the Retrieval of Interacting Genes (STRING) (https://string-

103 db.org/) were used to analyze PPI relationship, and PPI pairs with a combined score > 0.4 were

104 extracted. A cytoscape plugin, cytoHubba, was utilized to calculate the degree of protein nodes,

105 and the top rank ten genes were selected as hub genes.

106 **Hub genes validation and analysis**

107   The mRNA expression profiles in 502 LUSC samples and 49 normal samples from TCGA

108 database were analyzed by the "limma" R package. The statistical analysis was performed using

109   the Wilcox test, and adjusted P < 0.05 and |log2 (fold change)| ≥ 2.0 were selected as the cutoff

110   criteria. Subsequently, the expression levels of hub genes were validated in TCGA database.

111   Besides, the expression levels of the ten hub genes in LUSC were compared with other NSCLC

112   histologic subtypes using the Oncomine database (www.oncomine.org).

113   The Cox proportional hazards regression model was used for overall survival (OS) analysis

114   among the ten hub genes. Besides, Kaplan-Meier Plotter (http://kmplot.com/) was used to assess

115   the effect of the hub genes on LUSC prognosis. The option "only JetSet best probe set" was

116   selected for probes of genes, and only the LUSC patients were selected for analysis. At last, there

117   were 524 LUSC patients for OS analysis, 141 LUSC patients for FP analysis, and 20 LUSC

118   patients for PPS analysis in the Kaplan-Meier Plotter. P < 0.05 was regarded as statistically

119   significant. Then the genes with prognostic values were analyzed to identify their associations

120   with tumor grade using the Oncomine database.

121   **Survival analysis**

122   In order to identify the potential prognostic values of the DEGs in LUSC, univariate and

123   multivariate Cox regression were conducted using the R package. The clinical information of

124   LUSC patients was downloaded from TCGA database, and 488 LUSC patients were used for

125   analysis after removing the patients (14 of 502) with incomplete clinical data. Only 187 of 200

126   DEGs could be found and validated in TCGA database, and they were used for the univariate

127   Cox regression analysis. The genes associated with overall survival (P < 0.05) were subjected to

128   the multivariate Cox proportional hazards model to establish a robust gene prognostic signature

129   for LUSC. The LUSC patients were further grouped into "high-risk" and "low-risk" based on the

130   median risk score. What's more, a receiver operating characteristic (ROC) curve was constructed

131   to evaluate the predictive accuracy of the gene signature by using the R package "survival ROC".

132   Besides, the gene signature and clinicopathological parameters (age, gender, tumor stage, T/N/M

133   status) were submitted to Cox regression analysis to identify independent factors for OS in

134   LUSC patients.

135   **Results**

136   **Identification of DEGs in LUSC**

137   All the mRNA expression profiles of GSE19188, GSE33479, and GSE33532 were provided

138   in Table S1-3. According to the criteria of adjusted P-value < 0.05 and |log2 (fold change)| ≥ 2, a

139   total of 458 up-regulated genes and 668 down-regulated genes were identified in GSE19188 (Fig.

140   1A). From GSE33479, 1153 DEGs were identified, including 499 up-regulated genes and 654

141   down-regulated genes (Fig. 1B). From GSE33532, 1375 DEGs including 569 up-regulated genes

142   and 806 down-regulated genes were identified (Fig. 1C). All the DEGs were identified by

143   comparing LUSC samples and normal lung samples. Subsequently, the results from the three

144   studies were analyzed using the Venn diagram tool, and 116 up-regulated genes and 84 down-

145   regulated genes were identified (Fig. 1D-E, Table S4).

146   **Functional analysis and PPI network construction**

147   GO and KEGG pathway enrichment analysis of DEGs were carried out to explore the

148   biological functions of 200 DEGs (Table 1). The enriched GO terms of DEGs were classified

149   into three categories: molecular functions (MF), cellular components (CC), and biological

150   processes (BP). As shown in Fig. 2, the DEGs were mainly enriched in "mitotic nuclear

151   division", "cell division", and "sister chromatid cohesion" in the BP category. Within the CC

152   category, "condensed chromosome kinetochore", "chromosome", and "midbody" were

153   predominant. The results of KEGG analysis exhibited that the DEGs were mainly involved in

154   "Cell cycle" and "p53 signaling pathway".

155   PPI analysis could reflect the molecular mechanisms of cancer physiological and

156   pathological changes. The PPI network of common DEGs was constructed using the STRING

157   V11.0 tool and Cytoscape software (Fig. 3), which containing 199 nodes and 2363 edges.

158   Subsequently, the top ten genes with high node degree were selected as the hub genes for further

159   analysis (Fig. 4). Results showed that the top ten hub genes were significantly up-regulated in

160   LUSC tissues. Among them, cyclin-dependent kinases 1 (CDK1) was the most significant gene

161   with connectivity degree=57, followed by BUB1 mitotic checkpoint serine/threonine kinase

162   (BUB1; degree=48), cyclin B1 (CCNB2; degree=48), cyclin A2 (CCNA2; degree=46), kinesin

163   family member 2C (KIF2C; degree=41), aurora kinase B (AURKB; degree=41), kinesin family

164   member 11 (KIF11; degree=41), mitotic arrest deficient 2 like 1 (MAD2L1; degree=41), DNA

165   topoisomerase II alpha (TOP2A; degree=40), and DLG associated protein 5 (DLGAP5;

166   degree=39).

167   **Hub genes validation and analysis**

168   The expression of the ten hub genes was validated in TCGA database, which was consistent

169   with the results from the GEO database (Fig. 5). The Analysis indicated that their expression

170   levels in LUSC were significantly higher than those in other NSCLC (Fig. 6). Cox regression

171 model was used to analyze the ten hub genes, however, none of them were associated with OS,

172 and no significant gene signature could be established to predict OS in LUSC patients. Kaplan-

173 Meier Plotter showed that the up-regulated CCNA2, KIF11, MAD2L1, and DLGAP5 were all

174 related to worse first-progression survival (FP) in LUSC patients (Fig. 7, A-D). Only the up-

175 regulation of KIF2C was associated with unfavorable post-progression survival (PPS) in LUSC

176 patients (Fig. 7E). Among the five hub genes with prognostic values genes, four (CCNA2,

177 DLGAP5, MAD2L1, and KIF2C) were associated with LUSC tumor grade (Fig. 8).

178 **Cox progression analysis and construction of prognostic signature**

179   Only 187 of 200 DEGs were found and validated in TCGA database. Then the 187 DEGs

180 were used for the univariate Cox regression analysis, and 23 of 187 DEGs were significantly

181 associated with the OS (P < 0.05), as shown in Table S5. Subsequently, a seven-gene prognostic

182 signature was established by the multivariate Cox regression analysis (Table 2). Risk Score =

183 0.0341 × expression of FLRT3 + 0.0339 × expression of PPP2R2C + 0.0058 × expression of

184 MMP3 + 0.0031 × expression of MMP12 + 0.0518 × expression of CAPN8 + 0.1323 ×

185 expression of FILIP1 + 0.0002 × expression of SPP1. According to the signature, risk score was

186 calculated for each patient and each patient was grouped into "high-risk" and "low-risk"

187 according to the median risk score. The distribution of signature risk score (Fig. 9A), survival

188 status and survival time (Fig. 9B) in the model group were presented. The expression of the

189 seven genes in high and low-risk groups was shown in Fig. 9C. Compared with the low-risk

190 patients, the high-risk patients presented worse OS (Fig. 9D). According to the ROC curve on 5-

191 year OS (Fig. 9E), an area under curve (AUC) value of 0.707 (> 0.7) indicated that the

192 prognostic prediction value of the seven-gene signature was reliable.

193 **Independence assessment of seven-gene mRNA signature**

194   Univariate Cox progression analysis indicated that risk score, age, tumor stage and T status

195 were significantly associated with the OS among LUSC patients. Subsequently, multivariate Cox

196 analysis showed that the gene signature had a significantly independent prognostic value with *P*

197 < 0.001 (Table 3). These results revealed that the seven-gene risk signature was an independent

198 prognostic factor for LUSC after adjusting the confounding effects.

