Two predicted models based on ceRNAs and immune cells in lung adenocarcinoma

3 Miaomiao Zhang[1],Yuan Wang[1]

4 [1]The First Affiliated Hospital of Guangzhou Medical University, Guangzhou Institute of

5 Respiratory Diseases, State Key Laboratory of Respiratory Disease, 151 Yanjiang

6 Road,Guangzhou 510120, China

8 Correspondence Author:Yuan Wang

9 The First Affiliated Hospital of Guangzhou Medical University, Guangzhou Institute of

10 Respiratory Diseases, State Key Laboratory of Respiratory Disease, 151 Yanjiang Road,

11 Guangzhou 510120,China

12 Email:YuanWangxz1992@126.com

15 **Keywords** lung adenocarcinoma, immune microenvironment, prognosis,ceRNA (competitive

16 endogenous RNAs).

17   **Running Title:** TWO PREDICTED MODELS in LUAD

18   # Abstract

19   **Background** It is well accepted that both competitive endogenous RNAs (ceRNAs) and immune

20   microenvironment exert crucial roles in the tumor prognosis. The present study aimed to find

21   prognostic ceRNAs and immune cells in lung adenocarcinoma (LUAD). **Materials and Methods**

22   More specifically, we explored the associations of crucial ceRNAs with the immune

23   microenvironment. The cancer genome atlas (TCGA) database was employed to obtain expression

24   profiles of ceRNAs and clinical data. CIBERSORT was utilized to quantify the proportion of 22

25   immune cells in LUAD. **Results** We constructed two cox regression models based on crucial

26   ceRNAs and immune cells to predict prognosis in LUAD. Subsequently, seven ceRNAs and seven

27   immune cells were involved in prognostic models. We validated both predicted models via an

28   independent cohort GSE72094. Interestingly, both predicted models proved that the longer patients

29   were smoking, the higher risk scores would be obtained. We further investigated the relationships

30   between seven genes and immune/stromal scores via the ESTIMATE algorithm. The results

31   indicated that CDC14A and H1F0 expression were significantly related to stromal scores/Immune

32   scores in LUAD. Moreover, based on the result of the ceRNA model, single-sample gene set

33   enrichment analysis (ssGSEA) suggested that differences in immune status were evident between

34   high- and low-risk groups.

Introduction

37 Lung cancer is considered as the most common cause of survival-associated cancer globally(Bray

38 et al. 2018). Besides, lung adenocarcinoma (LUAD) approximately accounts for 50% of all lung

39 cancer types (Imielinski et al. 2012). Although the improvement of diagnosis and the

40 comprehensive treatment in the lung cancer has been made, the 5-year overall survival rate for

41 lung cancer is still approximately 18%, mainly due to invisible early symptoms (Siegel et al. 2018).

42 Thus, it is of great necessity to identify sensitive and accurate biomarkers in the early stage.

43 Competitive endogenous RNAs (ceRNAs), consisting of a group of regulatory RNA molecules,

44 bind with the specific miRNAs and then affect the protein level (Kartha & Subramanian 2014;

45 Phelps et al. 2016; Shukla et al. 2011). Previous studies have shown that ceRNA regulatory

46 networks play essential roles in tumor progression and migration (Bartel 2009; Qi et al. 2015;

47 Sanchez-Mejias & Tay 2015; Tay et al. 2014). Besides, increasing studies have demonstrated that

48 immune microenvironment (TIME) was also associated with the prognosis in various kinds of

49 tumors, including LUAD (Lavin et al. 2017; Liu et al. 2020a). Stromal cells and immune cells

50 constitute the central part of TIME. Stromal cells participate in the process of tumorigenesis and

51 metastasis and are generally conducive to the movement of cancer cells (Ni et al. 2018).

52 Additionally, various studies revealed that immune cells exhibit large effects on clinical outcomes

53 in many cancers. Donnem et al. summarized that different types of immune cells affected the

54 clinical stage and prognosis in non-small lung cancer (Donnem et al. 2016; Liu et al. 2020b). As a

55 result, to further investigate potential biomarkers based on immune cells and ceRNAs in LUAD,

56   we constructed Cox proportional hazards models and predicted nomograms.

57   However, the mechanisms of ceRNAs in the tumor microenvironment still remain unclear in

58   LUAD. It is well accepted that tumor stem cells (CSCs) are tumorigenic cells with self-renewal

59   ability. Besides, CSCs promote tumor immune to escape via expressing a variety of immune

60   factors (Bruttel & Wischhusen 2014). In the current study, RNA stem scores and DNA stem scores

61   were applied to assess tumor stem cell content (Malta et al. 2018). ESTIMATE algorithm was

62   utilized to evaluate the stromal and immune cells in every sample. Thus, we explored the

63   correlation between crucial prognostic genes with RNAss/DNAss scores and stromal/immune

64   scores.

Materials and Methods

65

Identification of differentially expressed ceRNAs

66

67   In the present study, expression profiles of LUAD and the corresponding clinical information were

68   derived from the TCGA database, including lncRNA,miRNA, and mRNA. The R package

69   "DEGseq2" was used to screen the differentially expressed mRNAs,miRNAs and lncRNAs

70   (DEmRNA, DEmiRNA and DElncRNA) between normal and tumor samples (Love et al.

71   2014).Specifically, a false discovery rate (FDR) <0.05 and log2 |fold change | > 1 were set as

72   cutoff criteria.

ceRNA network construction

73

74   Both lncRNA-miRNA and miRNA-mRNA interactions were predicted using, miRcode and

75   starBase(Chou et al. 2018; Jeggari et al. 2012; Li et al. 2014). DEmiRNAs, DEmRNAs, and DE

76   lncRNAs, which presented a significant difference in hypergeometric distribution detection and

77   correlation analysis (P<0.05), were selected to construct a lncRNA-miRNA-mRNA regulatory

78   network and then was visualized by Cytoscape 3.80 (Shannon et al. 2003).

79   **Construction of nomogram based on key ceRNAs**

80   Univariate Cox analysis was employed to screen the survival-associated ceRNAs. Lasso regression

81   analysis was applied to ensure that the Cox model was not overfitted. Eventually, the candidate

82   ceRNAs were integrated to establish the multivariate cox regression models. Risk scores were

83   described as following: risk score=(Exp RNA1* β 1) + (Exp RNA2* β 2) + ... + (Exp RNA3*

84   β 3). It is worth noting that Exp stands for the expression level, and β denotes the regression

85   coefficient. Based on the risk scores, patients with LUAD were divided into high-risk and low-risk

86   groups. The corresponding Kaplan-Meier survival curves were carried out to reveal the overall

87   survival (OS) within different groups. In addition, we also performed the ROC curve to evaluate

88   the specificity and sensitivity of the model. Finally, according to the multivariate model result, we

89   constructed a nomogram to predict patients' prognostic values. The calibration curve was applied

90   to evaluate the accuracy of the nomogram.

91   **Construction of nomogram based on immune cells**

92   The CIBERSORT was utilized to estimate immune cells in LUAD (Newman et al. 2015). Samples

93   with a CIBERSORT output of p < 0.05 presented that the predicted proportion of immune cells

94   were correct.

95   The Wilcoxon test was conducted to examine the significant difference of immune cells between

96   the tumor samples and healthy samples. We also performed multivariate Cox regression and

97   produced the corresponding risk scores. Based on the multivariate Cox model results, we finally

98   built a predicted nomogram. Similar to ceRNAs, patients were also stratified into high-risk and

99   low-risk groups on the basis of the mean risk score. Subsequently, the prognostic model's

100   sensitivity and specificity were examined by the ROC curve, and the calibration curve investigated

101   the accuracy of the nomogram.

102   **Validation of prognostic equations with GEO**

103   To better elucidate the accuracy of both multivariate Cox models, an external validation dataset

104   GSE72094 (n=442) was applied in the present study. We selected this dataset because it contained

105   most LUAD patients with clinical data in the Gene Expression Omnibus (GEO) database. Based

106   on the median risk score generated from TCGA, we also classified patients with high- and low-

107   risk patients. K-M and ROC curves were applied to evaluate the prognostic efficacy of multivariate

108   models in TCGA.

109   **Estimation of tumor stemness and microenvironment**

110   Stemness scores of "RNAss" and "DNAss" referred to the result based on mRNA and DNA-

111   methylation respectively. The larger stemness scores were, the higher likelihood more stem cells

112   were infiltrating. We utilized the "ESTIMATE" package to assess immune, stromal scores, and

113   the sum of both in individual patients. The higher scores, the higher proportion of corresponding

114   cells were. Additionally, correlation of hub ceRNAs with above scores was examined by Spearman

115   analysis. By applying the "gsva" package, we calculated infiltrating level of immune cells and

116   immune-associated functions in LUAD samples (Hänzelmann et al. 2013).

117   **Statistical analysis**

118   All statistical analyses were accomplished with R software 4.0.2. The Wilcoxon rank sum tests

119   were conducted for comparisions between two groups. Only a two-sided P-value <0.05 was

120   recognized as statistically significant.