199 **Discussion**

200   In this study, a total of 116 up-regulated DEGs and 84 down-regulated DEGs were

201 identified from the GEO database. KEGG enrichment analysis showed that these DEGs were

202   mainly involved in two pathways: "Cell cycle" and "p53 signaling pathway". Increasing

203   evidences suggest that disordered cell cycle has been a mark of tumors (*Hanahan & Weinberg,*

204   *2011*). In LUSC, cell cycle progression and cell proliferation could be inhibited by CCNB1

205   (*Wang et al., 2019*), promoted by DDA1 (*Cheng et al., 2017*), and arrested in the G phase by

206   SART3 (*Sherman et al., 2019*). Besides, the cell cycle is closely related with the p53 pathway

207   (*Joerger and Fershthe, 2016*). P53, a multifunctional transcription factor in cancer progression,

208   participates in the regulation of cell cycle, metabolic pathways, and so on (*Stegh, 2012; Fridman*

209   *& Lowe, 2003; Vogelstein et al., 2000*). Studies have found that p53 signaling pathway could

210   induce cancer cell apoptosis by targeting at Bax (*Jin et al., 2017*), p21, and HIF1α (*Yang et al.,*

211   *2018*) in LUSC. Besides, p53 plays the role in tumor suppression in LUSC via being regulated

212   by other genes, such as miR-223-3p (*Luo et al., 2019*), which has not been found in the studies of

213   other cancers. In short, the enriched pathways of these DEGs could interact with each other and

214   participate in the regulation of LUSC progression.

215   In order to predict the function associations between the 200 DEGs, a PPI network was

216   constructed and ten hub genes were identified, including CDK1, BUB1, CCNB2, CCNA2,

217   KIF2C, AURKB, KIF11, MAD2L1, TOP2A, and DLGAP5. Survival analysis revealed that five

218   of hub genes were associated with FP or PPS. Among the five hub genes, the altered CCNA2,

219   DLGAP5, MAD2L1 and KIF2C were associated with tumor grade, and were significantly up-

220   regulated in LUSC compared with other NSCLC, implicating crucial roles of them in LUSC

221   progression.

222   CCNA2, also known as CyclinA2, a highly conserved cyclin protein (*Ko et al., 2013*), has

223   been found significantly over-expressed in various cancers (*Gao et al., 2014*). Previous

224   researches showed that CCNA2 could regulate cell cycle in cancers via controlling the G1/S and

225   G2/M transitions (*Arsic et al., 2012*). CCNA2 is also associated with epithelial-mesenchymal

226   transition and cancer progression (*Bendris et al., 2012*). Previous studies have shown that diverse

227   molecules, such as coiled-coil domain containing 6, Zinc finger SWIM-type containing 5, and

228   miR-137, could regulate the proliferation, invasion, and migration of NSCLC cells by changing

229   the expression of CCNA2 (*Morra et al., 2015*; *Xu et al., 2018*; *Chen et al., 2017*). In the present

230   study, CCNA2 was found up-regulated in LUSC tissues compared with normal tissues, and over-

231   expression of CCNA2 was related to worse first-progression survival in LUSC patients. These

232 results indicated that CCNA2 might be a progression biomarker and prognostic indictor for
233 LUSC.

234   Kinesin family member 11 (KIF11), a kinesin-5-family protein, could affect tumor
235 development by controlling the correct arrangement of the microtubules, which was the key
236 stage in mitosis *(Blangy et al., 1995)*. Previous studies have reported that KIF11 was up-
237 regulated in lung cancer tissues compared with normal tissues, and associated with poor overall
238 survival *(Al-Khafaji et al., 2017*; *Schneider et al., 2017)*. Kato et al declared that high-level
239 KIF11 was as an independent prognostic factor in LUSC, and might be promising therapeutic
240 option for advanced lung cancer *(Kato et al., 2018)*. In this study, KIF11 was both identified as
241 one of over-expressed hub genes and related to unfavorable first-progression survival of LUSC
242 patients, which was consistent with previous studies.

243   MAD2 mitotic arrest deficient-like 1 (MAD2L1), an important spindle checkpoint protein,
244 plays important roles in protecting cells from abnormal chromosome segregation *(Yu, 2006)*.
245 MAD2L1 has been shown high expression levels and might be a biomaker for poor prognosis in
246 various cancers, such as breast cancer *(Wang et al., 2015)*, osteosarcoma *(Sun, Li & Yan, 2015)*,
247 and so on. MAD2L1 has been identified as a potential therapeutic target gene in NSCLC *(Zhou
248 et al., 2015)*, and could be a promising prognostic biomarker for LUAD *(MacDermed et al.,
249 2010)* and small cell lung cancer *(Liao et al., 2019)*, however, there was no similar results in
250 LUSC. In this study, over-expression of MAD2L1 and its prognostic value in LUSC were
251 identified.

252   Kinesin family member 2C (KIF2C) has been served as a modulator in bipolar spindle
253 formation, microtubule depolymerization, and chromosome segregation, and it could promote
254 the tumor proliferation and metastasis *(Gan et al., 2019)*. In this study, down-regulated KIF2C
255 had a significant value in post-progression survival in LUSC. These suggest that KIF2C might be
256 involved in tumor progression and be a potential prognostic factor in LUSC. Disc large
257 (drosophila) homolog-associated protein 5 (DLGAP5), as an important mitotic spindle protein,
258 participated in cancers development and progression *(Liao et al., 2013)*. DLGAP5 has been
259 found significantly over-expressed in different subtypes of lung cancer, and it could be a
260 promising prognostic biomarker and therapeutic target *(Qi et al., 2019)*, which are consistent
261 with the results of this study.

262   Proteins encoded by CDK1, BUB1 and AURKB belong to the serine/threonine kinases

263   family, and overexpression of them has been detected involved in various tumors prognosis via

264   regulating tumor cell cycle. CDK1 (*Xie et al., 2019*) and BUB1 (*Piao et al., 2019*) have been

265   proved to influence tumor progression by inducing cell cycle dysregulation. AURKB was proved

266   to be a therapeutic target by triggering G1/S arrest in NSCLC (*Bertran-Alamillo et al., 2019*).

267   CCNB2, encoding the member of the cyclin family, could promote tumor progression by

268   facilitating cell proliferation and maintain normal G2/M transition. CCNB2 and BUB1 might be

269   involved in the cancer stem cells to promote LUSC progression (*Qin et al., 2020*). TOP2A

270   encoding a DNA topoisomerase, could affect overall survival and clinicopathological features in

271   NSCLC by altering the transcription of DNA (*Hou et al., 2017*). Much further investigations on

272   the roles of the ten genes in LUSC are needed. In this report, functional analysis and their roles

273   in other tumors might provide valuable clues to investigate their roles in LUSC progression.

274   In this study, none of the ten hub gene was found associated with OS in LUSC by K-M

275   plotter analysis and cox progression analysis using TCGA database. Hence, cox progression

276   analysis was performed for all the DEGs. Compared to a single gene marker to predict patient

277   survival, a gene signature will provide a stable and effective prediction effect. Therefore, a novel

278   seven-gene signature (including FLRT3, PPP2R2C, MMP3, MMP12, CAPN8, FILIP1, and

279   SPP1) was established for LUSC prognostic prediction. According to the gene signature, each

280   LUSC patient with a risk score was classified as high-risk or low-risk. The risk scores of high-

281   risk patients were much higher than that of low-risk patients, which proved that the risk score

282   was significant. Results showed that high-risk patients presented significantly worse OS than

283   low-risk patients. ROC demonstrated that the seven-gene signature was efficient and sensitive in

284   the survival prediction in LUSC. In addition, clinical pathological features (including age, gender,

285   grade, and T/M/N state) are known prognostic factors in LUSC. In order to avoid the

286   interference effect of these known prognostic factors, multivariate Cox regression analysis was

287   carried out, and proved that the signature was an independent prognostic factor in LUSC. All

288   these results suggested that the signature could be an efficient and independent indicator for

289   LUSC prognosis.

290   Bluemn et al. demonstrated the loss of PPP2R2C is associated with cancer recurrence and a

291   poor survival in prostate cancer (*Bluemn et al., 2013*). Nevertheless, the role of PPP2R2C in

292   LUSC remains unclear. Previous studies have proved MMP3 (matrix metalloproteinase 3) played

293   crucial roles in invasion and metastasis in many cancers (*Ma et al., 2019*). MMP3 is associated

294   with poor survival in various cancers, and its polymorphism might increase the risk of lung

295   cancer (*Hu et al., 2013*). Up-regulated MMP12 has been found associated with the pathological

296   stage and tumor metastasis in lung cancer (*Lv et al., 2015*). CAPN8 has been proved up-

297   regulated in lung cancer and could be a prognostic biomarker (*Zhang et al., 2015*). FILIP1L

298   (FILamin A Interacting Protein 1-Like) is involved in cell proliferation and migration by

299   inhibiting the WNT signaling pathway (*Kwon et al., 2014*). Up-regulated SPP1 (secreted

300   phosphoprotein 1) could promote cell proliferation, migration, and invasion by Integrin

301   β1/FAK/AKT pathway (*Zeng et al.,2018*), and could be a prognostic indicator in many tumors

302   (*Choe et al.,2018*). Inhibition of SPP1 could enhance the invasion and might be a promising

303   target for NSCLC therapy (*Wang et al., 2019*). To our knowledge, the seven-gene signature for

304   prognostic prediction in LUSC has not been reported previously, and has been demonstrated as

305   an independent and useful prognostic signature in LUSC.

306   In addition, in the study, many more genes are specific to one database but not the others.

307   This might be caused by the heterogeneity between different datasets. Various factors might lead

308   to the heterogeneity, such as experimental methods and conditions, detection platform and

309   system, sample sources, reagent batches, and so on. Some aspects in the present study have been

310   considered to reduce the influences. Firstly, all of the samples in this study are from normal lung

311   tissues and LUSC tissues, and not from adjacent tissues, which ensures the consistency of the

312   experimental grouping. Secondly, all the data from the GEO database have been normalized,

313   which could reduce variation between groups and make them comparable. Lastly, the DEGs

314   were identified from each dataset, and then the common part of them were selected as the final

315   DEGs, which could reduce the influences resulted from the heterogeneity of the different

316   datasets. These aspects could make the analysis results more effective and objective.

Conclusions

318   In summary, this study identified 200 DEGs between LUSC and normal tissues by three

319   GEO datasets. Ten up-regulated hub genes were validated in TCGA database and were

320   associated with cell cycle and p53 signaling pathway. CCNA2, DLGAP5, MAD2L1, and KIF2C

321   were significantly up-regulated compared to other subtypes, and associated with tumor stage in

322   LUSC, suggesting that they might tightly be involved in progression and prognosis of LUSC. In

323   addition, a novel seven-gene signature was established to predict overall survival in LUSC,

324   which may help to provide clues in LUSC prognosis.

325

326   **Acknowledgements**

327   The authors would like to thank Dr. Shan Shan Li affiliated with CapitalBio Technology

328   Company (Beijing), for her technical assistance.

329

330

331

332   **References**

333   Al-Khafaji ASK, Marcus MW, Davies MPA, Risk JM, Shaw RJ, Field JK, Liloglou T. 2017.

334   AURKA mRNA expression is an independent predictor of poor prognosis in patients with

335   non-small cell lung cancer. Oncol Lett 13:4463-4468 DOI 10.3892/ol.2017.6012.

336   Arsic N, Bendris N, Peter M, Begon-Pescia C, Rebouissou C, Gadea G, Bouquier N, Bibeau F,

337   Lemmers B, Blanchard JM. 2012. A novel function for Cyclin A2: control of cell invasion via

338   RhoA signaling. J Cell Biol 196:147-162 DOI 10.1083/jcb.201102085.

339   Bendris N, Arsic N, Lemmers B, Blanchard JM. 2012. Cyclin A2, Rho GTPases and EMT.

340   Small GTPases 3:225-228 DOI 10.4161/sgtp.20791.

341   Bertran-Alamillo J, Cattan V, Schoumacher M, Codony-Servat J, Giménez-Capitán A, Cantero F,

342   Burbridge M, Rodríguez S, Teixidó C, Roman R, Castellví J, García-Román S, Codony-Servat

343   C, Viteri S, Cardona AF, Karachaliou N, Rosell R, Molina-Vila MA. 2019. AURKB as a

344   target in non-small cell lung cancer with acquired resistance to anti-EGFR therapy. Nat

345   Commun 10(1):1812. DOI 10.1038/s41467-019-09734-5

346   Blangy A, Lane HA, d'Herin P, Harper M, Kress M, Nigg EA. 1995. Phosphorylation by

347   p34cdc2 regulates spindle association of human Eg5, a kinesin-related motor essential for

348   bipolar spindle formation in vivo. Cell 83:1159-1169 DOI 10.1016/0092-8674(95)90142-6.

349   Bluemn EG, Spencer ES, Mecham B, Gordon RR, Coleman I, Lewinshtein D, Mostaghel E,

350   Zhang XT., Annis J, Grandori C, Porter C, Nelson PS. 2013. PPP2R2C loss promotes

351   castration-resistance and is associated with increased prostate cancer-specific mortality. Mol

352   Cancer Res 11(6): 568-78. DOI 10.1158/1541-7786.MCR-12-0710

353   Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, and Jemal A. 2018. Global cancer

354   statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in

355   185 countries. CA Cancer J Clin 68:394-424 DOI 10.3322/caac.21492.

356   Calvayrac O, Pradines A, Pons E, Mazieres J, Guibert N. 2017. Molecular biomarkers for lung

357   adenocarcinoma. Eur Respir J 49 DOI 10.1183/13993003.01734-2016.

358   Chen R, Zhang Y, Zhang C, Wu H, Yang S. 2017. miR-137 inhibits the proliferation of human

359   non-small cell lung cancer cells by targeting SRC3. Oncol Lett 13:3905-3911 DOI

360   10.3892/ol.2017.5904.

361   Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. 2014. Non-small-cell lung cancers:

362   a heterogeneous set of diseases. Nat Rev Cancer 14:535-546.

363   Choe EK, Yi JW, Chai YJ, Park KJ. 2018. Upregulation of the adipokine genes ADIPOR1 and

364   SPP1 is related to poor survival outcomes in colorectal cancer. J Surg Oncol 117(8): 1833-

365   1840. DOI 10.1002/jso.25078.

366   Fridman JS, Lowe SW. 2003. Control of apoptosis by p53. Oncogene 22:9030-9040 DOI

367   10.1038/nrc3775.

368   Gan H, Lin L, Hu N, Yang Y, Gao Y, Pei Y, Chen K, Sun B. 2019. KIF2C exerts an oncogenic

369   role in nonsmall cell lung cancer and is negatively regulated by miR-325-3p. Cell Biochem

370   Funct 37:424-431 DOI 10.1002/cbf.3420.

371   Gao T, Han Y, Yu L, Ao S, Li Z, Ji J. 2014. CCNA2 is a prognostic biomarker for ER+ breast

372   cancer and tamoxifen resistance. PLoS One 9:e91771 DOI 10.1371/journal.pone.0091771.

373   Hanahan D, and Weinberg RA. 2011. Hallmarks of cancer: the next generation. CELL 144:646-

374   674 DOI 10.1016/j.cell.2011.02.013.

375   Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ, Jr., Wu YL, Paz-Ares L. 2017.

376   Lung cancer: current therapies and new targeted treatments. Lancet 389:299-311 DOI

377   10.1016/S0140-6736(16)30958-8.

378   Hou GX, Liu PP, Yang J., Wen SJ. 2017. Mining expression and prognosis of topoisomerase

379   isoforms in non-small-cell lung cancer by using Oncomine and Kaplan-Meier plotter. PLoS

380   ONE 12(3): e0174515. DOI 10.1371/journal.pone.0174515

381   Hou Z, Xu C, Xie H, Xu H, Zhan P, Yu L, Fang X. 2014. Long noncoding RNAs expression

382   patterns associated with chemo response to cisplatin based chemotherapy in lung squamous

383   cell carcinoma patients. PLoS One 9:e108133 DOI 10.1371/journal.pone.0108133.

384   Hu CL, Wang JM, Xu YZ, Li XC, Chen HL, Bunjhoo H., Xiong WN, Xu YJ, Zhao JP. 2013.

385   Current evidence on the relationship between five polymorphisms in the matrix

386   metalloproteinases (MMP) gene and lung cancer risk: a meta-analysis. Gene 517(1):65-71

387   DOI 10.1016/j.gene.2012.12.085

388   Hwang JA, Song JS, Yu DY, Kim HR, Park HJ, Park YS, Kim WS, Choi CM. 2015.

389   Peroxiredoxin 4 as an independent prognostic marker for survival in patients with early-stage

390   lung squamous cell carcinoma. Int J Clin Exp Pathol 8:6627-6635.

391   Jin C, Zhang G, Zhang Y, Hua P, Song G, Sun M, Li X, Tong T, Li B, Zhang X. 2017.

392   Isoalantolactone induces intrinsic apoptosis through p53 signaling pathway in human lung

393   squamous carcinoma cells. PLoS One 12(8):e0181731 DOI 10.1371/journal.pone.0181731.