121   **Results**

122   **Identification of DEmRNAs,DEmiRNAs and DElncRNAs**

123   Figure1 presents our workflow for the bioinformatic analysis. Among the whole expression

124   profiles, totally1645 upregulated DEmRNAs, and 1344 downregulated DEmRNA were screened

125   out (Fig. 2A,2B). Furthermore, we also identified 111 upregulated DEmiRNA, 87 downregulated

126   DEmiRNA (Fig. 2C,2D), 163 upregulated DElncRNA, and 52 downregulated DElncRNA (Fig.

127   2E,2F).

128

129   **Construct Nomogram of ceRNAs**

130   To investigate the endogenously competitive relationships between the lncRNA and mRNA, we

131   constructed an intersected ceRNA network. Seven lncRNAs (H19, AC074117.1, FBXL19-AS1,

132   MAGI2-AS3, SNHG3, PVT1 and SNHG1), 94mRNA, and 15miRNA were involved in the

133   network (Supplementary Table.S1). Hsa-miR-130b-3p and hsa-miR-29b-3p were the top 2

134   regulated miRNA in the ceRNA network (Fig. 3). In addition, the initial univariate Cox regression

135   (Supplementary Fig. S1) and Lasso regression analysis were further conducted to investigate the

136   key biomarkers (Fig. 4B,4C). Seven survival-associated biomarkers were finally involved in a new

137   multiple Cox regression analysis (Fig. 4A). According to the result, all biomarkers belonged to

138   protein-coding RNAs. Consequently, the corresponding risk score was generated (Table 1): risk

139   score = (Exp CDC14A * 0.281888)+(Exp LOXL2*0.114969) + (Exp CCT6A * 0.174244)＋(Exp

140   E2F7 *0.187164)+(Exp GPR37*0.099737)+(Exp H1F0 *0.197590)+(Exp SMOC1 * 0.068385).

141   As shown in Fig.4D, the prognostic score could be a promising indicator to distinguish the LUAD

142   patients according to the individual prognosis (P=5.701e-8). The area under ROC reflected the

143   predictive efficacy of the prognostic signature (AUC of 3-year survival:0.772) (Fig.4E). By the

144   GEO validation set, the AUC of gene signature at 3-year was 0.711 (Supplementary Fig. S3A).

145   Additionally, K-M analysis further confirmed the difference between high- and low-risk groups

146   (P=9.963e-05) (Supplementary Fig. S3B). Eventually, we assessed the prognosis of patients with

147   LUAD by a nomogram (Fig. 5A), and the accuracy of the nomogram was confirmed by the

148   calibration curve (Fig. 5C). As shown in Fig.5E, based on this prognostic model, the risk score in

149   the group with fewer smoking years ($\leqq$20)was lower than the group with more smoking years (>

150   20) (P=0.025).

Construct Nomogram of immune Cells

151

152   By applying the CIBERSORT, we estimated the percentages of 22 immune cell types in LUAD

153   (Supplementary Fig. S2B,2C) (Chen et al. 2018). The violin plot indicated that 13 immune cell

154   types showed a significant difference in percentages of immune cells between tumor samples and

155   healthy samples by the Wilcoxon test (Supplementary Fig. S2D). The correlation analysis of

156   immune cells could be found in Supplementary Fig.2A.Subsequently, after the analysis of

157   univariate cox regression and lasso regression, seven immune cells were taken as prognostic

158   biomarkers and eventually integrated into a new multivariate model (Fig. 6A,6B,6C). The formula

159   was presented as following (Table 2): risk score =   [Exp CD8 *(-3.425649)]+(Exp

160   Tregs*9.175062)+[Exp Monocytes*(-5.318492)]＋(Exp M1 *6.134551)+(Exp Dendritic cells

161   activated*5.539130)+(Exp Mast cells activated *24.611110)+(Exp Eosinophils*37.351722). The

162   K-M analysis results showed a significant difference in overall survival between high- and low-

163   risk groups (P=3.342e-05) (Fig. 6D). The ROC curve (AUC of 3-years: 0.766) demonstrated that

164   the multivariate model was recognized as a fair potential to monitor the prognostic efficacy(Fig.

165   6E). In the external validation set, the model also carried out adequate predicted capacity. The

166   AUC of immune cell signature was 0.738 at 3-years (Supplementary Fig. S3C).K-M curves

167   reflected that low-risk group had longer overall survival than the high-risk group (P=2.922e-04)

168   (Supplementary Fig. S3D). A predicted nomogram was also performed, and the discrimination

169   was conducted to test the calibration quality (Fig. 5B,5D). Similarly, we could also found that

170   more somking group had more risk score based on the predicted model of immune cells (P=0.0046)

171   (Fig.5F)

172   **Correlation of hub ceRNAs expression with tumor stemness and microenvironment**

173   As shown in Fig.7, CCT6A, H1F0 and E2F7 were positively related to RNAss and DNAss (R=0.22

174   to 0.53). A significant negative association was detected between CDC14A and RNAss (R=-0.25,

175   P=7.5e-8). Nevertheless, we did not find a significant correlation between LOXL2, GPR37 and

176   SMOC1 with tumor stemness. In addition, H1F0 displayed a negative association with stromal

177   scores, immune scores and estimate scores (R=-0.34 to -0.27). However, CDC14A showed a

178   positive association with the microenvironment-related scores (R=0.31 to 0.35).

179   **Multiple databases validation**

180   To validate CDC14A and H1F0 expression in LUAD, we utilized multiple databases. Based on

181   the result of UCLAN, GEPIA and TIMER, it could be concluded that H1F0 was highly expressed

182   while CDC14A was weakly expressed in LUAD (Supplementary Fig.4). Furthermore, the

183   CDC14A and H1F0 protein expression were obtained from The Human Protein Atlas data (HPA),

184   indicating that H1F0 was high staining in tumor cells, while it was medium staining in normal

185   tissues (Supplementary Fig.S5A, S5B). The opposite trend was observed in CDC14A. Besides,

186   the association of H1F0 and CDC14A with immune cells were validated in TISIDB. CDC14A

187   expression was positively associated with Immature B cell (R=0.345), NK cell (R=0.308),

188   Eosinophils (R=0.357), and Mast cell (R=0.332) (Supplementary Fig.S6A-S6E).

189   Controversially, H1F0 expression was negatively related to Eosinophil (R=-0.34),

190   Macrophage(R=-0.359), NKT(R=-0.323), Tem_CD8 (R=-0.323) and Th1(R=-0.407)

191   (Supplementary Fig.S6F-S6J).

Differences in ssGSEA scores in TCGA/GEO cohort

193   To further elucidate the immune status in risk- and low groups of ceRNA signatures, we performed

194   the ssGSEA algorithm to estimate the immune infiltrating between two groups (Supplementary

195   Fig.7).Of note, low-risk group were associated with higher scores of aDCs, B_cells,

196   CD8+_T_cells,iDCs, Mast_cells, Neutrophils, pDCs, T_helper_cells and TIL

197   (Fig.8A,8B).Similarly, in lower-risk group, the scores of Check-point, HLA, T_cell_co-

198   stimulation and Type_II_IFN_Reponse in still higher in the low-risk group.Additionally, the

199   differences above were confirmed by the GEO cohort (Fig. 8C,8D).

Discussion

201   With the reduced overall survival rate, lung adenocarcinoma is the most common pathological

202   type since most patients were diagnosed at an advanced stage. Currently, there is a general

203   consensus that both molecular and cellular characteristics play an imposing role in oncogenesis

204   and metastasis, and thus they were suggested as underlying prognostic signatures. However, the

205   integrative analysis of ceRNAs and microenvironments has yet to be fully explored. In the present

206   study, we constructed nomogram survival models of both ceRNAs and immune cells.   GSE72094

207   dataset further proved the feasibility of predicted equations. Besides, we stratified TCGA patients

208   with smoking history into two groups and calculated risk scores in each sample via both predicted

209 formulas, respectively. Not unexpectedly, both predicted formulas confirmed that the more

210 extended smoking group (smoking years >20) have a higher risk than the shorter smoking group

211 (smoking years <=20). Thus, it could be concluded that smoking was an adverse factor for

212 prognosis in both models.

213   E2F7, a member of EF transcription factors, was known as a transcriptional repressor that

214 influences the prognosis in a lot of tumors, such as breast cancer, endometrial carcinoma,

215 gallbladder cancer, and colon cancer. Previous studies have proved that miR-30a-5p prevents

216 tumor migration and metastasis by targeting E2F7 in gallbladder cancer (Ye et al. 2018). Liu et al.

217 have revealed that E2F7 was reported significantly overexpressed in ER-positive breast cancer

218 than healthy breast tissues and then led to tamoxifen resistance in breast cancer cells (Liu et al.

219 2018). Additionally, overexpressed E2F7 could be suggested as a significant biomarker to identify

220 the high-risk and low-risk groups of patients with lung cancer, which is also in consistence with

221 our report (Sun et al. 2018).