394   Joerger AC, and Fersht AR. 2016. The p53 Pathway: Origins, Inactivation in Cancer, and

395   Emerging Therapeutic Approaches. Annu. Rev. Biochem 85: 375-404 DOI 10.1146/annurev-

396   biochem-060815-014710

397   Kato T, Lee D, Huang H, Cruz W, Ujiie H, Fujino K, Wada H, Patel P, Hu HP, Hirohashi K,

398   Nakajima T, Sato M, Kaji M, Kaga K, Matsui Y, Chen J, Zheng G, Yasufuku K. 2018.

399   Personalized siRNA-Nanoparticle Systemic Therapy using Metastatic Lymph Node

400   Specimens Obtained with EBUS-TBNA in Lung Cancer. Mol Cancer Res 16:47-57

401   DOI 10.1158/1541-7786.MCR-16-0341.

402   Ko E, Kim Y, Cho EY, Han J, Shim YM, Park J, Kim DH. 2013. Synergistic effect of Bcl-2 and

403   cyclin A2 on adverse recurrence-free survival in stage I non-small cell lung cancer. Ann Surg

404   Oncol 20:1005-1012 DOI 10.1245/s10434-012-2727-2.

405   Kuo CN, Liao YM, Kuo LN, Tsai HJ, Chang WC, Yen Y. 2019. Cancers in Taiwan: Practical

406   insight from epidemiology, treatments, biomarkers, and cost. J Formos Med Assoc S0929-

407   6646(19):30018-X DOI 10.1016/j.jfma.2019.08.023.

408   Kwon M, Lee SJ, Wang Y, Rybak Y, Luna A, Reddy S, Adem A, Beaty BT, Condeelis JS,

409   Libutti SK. 2014. Filamin A interacting protein 1-like inhibits WNT signaling and MMP

410   expression to suppress cancer cell invasion and metastasis. Int J Cancer 135(1):48-60 DOI

411   10.1002/ijc.28662.

412   Li S, Sun X, Miao S, Liu J, Jiao W. 2017. Differential protein-coding gene and long noncoding

413   RNA expression in smoking-related lung squamous cell carcinoma. Thorac Cancer 8:672-681

414   DOI 10.1111/1759-7714.12510.

415   Liao W, Liu W, Yuan Q, Liu X, Ou Y, He S, Yuan S, Qin L, Chen Q, Nong K, Mei M, Huang J.

416   2013. Silencing of DLGAP5 by siRNA significantly inhibits the proliferation and invasion of

417   hepatocellular carcinoma cells. PLoS One 8:e80789 DOI 10.1371/journal.pone.0080789.

418   Liao Y, Yin G, Wang X, Zhong P, Fan X, Huang C. 2019. Identification of candidate genes

419   associated with the pathogenesis of small cell lung cancer via integrated bioinformatics

420   analysis. Oncol Lett 18:3723-3733 DOI 10.3892/ol.2019.10685.

421   Luo P, Wang Q, Ye YY., Zhang J, Lu DP, Cheng LQ, Zhou HC, Xie MG, Wang BL. 2019.

422   MiR-223-3p functions as a tumor suppressor in lung squamous cell carcinoma by miR-223-

423   3p-mutant p53 regulatory feedback loop. J Exp Clin Cancer Res 38(1):74 DOI

424   10.1186/s13046-019-1079-1.

425   Lv FZ, Wang JL, Wu Y, Chen HF, Shen XY. 2015. Knockdown of MMP12 inhibits the growth

426   and invasion of lung adenocarcinoma cells. Int J Immunopathol Pharmacol 28(1): 77-84 DOI

427   10.1177/0394632015572557.

428   Ma YH, Cang SD, Li GQ, Su Y, Zhang HF, Wang LM., Yang JP, Shi XY, Qin GJ, Yuan HJ.

429   2019. Integrated analysis of transcriptome data revealed MMP3 and MMP13 as critical genes

430   in anaplastic thyroid cancer progression. J Cell Physiol 234(12): 22260-22271 DOI

431   10.1002/jcp.28793.

432   MacDermed DM, Khodarev NN, Pitroda SP, Edwards DC, Pelizzari CA, Huang L, Kufe DW,

433   Weichselbaum RR. 2010. MUC1-associated proliferation signature predicts outcomes in lung

434   adenocarcinoma patients. BMC Med Genomics 3:16 DOI 10.1186/1755-8794-3-16.

435   Morra F, Luise C, Visconti R, Staibano S, Merolla F, Ilardi G, Guggino G, Paladino S, Sarnataro

436   D, Franco R, Monaco R, Zitomarino F, Pacelli R, Monaco G, Rocco G, Cerrato A,

437   Linardopoulos S, Muller MT, Celetti A. 2015. New therapeutic perspectives in CCDC6

438   deficient lung cancer cells. Int J Cancer 136:2146-2157 DOI 10.1002/ijc.29263.

439   Piao JJ, Zhu LH, Sun J, Li N, Dong B, Yang Y, Chen LY. 2019. High expression of CDK1 and

440   BUB1 predicts poor prognosis of pancreatic ductal adenocarcinoma. Gene 701:15-22 DOI

441   10.1016/j.gene.2019.02.081.

442   Qi L, Gao C, Feng F, Zhang T, Yao Y, Wang X, Liu C, Li J, Li J, Sun C. 2019. MicroRNAs

443   associated with lung squamous cell carcinoma: New prognostic biomarkers and therapeutic

444   targets. J Cell Biochem 120:18956-18966 DOI 10.1002/jcb.29216.

445 Qin SB, Long X, Zhao Q, Zhao WX. 2020. Co-Expression Network Analysis Identified Genes
446   Associated with Cancer Stem Cell Characteristics in Lung Squamous Cell Carcinoma. Cancer
447   Invest. 38(1):13-22 DOI 10.1080/07357907.2019.1697281.
448 Schneider MA, Christopoulos P, Muley T, Warth A, Klingmueller U, Thomas M, Herth FJ,
449   Dienemann H, Mueller NS, Theis F, Meister M. 2017. AURKA, DLGAP5, TPX2, KIF11 and
450   CKAP5: Five specific mitosis-associated genes correlate with poor prognosis for non-small
451   cell lung cancer patients. Int J Oncol 50:365-372 DOI 10.3892/ijo.2017.3834.
452 Stegh AH. 2012. Targeting the p53 signaling pathway in cancer therapy - the promises,
453   challenges and perils. Expert Opin Ther Targets 16(1): 67-83 DOI
454   10.1517/14728222.2011.643299.
455 Sun KX, Zheng RS, Zeng HM, Zhang SW, Zou XN, Gu XY, Xia CF, Yang ZX, Li H, Chen
456   WQ, He J. 2018. The incidence and mortality of lung cancer in China, 2014. Zhonghua Zhong
457   Liu Za Zhi 40:805-811 DOI 10.3760/cma.j.issn.0253-3766.2018.11.002.
458 Sun L, Li J, Yan B. 2015. Gene expression profiling analysis of osteosarcoma cell lines. Mol
459   Med Rep 12:4266-4272 DOI 10.3892/mmr.2015.3958.
460 Vogelstein B, Lane D, Levine AJ, Levine AJ. 2000. Surfing the p53 network. Nature (London)
461   408:307–310. DOI 10.1038/35042675
462 Wang X, Zhang F, Yang X, Xue M, Li X, Gao Y, Liu L. 2019. Secreted Phosphoprotein 1 (SPP1)
463   Contributes to Second-Generation EGFR Tyrosine Kinase Inhibitor Resistance in Non-Small
464   Cell Lung Cancer. Oncol Res 27(8):871-877 DOI 10.3727/096504018X15426271404407.
465 Wang X, Zhang F, Yang X, Xue M, Li X, Gao Y, Liu L. 2019. Secreted Phosphoprotein 1 (SPP1)
466   Contributes to Second-Generation EGFR Tyrosine Kinase Inhibitor Resistance in Non-Small
467   Cell Lung Cancer. Oncol Res 27(8): 871-877 DOI 10.3727/096504018X15426271404407.
468 Wang Z, Katsaros D, Shen Y, Fu Y, Canuto EM, Benedetto C, Lu L, Chu WM, Risch HA, Yu H.
469   2015. Biological and Clinical Significance of MAD2L1 and BUB1, Genes Frequently
470   Appearing in Expression Signatures for Breast Cancer Prognosis. PLoS One 10:e0136246
471   DOI 10.1371/journal.pone.0136246.
472 Xie B, Wang SY, Jiang N, Li JJ. 2019. Cyclin B1/CDK1-regulated mitochondrial bioenergetics
473   in cell cycle progression and tumor resistance. Cancer Lett 443: 56-66 DOI
474   10.1016/j.canlet.2018.11.019.