222   SMOC1 is a cancer-associated protein, identified as an extracellular glycoprotein of the

223 SPARC-related modular calcium-binding protein family. SMOC1 was overexpressed in brain

224 cancer, including oligodendrogliomas, astrocytomas, and glioblastomas (Brellier et al. 2011;

225 Fackler et al. 2011). However, till now, there is no enough evidence for the impact of SMOC1 in

226 LUAD reported in the literature. Moreover, the current study may provide a study direction for

227 researches of SMOC1 in the LUAD. Additionally, CCTA6 is reported to encode a molecular

228 chaperone and play critical roles in damaged proteins repair, cytoskeletal organization, and the cell

229   cycle (Van Hove et al. 2015; Yam et al. 2008). Ying et al. found that CCTA6 served as a SMAD2-

230   binding protein, which could inhibit the function of SMAD2, and promote metastasis (Ying et al.

231   2017).

232   CDC14A, which is of considerable significance to regulate the actin, was reported

233   downregulated in many tumors and then reduced CDC14A was associated with the poor clinical

234   outcomes (Chen et al. 2016). Our results indicated that CDC14A was a protective biomarker in

235   LUAD (HR=0.75, P=0.0066), which corresponded well with previous studies. Similarly, based on

236   the correlation with stemness scores and immune socres, we could also draw inferences that

237   CDC14A suppressed tumor progression.

238   As one of H1 histone genes, H1F0 (known as H1.0) is heterogeneous in many tumors (Di Liegro

239   et al. 2018). Knockdown of H1.0 influenced the differentiation of embryonic stem cells (Terme et

240   al. 2011). Recent studies revealed that H1F0 were prognostic indicators in many cancers, such as

241   breast cancer, liver cancer, and kidney cancer (Torres et al. 2016).In the current study, the

242   relationships of H1F0 expression with stemness scores and Estamine scores suggested that H1F0

243   might promote LUAD.

244   T cells regulatory (Tregs) account for 5%-10% of peripheral CD4+ T cells in humans.

245   Increasing evidence demonstrated that the immunosuppression mediated by Tregs is one of the

246   essential mechanisms of immune evasion in tumors (Dunn et al. 2004; Shevach 2002; Zou 2005).

247   Mast cells are increased in several tumors, and their accumulation was associated with a low

248   survival rate in many cancers, such as pancreatic adenocarcinoma (Strouch et al. 2010) and

249   melanoma (Ribatti et al. 2003). Macrophages M1 was historically recognized as the

250   proinflammatory subgroup, expressing a series of chemokines and consequently play antitumor

251   roles (Hao et al. 2012). Controversially, our results suggested that M1 was highly expressed in

252   LUAD (Fig.8). Actually, heterogeneity of environment signals could influence TAM

253   development, and comprehensive nomenclatures of TAM have consequently been suggested

254   (Ostuni et al. 2015).

255   Additionally, our study found that there were Obviously differences in immune status

256   between high- and low risk groups in both TCGA and GEO cohorts.This indicated that decreased

257   risk scores of ceRNAs were related to the enhancement of antitumor immunity.Furthermore, we

258   could say that diminished immune function is responsible for poor prognosis in LUAD patients.

259   However, there are inevitably some disadvantages in our study that should be taken into

260   consideration. Firstly, considering the limit data from the TCGA and GEO, it could result in large

261   analysis deviation and be verified by other large cohorts. Secondly, our follow-up experiment will

262   explore the molecular mechanisms of prognostic signatures in the present study.

263   **Conclusion**

264   To conclude, the present study firstly provided a combined analysis of ceRNA and immune cells

265   and then established nomograms to predict the prognosis in LUAD reliably. Based on the present

266   study result, we will carry out the biological experiments to validate key ceRNAs and mechanisms

267   of immune cells in the future.

Acknowledgments

The authors would like to thank TCGA, GEO, TIMER,GEPIA,HPA,UCLAN and TISIDB

databases

Disclosure Statement

No competing financial interests exist

Funding Information

No funding was received for this article

References

Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* 136:215-233. 10.1016/j.cell.2009.01.002

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, and Jemal A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68:394-424. 10.3322/caac.21492

Brellier F, Ruggiero S, Zwolanek D, Martina E, Hess D, Brown-Luedi M, Hartmann U, Koch M, Merlo A, Lino M, and Chiquet-Ehrismann R. 2011. SMOC1 is a tenascin-C interacting protein over-expressed in brain tumors. *Matrix Biol* 30:225-233. 10.1016/j.matbio.2011.02.001

Bruttel VS, and Wischhusen J. 2014. Cancer stem cell immunology: key to understanding tumorigenesis and tumor immune escape? *Front Immunol* 5:360. 10.3389/fimmu.2014.00360

Chen B, Khodadoust MS, Liu CL, Newman AM, and Alizadeh AA. 2018. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol* 1711:243-259. 10.1007/978-1-4939-7493-1_12

Chen NP, Uddin B, Voit R, and Schiebel E. 2016. Human phosphatase CDC14A is recruited to the cell leading edge to regulate cell migration and adhesion. *Proc Natl Acad Sci U S A* 113:990-995. 10.1073/pnas.1515605113

Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH, Chiew MY, Tai CS, Wei TY, Tsai TR, Huang HT, Wang CY, Wu HY, Ho SY, Chen PR, Chuang CH, Hsieh PJ, Wu YS, Chen WL, Li MJ, Wu YC, Huang XY, Ng FL, Buddhakosai W, Huang PC,

297   Lan KC, Huang CY, Weng SL, Cheng YN, Liang C, Hsu WL, and Huang HD. 2018. miRTarBase
298   update 2018: a resource for experimentally validated microRNA-target interactions.
299   *Nucleic Acids Res* 46:D296-d302. 10.1093/nar/gkx1067
300   Di Liegro CM, Schiera G, and Di Liegro I. 2018. H1.0 Linker Histone as an Epigenetic Regulator of
301   Cell Proliferation and Differentiation. *Genes (Basel)* 9. 10.3390/genes9060310
302   Donnem T, Kilvaer TK, Andersen S, Richardsen E, Paulsen EE, Hald SM, Al-Saad S, Brustugun OT,
303   Helland A, Lund-Iversen M, Solberg S, Gronberg BH, Wahl SG, Helgeland L, Fløtten O, Pohl
304   M, Al-Shibli K, Sandanger TM, Pezzella F, Busund LT, and Bremnes RM. 2016. Strategies
305   for clinical implementation of TNM-Immunoscore in resected nonsmall-cell lung cancer.
306   *Ann Oncol* 27:225-232. 10.1093/annonc/mdv560
307   Dunn GP, Old LJ, and Schreiber RD. 2004. The immunobiology of cancer immunosurveillance and
308   immunoediting. *Immunity* 21:137-148. 10.1016/j.immuni.2004.07.017
309   Fackler MJ, Umbricht CB, Williams D, Argani P, Cruz LA, Merino VF, Teo WW, Zhang Z, Huang P,
310   Visvananthan K, Marks J, Ethier S, Gray JW, Wolff AC, Cope LM, and Sukumar S. 2011.
311   Genome-wide methylation analysis identifies genes specific to breast cancer hormone
312   receptor status and risk of recurrence. *Cancer Res* 71:6195-6207. 10.1158/0008-
313   5472.can-11-1630
314   Hänzelmann S, Castelo R, and Guinney J. 2013. GSVA: gene set variation analysis for microarray
315   and RNA-seq data. *BMC Bioinformatics* 14:7. 10.1186/1471-2105-14-7
316   Hao NB, Lü MH, Fan YH, Cao YL, Zhang ZR, and Yang SM. 2012. Macrophages in tumor
317   microenvironments and the progression of tumors. *Clin Dev Immunol* 2012:948098.
318   10.1155/2012/948098
319   Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti
320   M, Sivachenko A, Sougnez C, Auclair D, Lawrence MS, Stojanov P, Cibulskis K, Choi K, de
321   Waal L, Sharifnia T, Brooks A, Greulich H, Banerji S, Zander T, Seidel D, Leenders F, Ansén
322   S, Ludwig C, Engel-Riedel W, Stoelben E, Wolf J, Goparju C, Thompson K, Winckler W,
323   Kwiatkowski D, Johnson BE, Jänne PA, Miller VA, Pao W, Travis WD, Pass HI, Gabriel SB,
324   Lander ES, Thomas RK, Garraway LA, Getz G, and Meyerson M. 2012. Mapping the
325   hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 150:1107-
326   1120. 10.1016/j.cell.2012.08.029
327   Jeggari A, Marks DS, and Larsson E. 2012. miRcode: a map of putative microRNA target sites in
328   the long non-coding transcriptome. *Bioinformatics (Oxford, England)* 28:2062-2063.
329   10.1093/bioinformatics/bts344
330   Kartha RV, and Subramanian S. 2014. Competing endogenous RNAs (ceRNAs): new entrants to
331   the intricacies of gene regulation. *Frontiers in genetics* 5:8-8. 10.3389/fgene.2014.00008
332   Lavin Y, Kobayashi S, Leader A, Amir ED, Elefant N, Bigenwald C, Remark R, Sweeney R, Becker
333   CD, Levine JH, Meinhof K, Chow A, Kim-Shulze S, Wolf A, Medaglia C, Li H, Rytlewski JA,
334   Emerson RO, Solovyov A, Greenbaum BD, Sanders C, Vignali M, Beasley MB, Flores R,
335   Gnjatic S, Pe'er D, Rahman A, Amit I, and Merad M. 2017. Innate Immune Landscape in
336   Early Lung Adenocarcinoma by Paired Single-Cell Analyses. *Cell* 169:750-765.e717.
337   10.1016/j.cell.2017.04.014