475 Xu J, Shu Y, Xu T, Zhu W, Qiu T, Li J, Zhang M, Xu J, Guo R, Lu K, Zhu L, Yin Y, Gu Y, Liu
476   L, Liu P, Wang R. 2018. Microarray expression profiling and bioinformatics analysis of
477   circular RNA expression in lung squamous cell carcinoma. Am J Transl Res 10:771-783.

478 Xu J, Shu Y, Xu T, Zhu W, Qiu T, Li J, Zhang M, Xu J, Guo R, Lu K, Zhu L, Yin Y, Gu Y, Liu
479   L, Liu P, Wang R. 2018. Microarray expression profiling and bioinformatics analysis of
480   circular RNA expression in lung squamous cell carcinoma. Am J Transl Res 10:771-783.

481 Yang D, Cheng DM, Tu Q, Yang HH, Sun B, Yan LZ, Dai HJ, Luo JR, Mao BY, Cao Y, Yu XP,
482   Jiang H, Zhao XD. 2018. HUWE1 controls the development of non-small cell lung cancer
483   through down-regulation of p53. Theranostics 8(13):3517-3529 DOI 10.7150/thno.24401.

484 Yeo MK, Choi SY, Seong IO, Suh KS, Kim JM, Kim KH. 2017. Association of PD-L1
485   expression and PD-L1 gene polymorphism with poor prognosis in lung adenocarcinoma and
486   squamous cell carcinoma. Hum Pathol 68:103-111 DOI 10.1016/j.humpath.2017.08.016.

487 Yu H. 2006. Structural activation of Mad2 in the mitotic spindle checkpoint: the two-state Mad2
488   model versus the Mad2 template model. J Cell Biol 173:153-157 DOI 10.1083/jcb.200601172.

489 Zeng B, Zhou M, Wu H, Xiong Z. 2018. SPP1 promotes ovarian cancer progression via Integrin
490   β1/FAK/AKT signaling pathway. Onco Targets Ther 11:1333-1343 DOI
491   10.2147/OTT.S154215. eCollection 2018.

492 Zhang Y, Wang H, Wang J, Bao L, Wang L, Huo J, Wang X. 2015. Global analysis of
493   chromosome 1 genes among patients with lung adenocarcinoma, squamous carcinoma, large-
494   cell carcinoma, small-cell carcinoma, or non-cancer. Cancer Metastasis Rev 34(2):249-64 DOI
495   10.1007/s10555-015-9558-0.

496 Zhou W, Yin M, Cui H, Wang N, Zhao LL, Yuan LZ, Yang XP, Ding XM, Men FZ, Ma X, Na
497   JR. 2015. Identification of potential therapeutic target genes and mechanisms in non-small-cell
498   lung carcinoma in non-smoking women based on bioinformatics analysis. Eur Rev Med
499   Pharmacol Sci 19:3375-3384.

500

Significantly enriched GO terms and KEGG pathways of DEGs.

1   **Table 1: Significantly enriched GO terms and KEGG pathways of DEGs.**

Category	Term	Description	Count	FDR
BP	GO:0007067	mitotic nuclear division	27	8.64E-15
BP	GO:0051301	cell division	28	4.98E-12
BP	GO:0007062	sister chromatid cohesion	15	1.60E-08
BP	GO:0007059	chromosome segregation	12	4.09E-07
BP	GO:0000082	G1/S transition of mitotic cell cycle	12	3.40E-05
BP	GO:0006260	DNA replication	11	0.01728675
BP	GO:0008283	cell proliferation	16	0.029764593
BP	GO:0000086	G2/M transition of mitotic cell cycle	10	0.041628856
CC	GO:0000777	condensed chromosome kinetochore	14	9.87E-09
CC	GO:0000775	chromosome, centromeric region	11	5.29E-07
CC	GO:0030496	midbody	14	1.54E-06
CC	GO:0000776	kinetochore	10	2.51E-04
CC	GO:0005819	spindle	11	8.59E-04
CC	GO:0005578	proteinaceous extracellular matrix	15	0.00144238
CC	GO:0005654	nucleoplasm	55	0.006300472
CC	GO:0005829	cytosol	61	0.01409558
MF	GO:0003777	microtubule motor activity	10	2.36E-04
KEGG	hsa04110	Cell cycle	12	1.59E-05
KEGG	hsa04115	p53 signaling pathway	7	0.037570687

2   **Abbreviations:** BP, biological process; CC, cellular component; MF, molecular function; DEG, differentially

3   expressed gene; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; FDR, adjust *P*

4   value

Prognostic values for the seven-genes prognostic signature in 488 LUSC patients.

2   **Table 2 Prognostic values of the seven-genes prognostic signature in 488 LUSC patients**

id	coef	HR	HR.95L	HR.95H	*P* value
FLRT3	0.034099635	1.034687693	1.004943783	1.065311951	0.021943406
PPP2R2C	0.033850795	1.034430253	1.000444568	1.069570452	0.047028752
MMP3	0.005741905	1.005758421	1.000626781	1.010916378	0.027803799
MMP12	0.003057819	1.003062498	1.000893676	1.00523602	0.005626057
CAPN8	0.051756734	1.053119523	1.013625297	1.094152576	0.007956628
FILIP1	0.132257889	1.141402636	1.038057415	1.255036532	0.006308144
SPP1	0.0002455	1.00024553	1.000013897	1.000477217	0.037749151

Univariate and multivariate Cox regression analysis of OS in LUSC.

Clinical feature	Univariate analysis			Multivariate analysis		
	HR	95%CI	*P*-value	HR	95%CI	*P*-value
Risk score(high/low)	1.655	(1.431-1.913)	<0.001	1.642	(1.412-1.412)	<0.001
age(year)	1.024	(1.004-1.045)	0.02	1.028	(1.006-1.049)	0.011
gender(male/female)	1.263	(0.863-1.849)	0.229	1.289	(0.878-1.890)	0.195
stage(IV, III/II, I)	1.261	(1.044-1.524)	0.016	0.982	(0.636-1.518)	0.936
T(T4, T3/T2, T1)	1.311	(1.059-1.623)	0.013	1.272	(0.925-1.749)	0.138
M(M1/M0)	1.829	(0.581-5.761)	0.302	1.816	(0.455-7.253)	0.399
N(N3, N2, N1/N0)	1.159	(0.93-1.442)	0.189	1.182	(0.802-1.742)	0.398

Table 3 Univariate and multivariate Cox regression analysis of OS in LUSC

Figure 1

Volcano plots (A-C) and Venn diagrams (D-E).

(A-C): The DEGs of GSE19188, GSE33479, and GSE33532. The orange dots represent DEGs filtered based on the cutoff criteria of adjusted *P*-value < 0.05 and |log2 (fold change)| ≥ 2. while the blue dots represent genes that are not satisfied the cutoff criteria. The genes with log2 (fold change) ≥ 2 were up-regulated and the genes with log2 (fold change) ≤ -2 were down-regulated. (D-E): The common up-regulated and down-regulated DEGs in the three datasets.
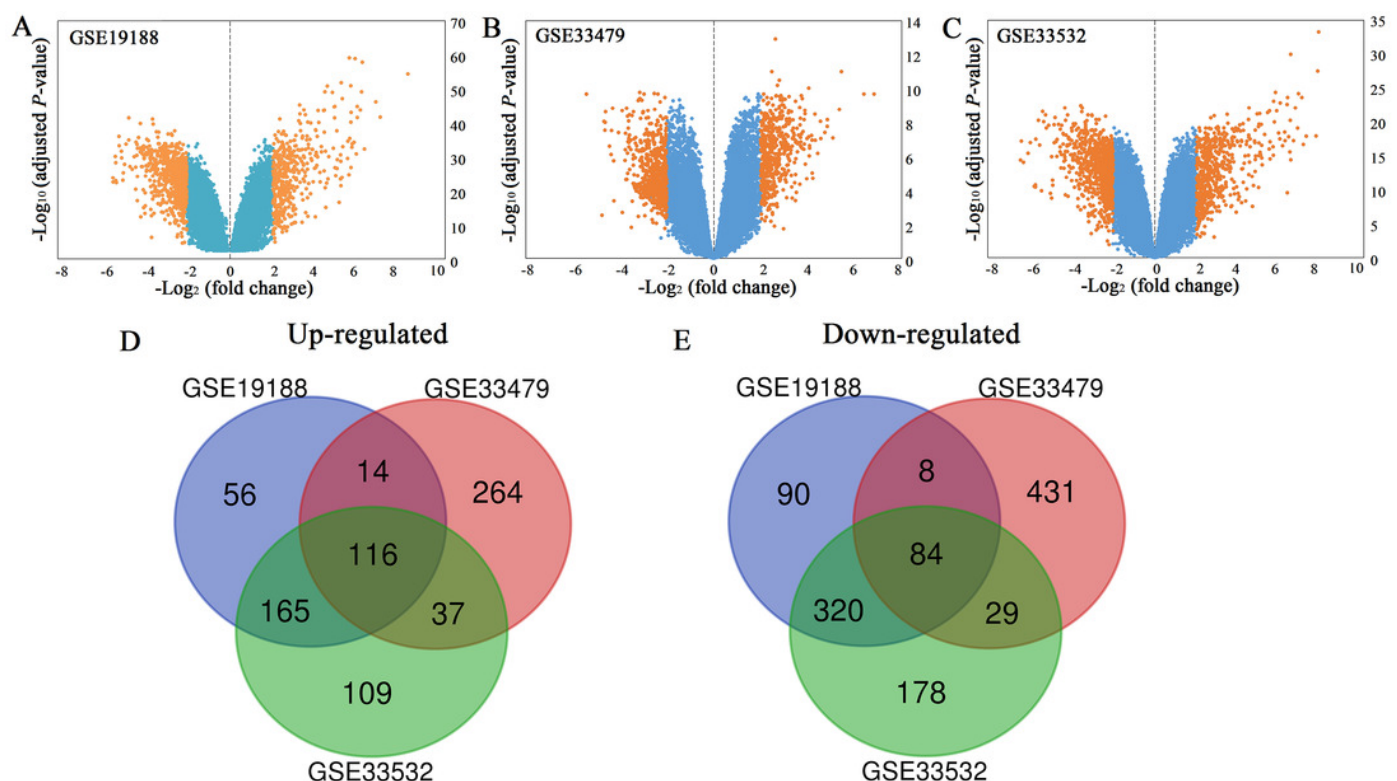
Figure 2

Enriched GO terms and KEGG pathways of the DEGs.

MF: molecular function; CC: cellular component; BP: biological process. Adjusted *P* value
(FDR) < 0.05 was considered significantly.
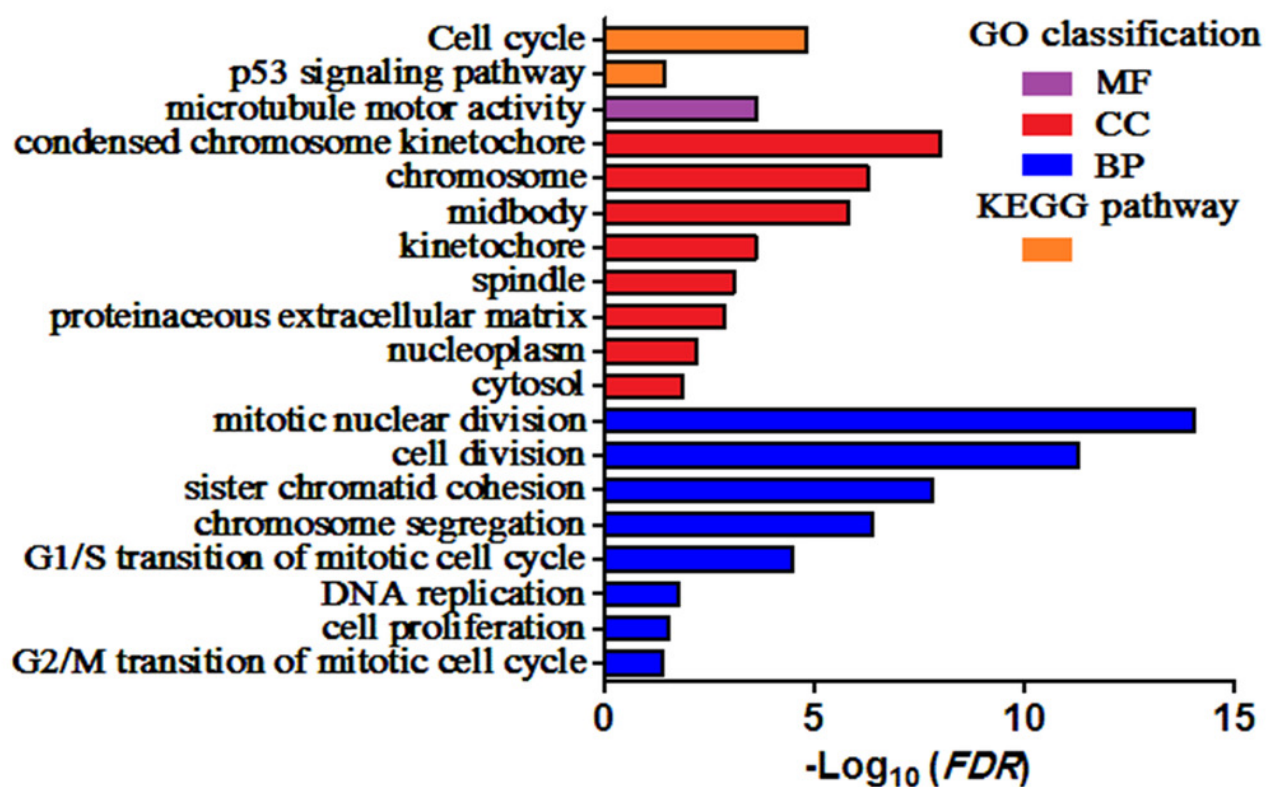
Figure 3

Protein-protein interaction network of DEGs.

The red nodes represent the up-regulated DEGs. The blue nodes represent the down-regulated DEGs.

Figure 4

A PPI network for the hub genes identified with the Cytoscape.

The top ten hub genes were ranked by the degree, and advanced ranking is represented by a redder color.
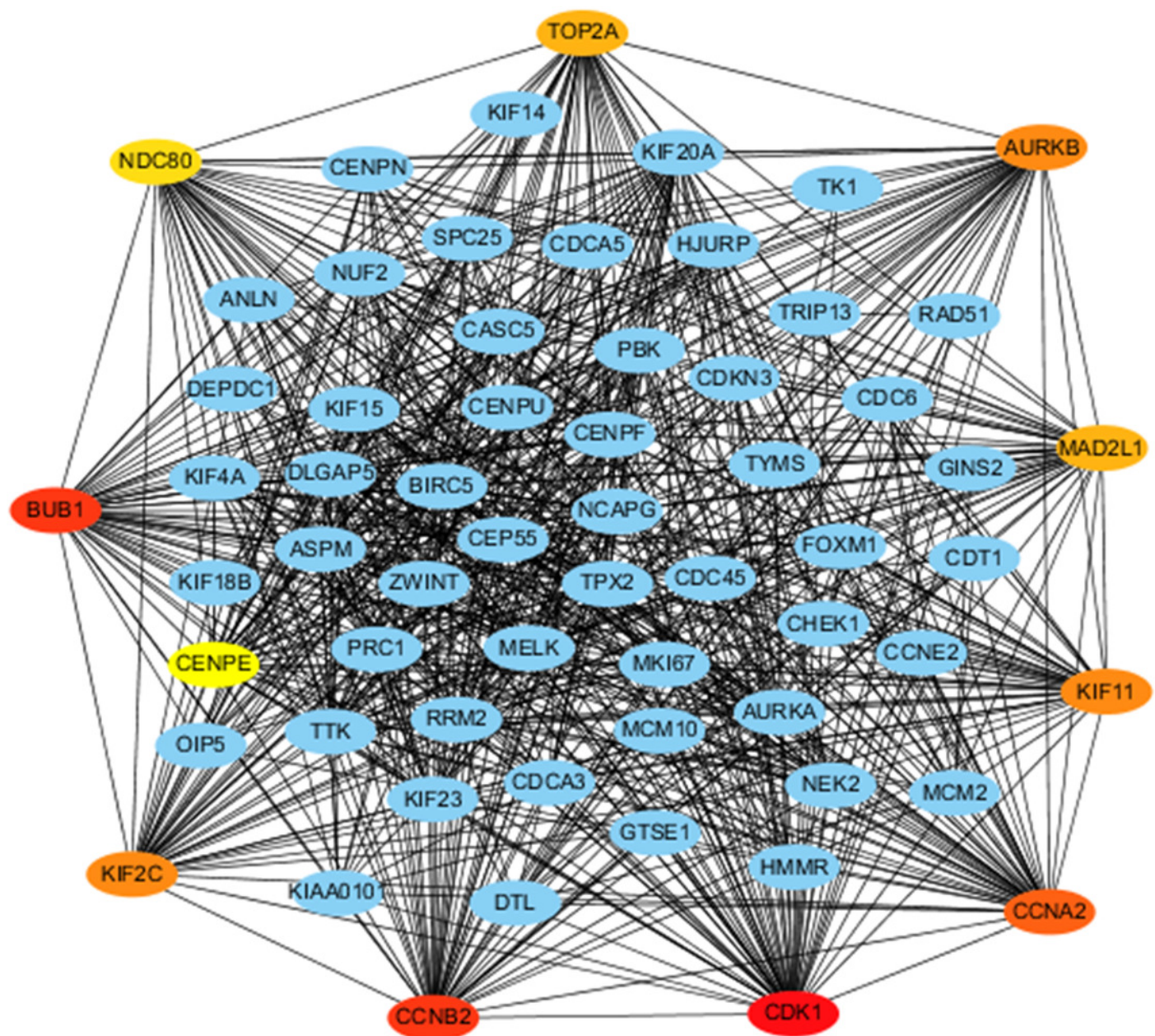
Figure 5