338 Li J-H, Liu S, Zhou H, Qu L-H, and Yang J-H. 2014. starBase v2.0: decoding miRNA-ceRNA, miRNA-
339   ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic*
340   *Acids Res* 42:D92-D97. 10.1093/nar/gkt1248
341 Liu J, Li X, Wang M, Xiao G, Yang G, Wang H, Li Y, Sun X, Qin S, Du N, Ren H, and Pang Y. 2018. A
342   miR-26a/E2F7 feedback loop contributes to tamoxifen resistance in ER-positive breast
343   cancer. *Int J Oncol* 53:1601-1612. 10.3892/ijo.2018.4492
344 Liu X, Niu X, and Qiu Z. 2020a. A Five-Gene Signature Based on Stromal/Immune Scores in the
345   Tumor Microenvironment and Its Clinical Implications for Liver Cancer. *DNA Cell Biol*
346   39:1621-1638. 10.1089/dna.2020.5512
347 Liu Z, Wan Y, Qiu Y, Qi X, Yang M, Huang J, and Zhang Q. 2020b. Development and validation of a
348   novel immune-related prognostic model in lung squamous cell carcinoma. *Int J Med Sci*
349   17:1393-1405. 10.7150/ijms.47301
350 Love MI, Huber W, and Anders S. 2014. Moderated estimation of fold change and dispersion for
351   RNA-seq data with DESeq2. *Genome Biol* 15:550. 10.1186/s13059-014-0550-8
352 Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kamińska B, Huelsken
353   J, Omberg L, Gevaert O, Colaprico A, Czerwińska P, Mazurek S, Mishra L, Heyn H, Krasnitz
354   A, Godwin AK, Lazar AJ, Stuart JM, Hoadley KA, Laird PW, Noushmehr H, and Wiznerowicz
355   M. 2018. Machine Learning Identifies Stemness Features Associated with Oncogenic
356   Dedifferentiation. *Cell* 173:338-354.e315. 10.1016/j.cell.2018.03.034
357 Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, and Alizadeh AA.
358   2015. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*
359   12:453-457. 10.1038/nmeth.3337
360 Ni SJ, Zhao LQ, Wang XF, Wu ZH, Hua RX, Wan CH, Zhang JY, Zhang XW, Huang MZ, Gan L, Sun
361   HL, Dimri GP, and Guo WJ. 2018. CBX7 regulates stem cell-like properties of gastric cancer
362   cells via p16 and AKT-NF-κB-miR-21 pathways. *J Hematol Oncol* 11:17. 10.1186/s13045-
363   018-0562-z
364 Ostuni R, Kratochvill F, Murray PJ, and Natoli G. 2015. Macrophages and cancer: from
365   mechanisms to therapeutic implications. *Trends Immunol* 36:229-239.
366   10.1016/j.it.2015.02.004
367 Phelps M, Coss C, Wang H, and Cook M. 2016. Registered report: Coding-independent regulation
368   of the tumor suppressor PTEN by competing endogenous mRNAs. *Elife* 5.
369   10.7554/eLife.12470
370 Qi X, Zhang DH, Wu N, Xiao JH, Wang X, and Ma W. 2015. ceRNA in cancer: possible functions
371   and clinical implications. *J Med Genet* 52:710-718. 10.1136/jmedgenet-2015-103334
372 Ribatti D, Ennas MG, Vacca A, Ferreli F, Nico B, Orru S, and Sirigu P. 2003. Tumor vascularity and
373   tryptase-positive mast cells correlate with a poor prognosis in melanoma. *Eur J Clin Invest*
374   33:420-425. 10.1046/j.1365-2362.2003.01152.x
375 Sanchez-Mejias A, and Tay Y. 2015. Competing endogenous RNA networks: tying the essential
376   knots for cancer biology and therapeutics. *J Hematol Oncol* 8:30. 10.1186/s13045-015-
377   0129-1
378 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker

379   T. 2003. Cytoscape: a software environment for integrated models of biomolecular
380   interaction networks. *Genome Res* 13:2498-2504. 10.1101/gr.1239303
381 Shevach EM. 2002. CD4+ CD25+ suppressor T cells: more questions than answers. *Nat Rev*
382   *Immunol* 2:389-400. 10.1038/nri821
383 Shukla GC, Singh J, and Barik S. 2011. MicroRNAs: Processing, Maturation, Target Recognition
384   and Regulatory Functions. *Mol Cell Pharmacol* 3:83-92.
385 Siegel RL, Miller KD, and Jemal A. 2018. Cancer statistics, 2018. *CA Cancer J Clin* 68:7-30.
386   10.3322/caac.21442
387 Strouch MJ, Cheon EC, Salabat MR, Krantz SB, Gounaris E, Melstrom LG, Dangi-Garimella S, Wang
388   E, Munshi HG, Khazaie K, and Bentrem DJ. 2010. Crosstalk between mast cells and
389   pancreatic cancer cells contributes to pancreatic tumor progression. *Clin Cancer Res*
390   16:2257-2265. 10.1158/1078-0432.ccr-09-1230
391 Sun CC, Zhou Q, Hu W, Li SJ, Zhang F, Chen ZL, Li G, Bi ZY, Bi YY, Gong FY, Bo T, Yuan ZP, Hu WD,
392   Zhan BT, Zhang Q, Tang QZ, and Li DJ. 2018. Transcriptional E2F1/2/5/8 as potential
393   targets and transcriptional E2F3/6/7 as new biomarkers for the prognosis of human lung
394   carcinoma. *Aging (Albany NY)* 10:973-987. 10.18632/aging.101441
395 Tay Y, Rinn J, and Pandolfi PP. 2014. The multilayered complexity of ceRNA crosstalk and
396   competition. *Nature* 505:344-352. 10.1038/nature12986
397 Terme JM, Sesé B, Millán-Ariño L, Mayor R, Izpisúa Belmonte JC, Barrero MJ, and Jordan A. 2011.
398   Histone H1 variants are differentially expressed and incorporated into chromatin during
399   differentiation and reprogramming to pluripotency. *J Biol Chem* 286:35347-35357.
400   10.1074/jbc.M111.281923
401 Torres CM, Biran A, Burney MJ, Patel H, Henser-Brownhill T, Cohen AS, Li Y, Ben-Hamo R, Nye E,
402   Spencer-Dene B, Chakravarty P, Efroni S, Matthews N, Misteli T, Meshorer E, and Scaffidi
403   P. 2016. The linker histone H1.0 generates epigenetic and functional intratumor
404   heterogeneity. *Science* 353. 10.1126/science.aaf1644
405 Van Hove I, Verslegers M, Hu TT, Carden M, Arckens L, and Moons L. 2015. A proteomic approach
406   to understand MMP-3-driven developmental processes in the postnatal cerebellum:
407   Chaperonin CCT6A and MAP kinase as contributing factors. *Dev Neurobiol* 75:1033-1048.
408   10.1002/dneu.22272
409 Yam AY, Xia Y, Lin HT, Burlingame A, Gerstein M, and Frydman J. 2008. Defining the TRiC/CCT
410   interactome links chaperonin function to stabilization of newly made proteins with
411   complex topologies. *Nat Struct Mol Biol* 15:1255-1262. 10.1038/nsmb.1515
412 Ye YY, Mei JW, Xiang SS, Li HF, Ma Q, Song XL, Wang Z, Zhang YC, Liu YC, Jin YP, Hu YP, Jiang L, Liu
413   FT, Zhang YJ, Hao YJ, and Liu YB. 2018. MicroRNA-30a-5p inhibits gallbladder cancer cell
414   proliferation, migration and metastasis by targeting E2F7. *Cell Death Dis* 9:410.
415   10.1038/s41419-018-0444-x
416 Ying Z, Tian H, Li Y, Lian R, Li W, Wu S, Zhang HZ, Wu J, Liu L, Song J, Guan H, Cai J, Zhu X, Li J, and
417   Li M. 2017. CCT6A suppresses SMAD2 and promotes prometastatic TGF-β signaling. *J Clin*
418   *Invest* 127:1725-1740. 10.1172/jci90439
419 Zou W. 2005. Immunosuppressive networks in the tumour environment and their therapeutic

420   relevance. *Nat Rev Cancer* 5:263-274. 10.1038/nrc1586

421

422

Figure 1

Figure 1 The workflow for bioinformatic analysis
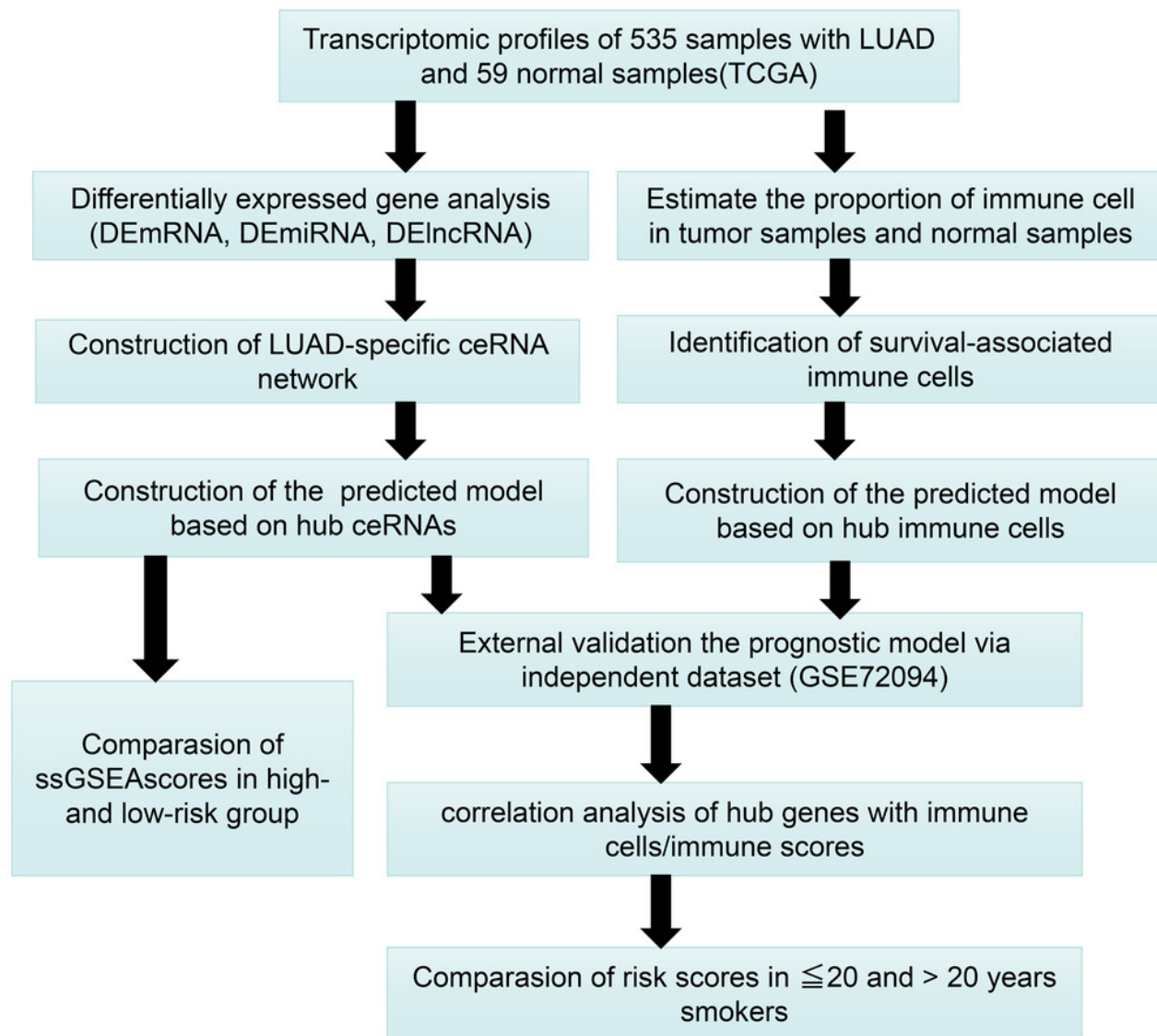
Figure 2

Figure 2 Differentially expressed genes.

Heatmap (**A**) and volcano plot (**B**) of DEmRNA.Heatmap (**C**) and volcano plot (**D**) of DEmiRNA.Heatmap (**E**) and volcano plot (**F**) of DElncRNA. Red dots and green dots represent up -regulated, down-regulated differentially expressed genes, respectively. DEmRNA, DEmiRNA and DElncRNA represent differentially expressed mRNA,miRNA and lncRNA, respectively.
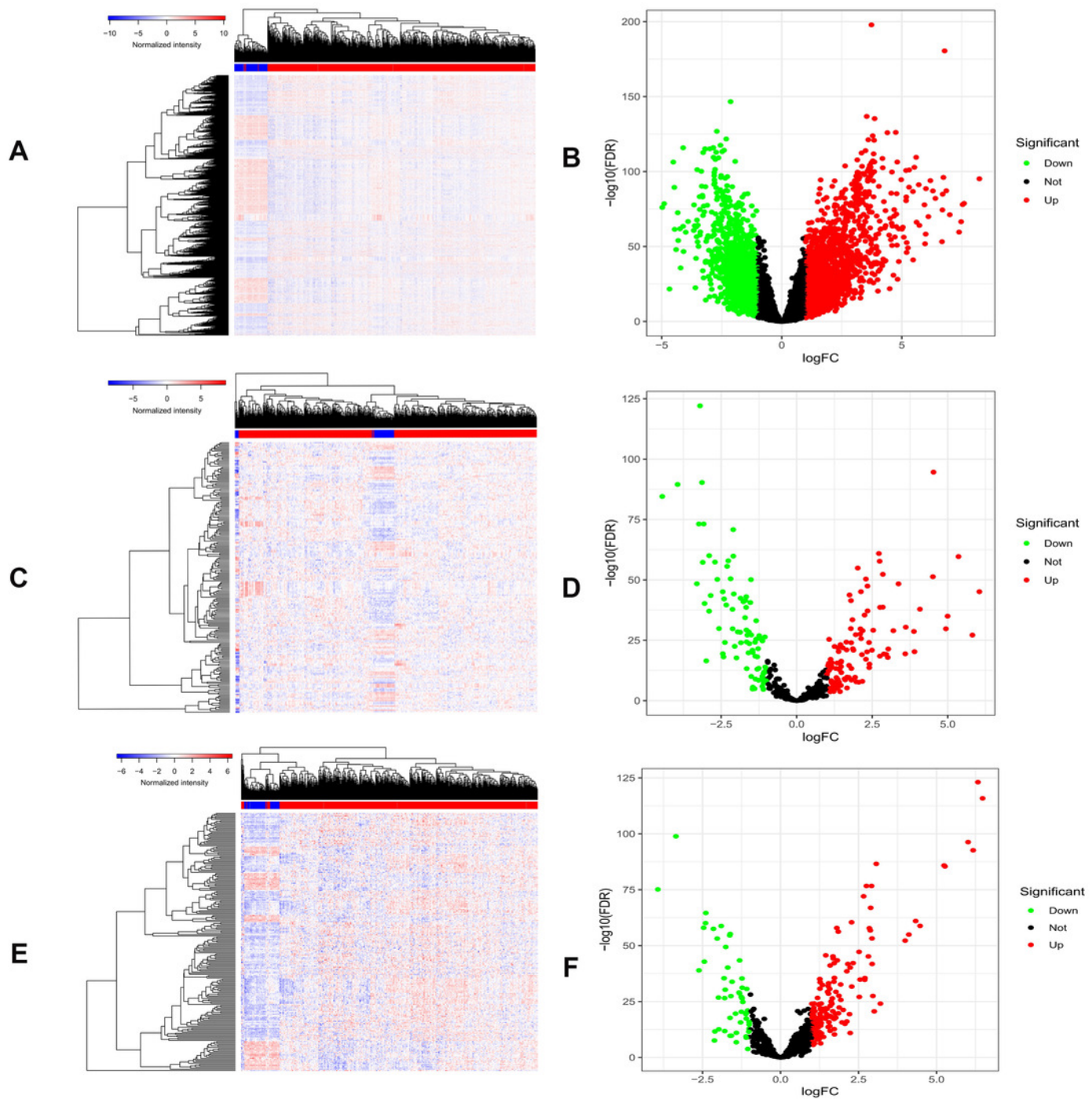
Figure 3