The expression levels of 10 hub genes in the TCGA database.

The expression of the ten hub genes in 502 LUSC tissues and 49 normal tissues from TCGA database. Gene expression values are $\log_2$-transformed.
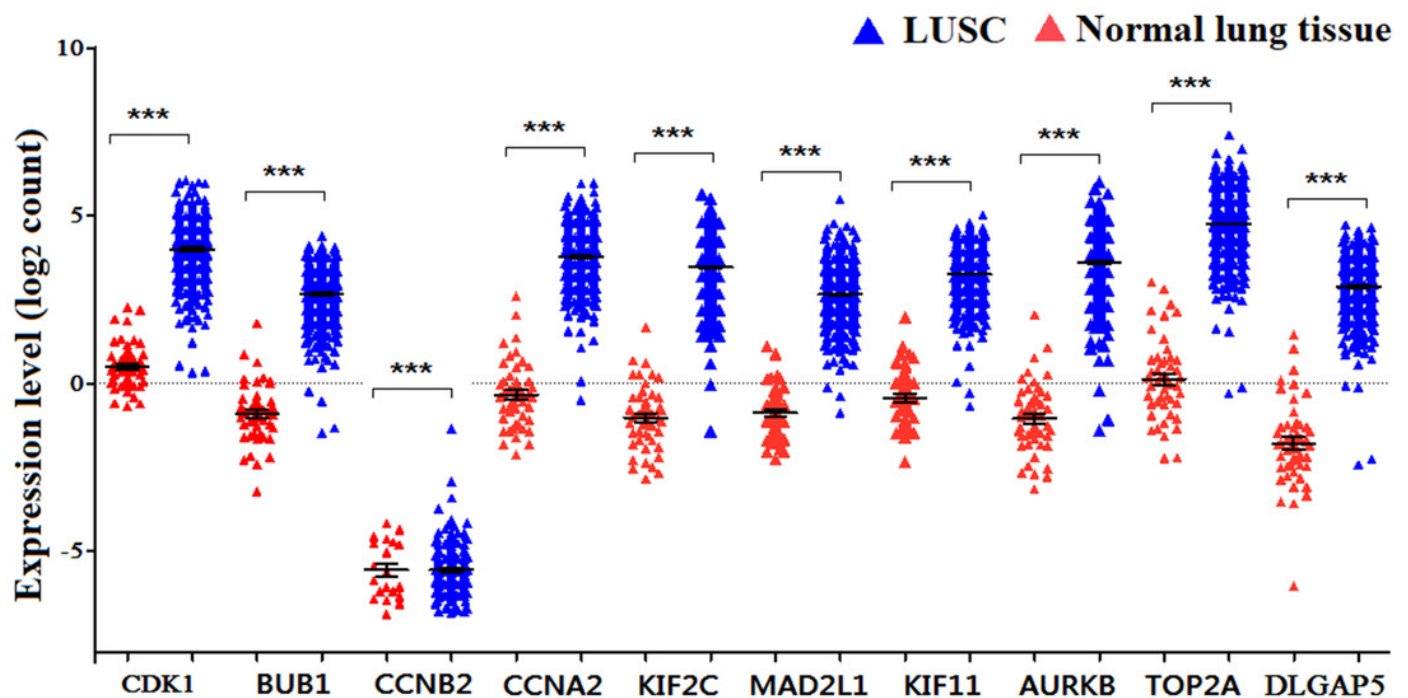
Figure 6

The expression levels of 10 hub genes between different subtypes.

The expression of the ten hub genes in LUSC was significantly different from that in other subtypes, which was analyzed using Oncomine database. Gene expression values are $\log_2$-transformed.
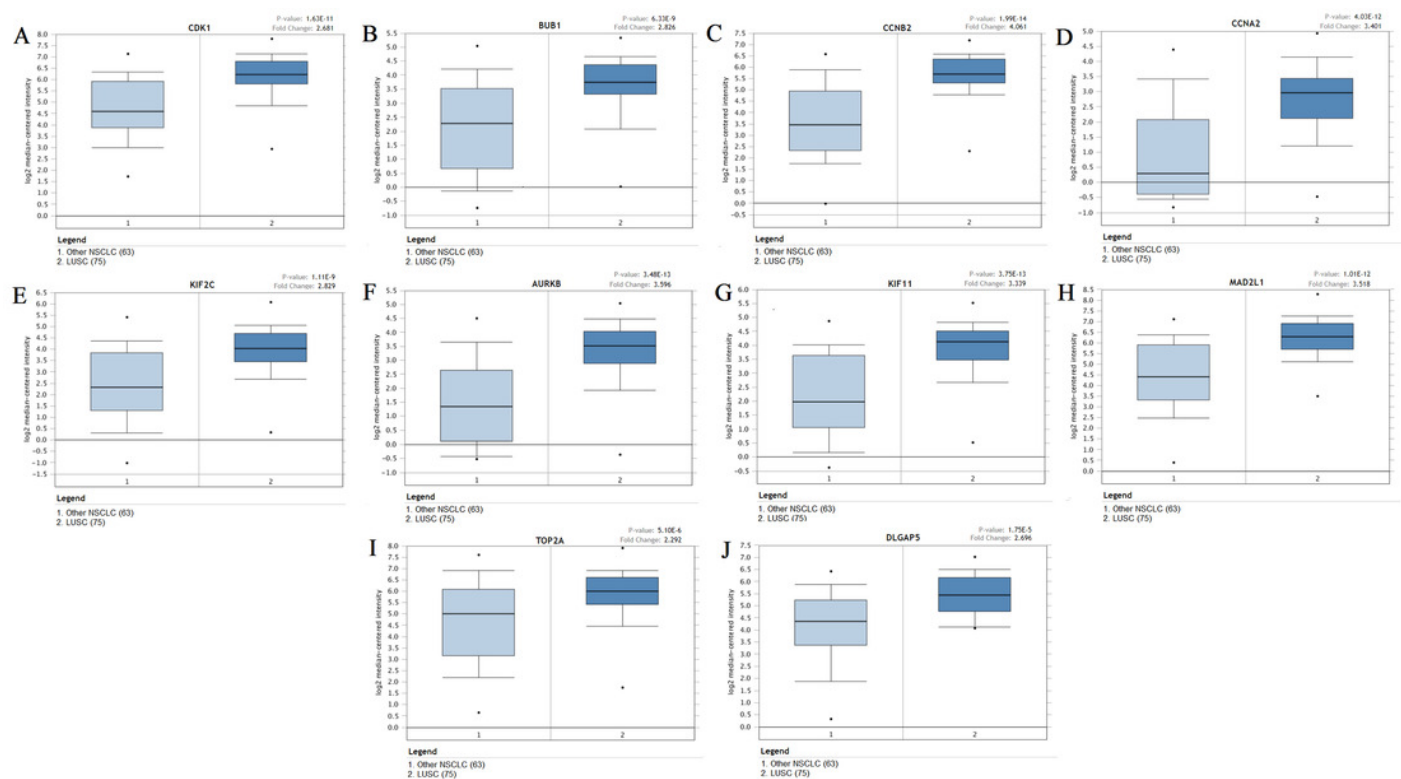
Figure 7

Kaplan–Meier survival analyses of hub genes in LUSC patients.