Figure 3 Construction of ceRNA network in LUAD.LUAD, lung adenocarcinoma. ceRNA, competing endogenous RNA.
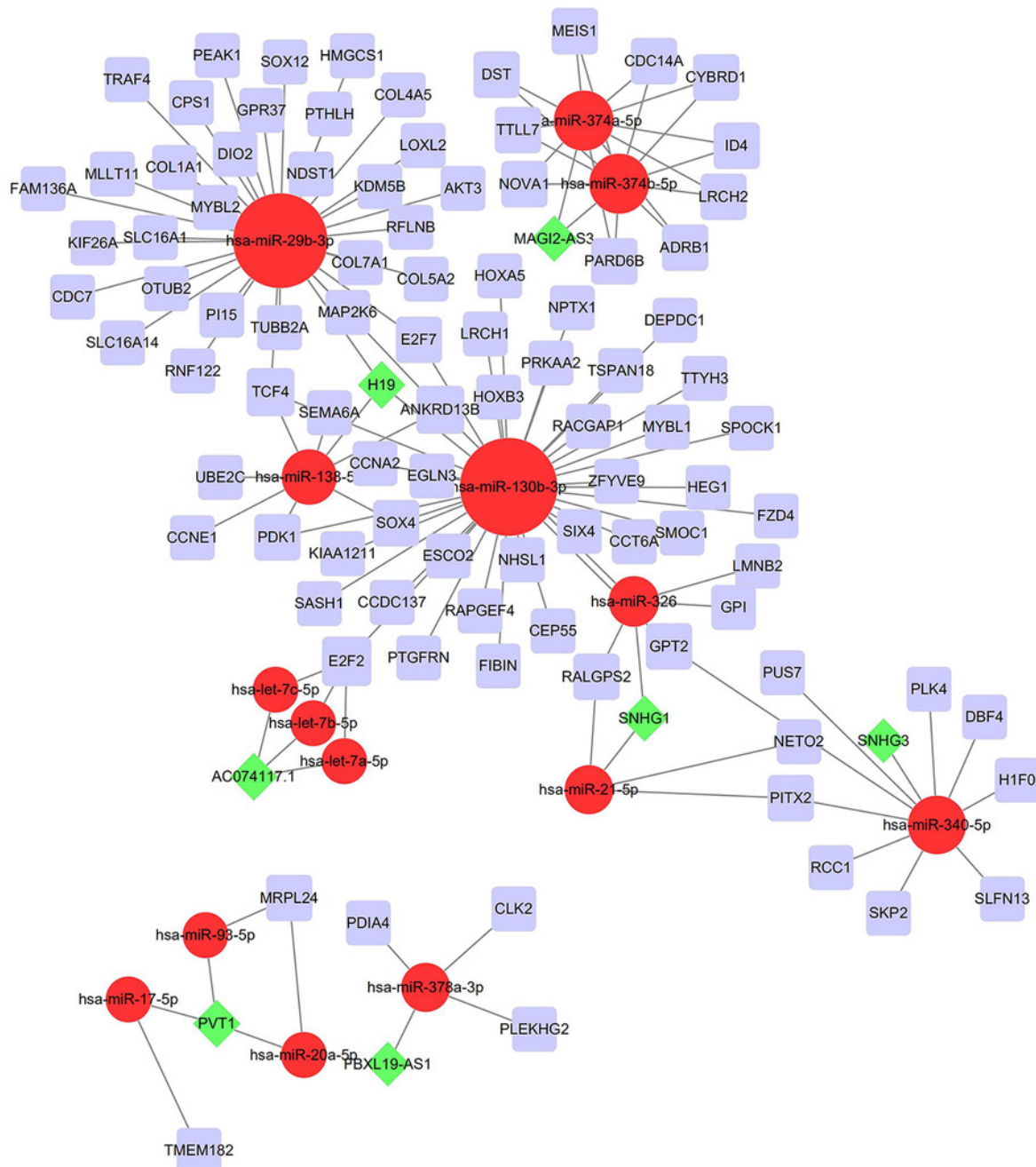
Figure 4

Figure 4 Construction of prognostic model involving the key RNAs of ceRNA network in the LUAD.

(**A**), multivariate Cox regression. (**B**), (**C**) lasso regression.(**D**) the Kaplan–Meier survival curve. (**E**) the ROC curve. ROC,the Receiver Operating Characteristic. LUAD, lung adenocarcinoma.
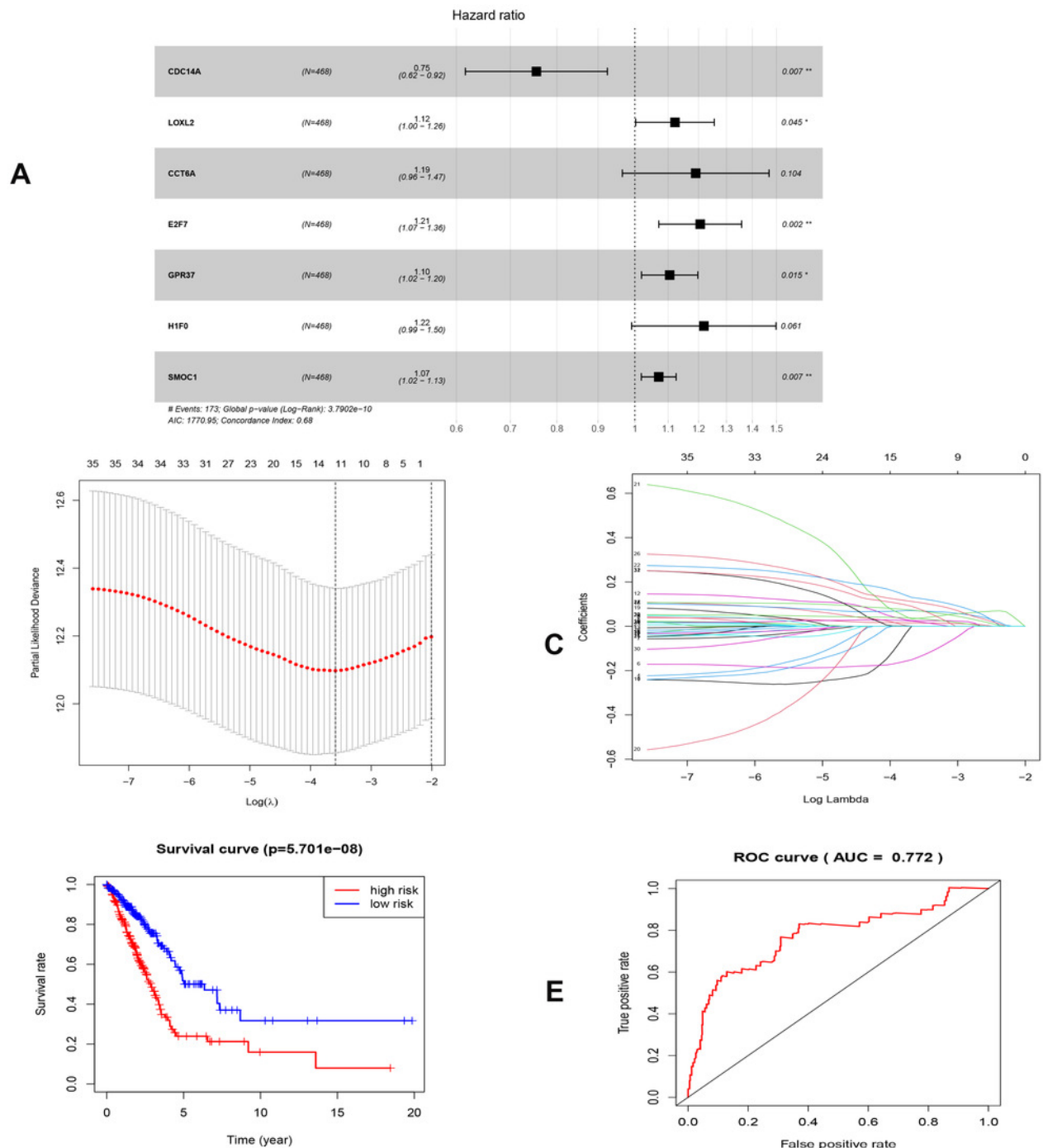
Figure 5

Figure 5 Nomograms based on key ceRNAs and immune cells.

Nomograms predicting 1-,2- and 3-year overall survival of ceRNAs signatures (**A**) and immune cells signatures (**B**).Calibration curves of the nomogram at the 3 year of ceRNAs (**C**) signatures and immune cells signatures (**D**). Differences in risk scores between >20 smoking years and ≤ 20 smoking years based on ceRNAs (**E**) and immune cells models (**F**).
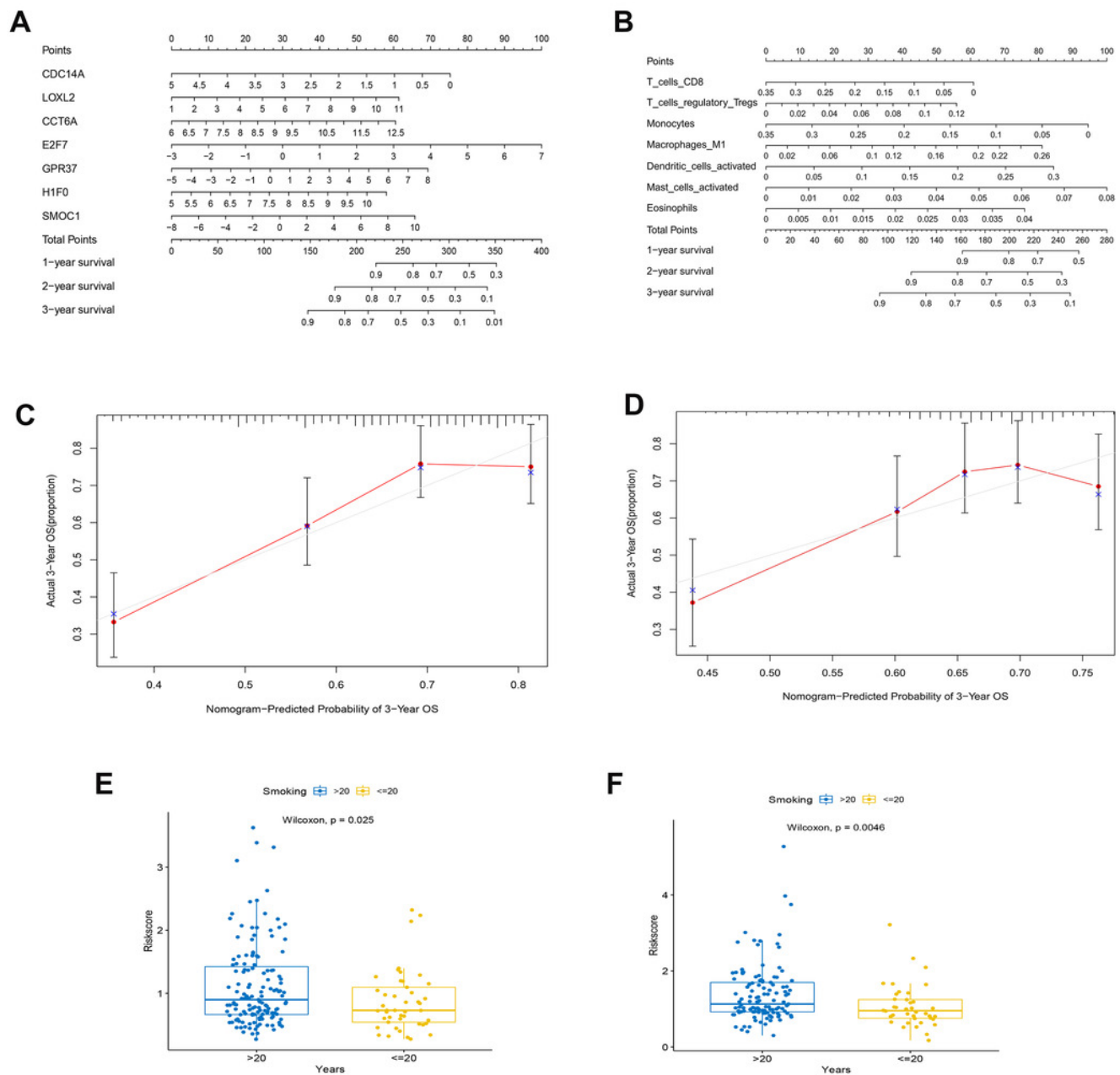
Figure 6

Figure 6 Construction of prognostic model based on key immune cells.

(**A**), multivariate Cox regression. (**B**), (**C**) lasso regression.(**D**) the Kaplan–Meier survival curve. (**E**) the ROC curve. ROC,the Receiver Operating Characteristic.
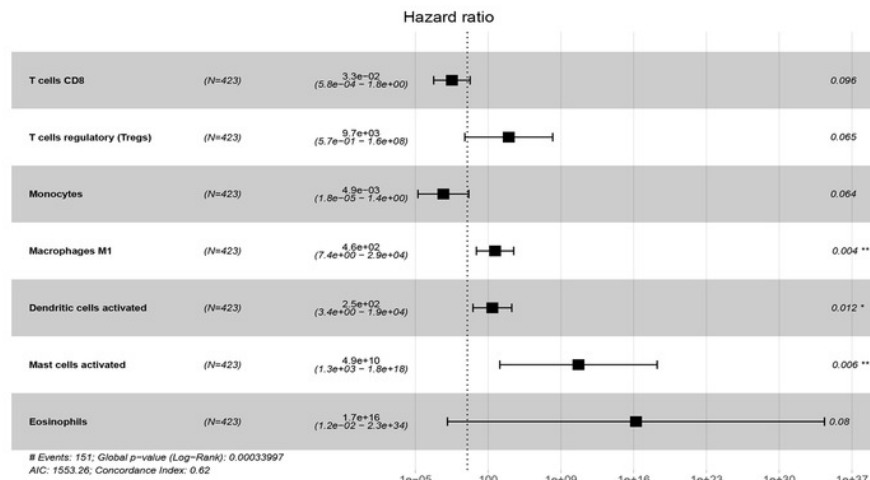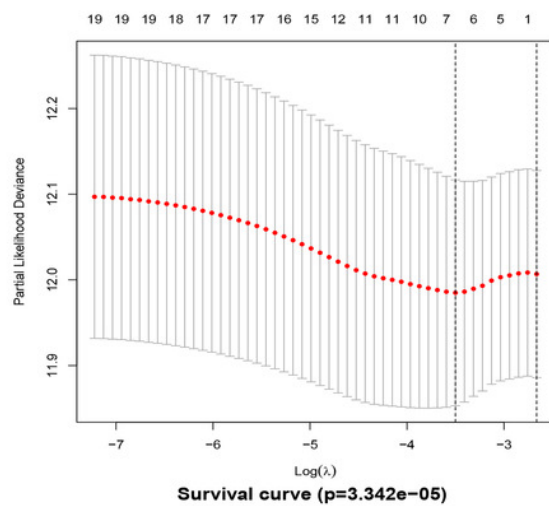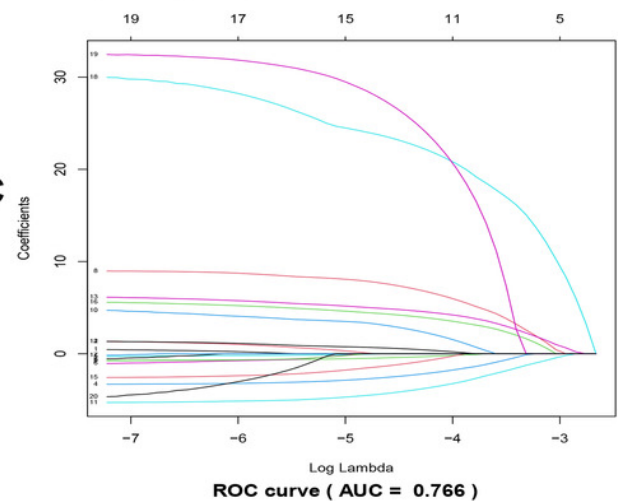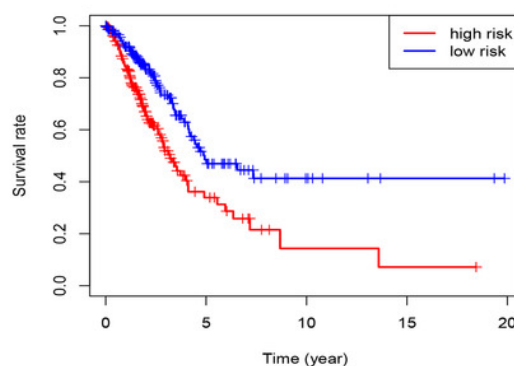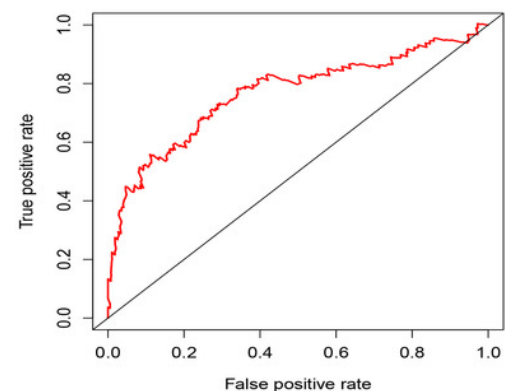
Figure 7

Figure 7 Correlation between crucial genes and RNAss/DNAss score, stromal score, immune score, and Estimate Score.