(A-D) The first progression survival analyses of the hub genes (CCNA2, KIF11, MAD2L1, and DLGAP5) were performed. (E) The post progression survival analyses of the hub gene (KIF2C) were performed. Logrank *P* < 0.05 was considered statistically significant.
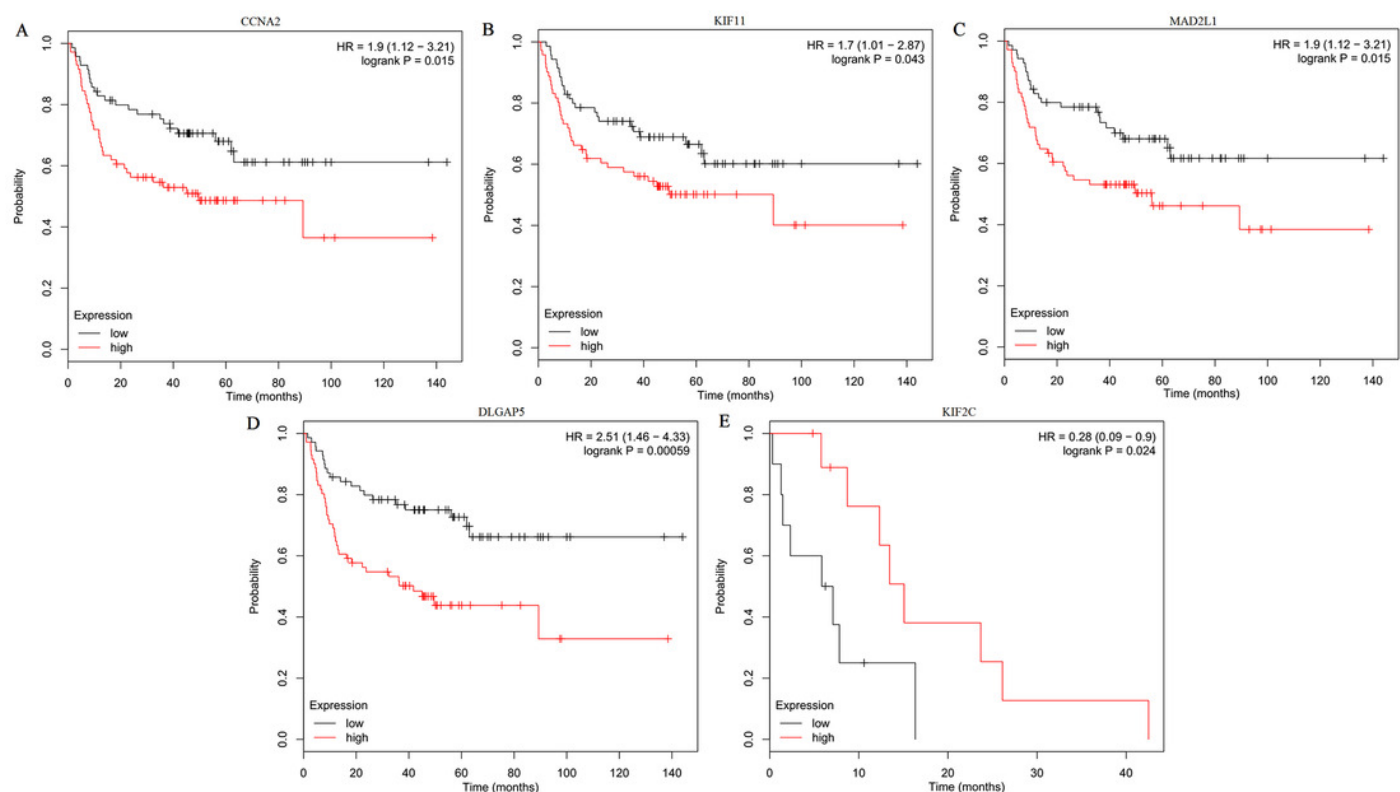
Figure 8

The association between the expression of prognosis-related hub genes and tumor stage (A-D).

In the Lee Lung dataset, the altered CCNA2, DLGAP5, MAD2L1 and KIF2C were associated with tumor grade in the LUSC progression. 0: No value; 1: Grade 1: high differentiation; 2: Grade 2: middle differentiation; Grade 3: poor differentiation.
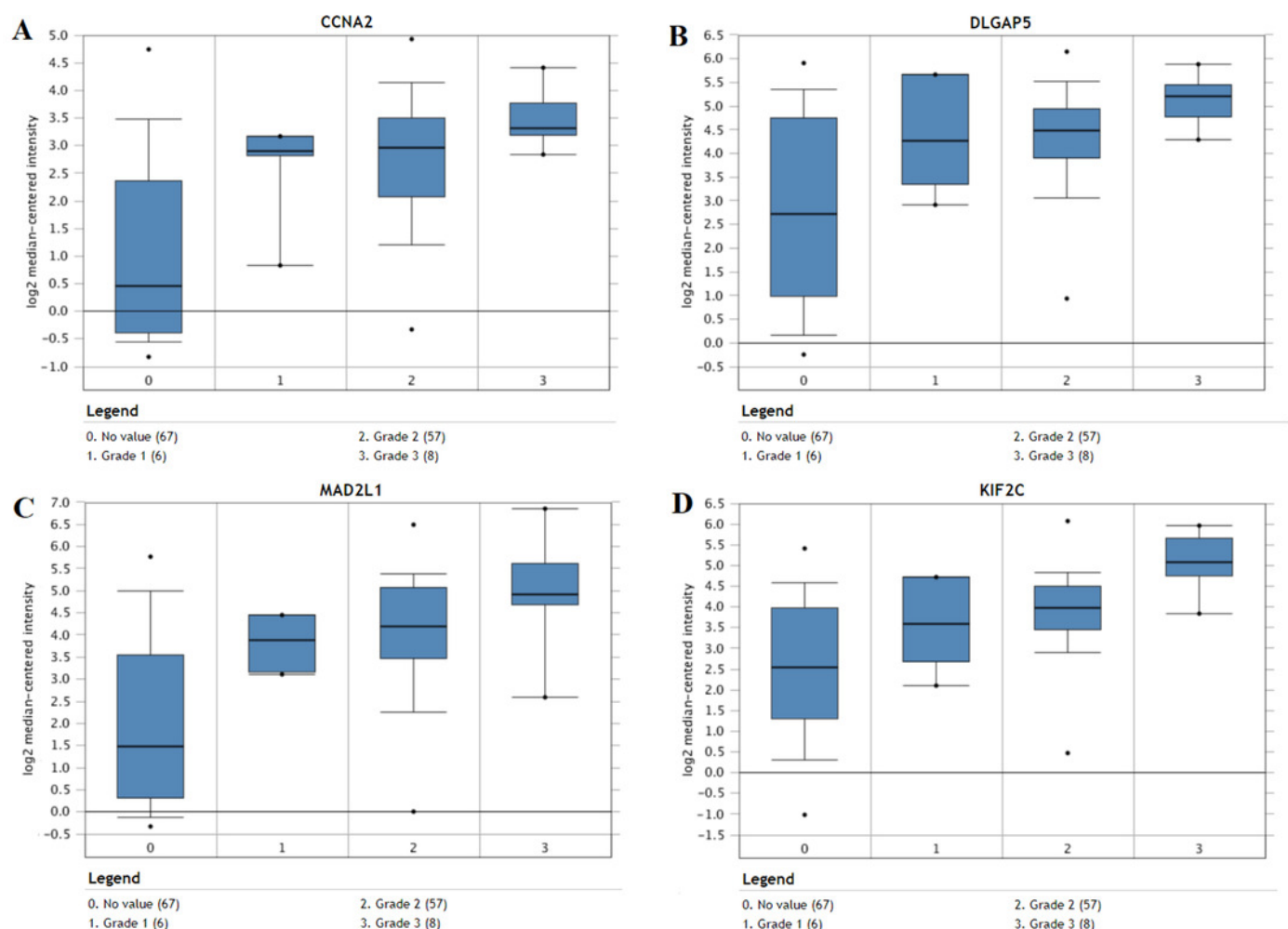
Figure 9

Prognosticgene signature of the seven genes in 488 LUSC patients.

(A) Risk score distribution. (B) Patients' survival status distribution. (C) Expression levels of the seven genes in low and high-risk groups (TCGA database).Gene expression values are log2-transformed. (D) The survival curves of LUSC patients in high- and low-risk groups. (E) time-dependent ROC curves for predicting OS in LUSC patients by the risk score.