Cancer: LUAD
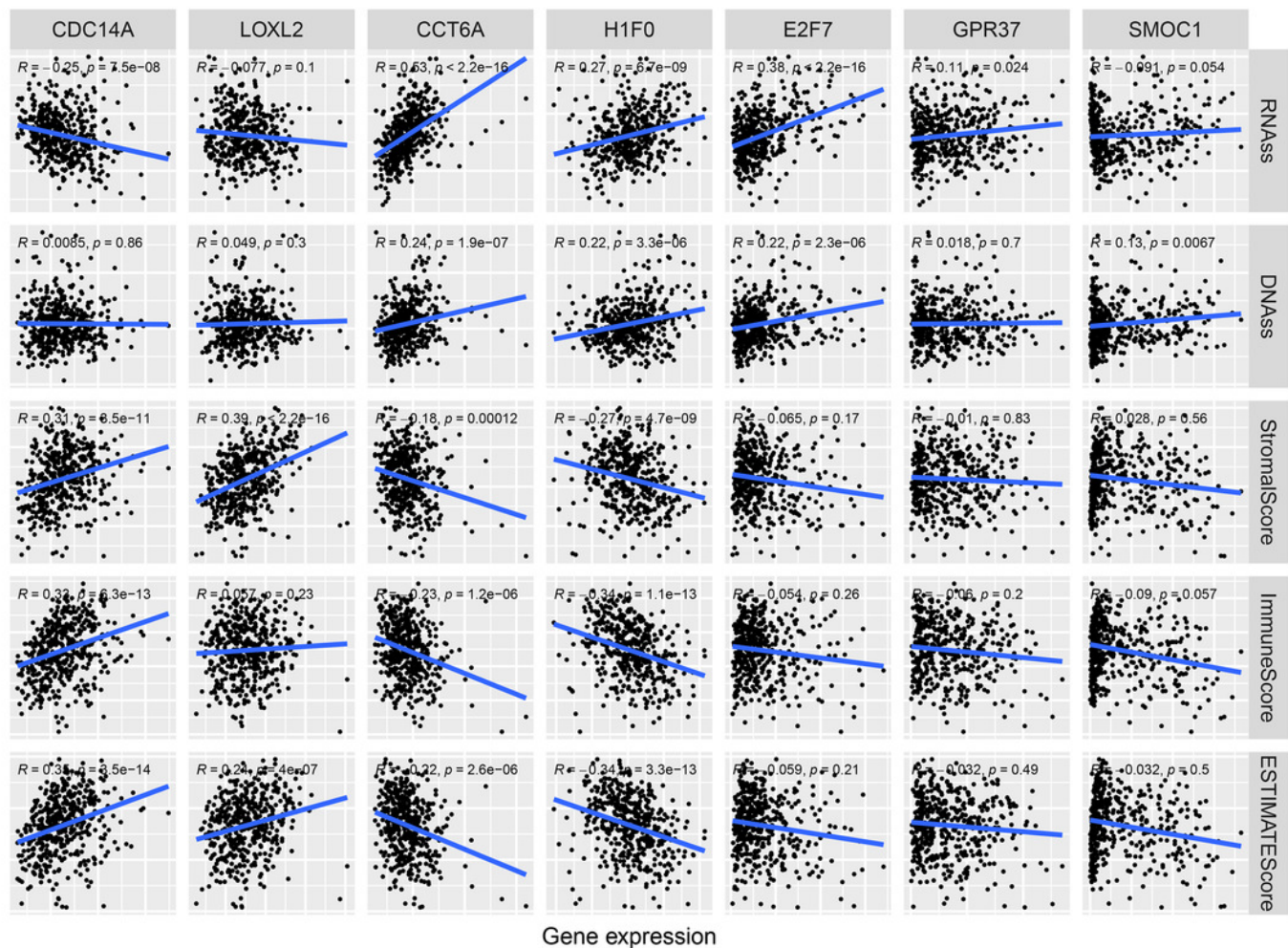
Figure 8

Figure 8 Differences of ssGSEA scores in high- and low- risk groups in TCGA

(**A**)(**B**) cohort and GEO (**C**)(**D**) cohort. (**A**),(**C**) refer to the difference in immune cells.(**B**),(**D**) refer to the difference in immune-associated functions *p < 0.05, **p < 0.01, and ***p < 0.001
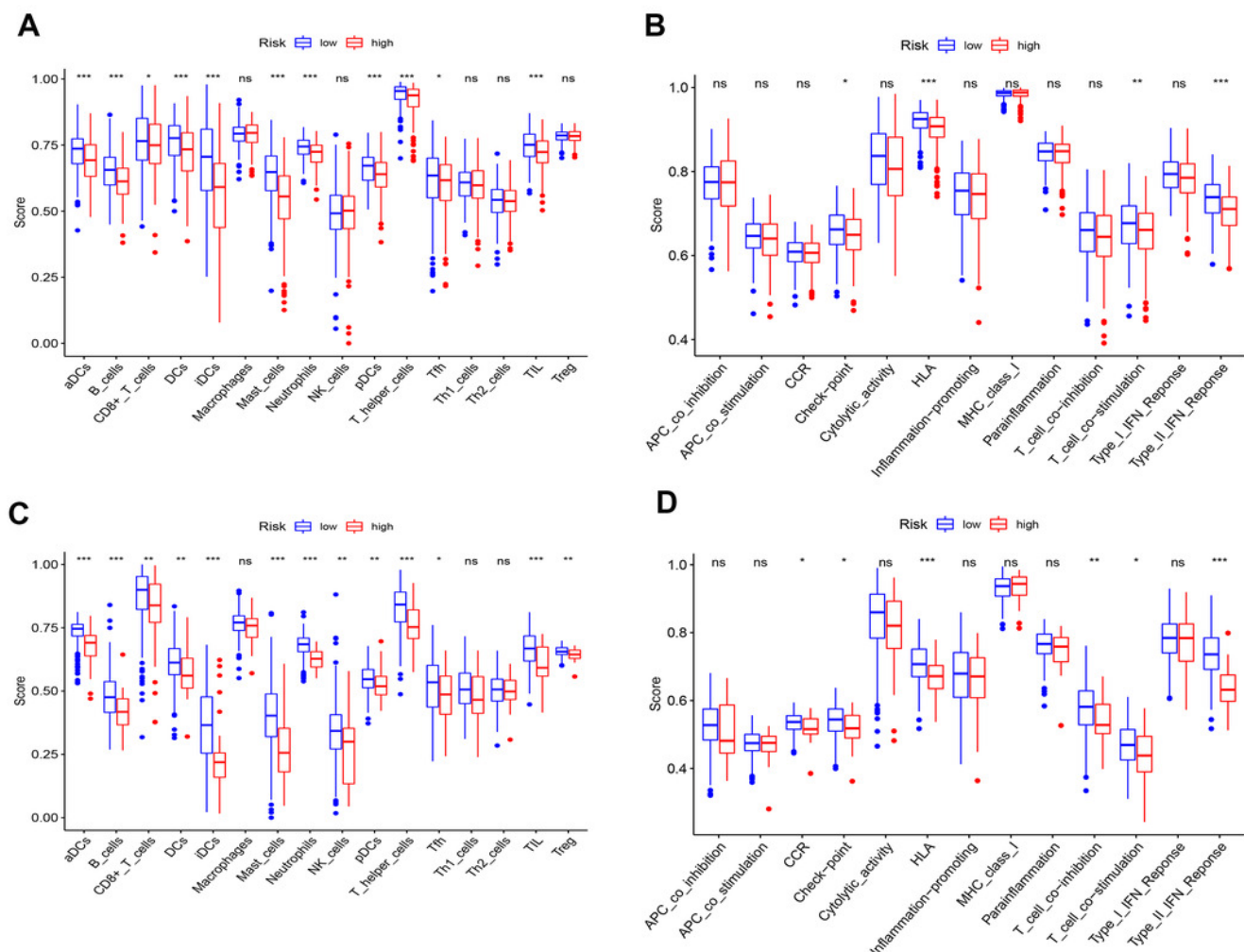
Table 1 Multivariate Cox proportional hazards regression model based on ceRNAs in LUAD.

LUAD lung adenocarcinoma,, HR hazrd ratio,CI confidence interval.

1 **Table 1   Multivariate Cox proportional hazards regression model based on ceRNAs in**
2 **LUAD.**

Genes	Coef	HR	95%CI	P value
CDC14A	-0.282	0.754	(0.616-0.924)	0.007
LOXL2	0.115	1.122	(1.002-1.256)	0.045
CCT6A	0.174	1.190	(0.965-1.468)	0.104
E2F7	0.187	1.206	(1.071-1.358)	0.002
GPR37	0.100	1.105	(1.019-1.198)	0.015
H1F0	0.198	1.218	(0.991-1.498)	0.061
SMOC1	0.068	1.071	(1.019-1.125)	0.007

3 LUAD lung adenocarcinoma,, HR hazrd ratio,CI confidence interval.

Table 2 Multivariate Cox proportional hazards regression model based on Immune cells in LUAD.

LUAD lung adenocarcinoma,Coef correlation coefficients, HR hazrd ratio,CI confidence interval.

1 **Table 2   Multivariate Cox proportional hazards regression model based on Immune cells**
2 **in LUAD.**

Immune cells	Coef	HR	95%CI	P value
T cells CD8	-3.426	0.033	(5.767e-04-1.835)	0.096
T cells regulatory (Tregs)	9.175	9653.370	(0.571-1.631e+08)	0.065
Monocytes	-5.318	0.005	(1.778e-05-1.351)	0.064
Macrophages M1	6.135	461.532	(7.447-2.860e+04)	0.004
Dendritic cells activated	5.539	254.456	(3.446-1.879e+04)	0.012
Mast cells activated	24.611	4.881e+10	(1.320e+03-1.805e+18)	0.006
Eosinophils	37.352	1.66589E+16	(0.012-2.282e+24)	0.080

3 LUAD lung adenocarcinoma,Coef correlation coefficients, HR hazrd ratio,CI confidence interval.