Bioinformatics analysis of microarray data to identify the candidate biomarkers of lung adenocarcinoma

Tingting Guo [Equal first author, 1], **Hongtao Ma** [Equal first author, 1], **Yubai Zhou** [Corresp. 1]

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

Corresponding Author: Yubai Zhou
Email address: zhouyubai@bjut.edu.cn

Background: Lung adenocarcinoma (LUAD) is the major subtype of lung cancer and the most lethal malignant disease worldwide. However, the molecular mechanisms underlying LUAD are not fully understood.

Methods: Four datasets GSE118370, GSE85841, GSE43458 and GSE32863 were obtained from the Gene Expression Omnibus (GEO). Identification of differentially expressed genes (DEGs) and functional enrichment analysis were performed using limma and clusterProfiler packages, respectively. Protein-protein interaction (PPI) network was constructed via STRING database and the module analysis was performed by Cytoscape. Then, overall survival analysis was performed using Kaplan-Meier curve, and prognostic candidate biomarkers were further analyzed using oncomine database.

Results: Totally, 349 DEGs were identified, including 275 downregulated and 74 upregulated genes which were significantly enriched in the biological process of extracellular structure organization, leukocyte migration and response to peptide. The mainly enriched pathways were complement and coagulation cascades, malaria and prion diseases. By extracting key modules from the PPI network, 11 hub genes were screened out. Survival analysis showed that except VSIG4, other hub genes may be involved in the development of LUAD, in which MYH10, METTL7A, FCER1G and TMOD1 have not been reported previously to correlated with LUAD. Briefly, novel hub genes identified in this study will help to deepen our understanding of the molecular mechanisms of LUAD carcinogenesis and progression, and to discover candidate targets for early detection and treatment of LUAD.

1 Bioinformatics analysis of microarray data to identify
2 the candidate biomarkers of lung adenocarcinoma

5 Tingting Guo, Hongtao Ma, Yubai Zhou

7 Department of Biotechnology, College of Life Science & Bioengineering, Beijing University of
8 Technology, Chaoyang, Beijing, China

10 Corresponding Author:
11 Yubai Zhou
12 Pingleyuan 100#, District of Chaoyang, Beijing 100124, China.
13 Email address: zhouyubai@bjut.edu.cn

15 Abstract

16 **Background:** Lung adenocarcinoma (LUAD) is the major subtype of lung cancer and the most

17 lethal malignant disease worldwide. However, the molecular mechanisms underlying LUAD are

18 not fully understood.

19 **Methods:** Four datasets GSE118370, GSE85841, GSE43458 and GSE32863 were obtained from

20 the Gene Expression Omnibus (GEO). Identification of differentially expressed genes (DEGs) and

21 functional enrichment analysis were performed using limma and clusterProfiler packages,

22 respectively. Protein-protein interaction (PPI) network was constructed via STRING database and

23 the module analysis was performed by Cytoscape. Then, overall survival analysis was performed

24 using Kaplan-Meier curve, and prognostic candidate biomarkers were further analyzed using

25 oncomine database.

26 **Results:** Totally, 349 DEGs were identified, including 275 downregulated and 74 upregulated

27 genes which were significantly enriched in the biological process of extracellular structure

28 organization, leukocyte migration and response to peptide. The mainly enriched pathways were

29 complement and coagulation cascades, malaria and prion diseases. By extracting key modules

30   from the PPI network, 11 hub genes were screened out. Survival analysis showed that except

31   VSIG4, other hub genes may be involved in the development of LUAD, in which MYH10,

32   METTL7A, FCER1G and TMOD1 have not been reported previously to correlated with LUAD.

33   Briefly, novel hub genes identified in this study will help to deepen our understanding of the

34   molecular mechanisms of LUAD carcinogenesis and progression, and to discover candidate targets

35   for early detection and treatment of LUAD.

36   **Introduction**

37   Lung cancer remains the one of leading healthy issues worldwide, with estimated 2.1 million new

38   cases and 1.8 million deaths in 2018(Bray et al. 2018). It has been ranked the first and second

39   cancer morbidity of male and female in China respectively, and has the highest mortality rate (Sun

40   et al. 2018). Lung adenocarcinoma (LUAD) is the most common subtype of lung cancer (Maemura

41   et al. 2018; Walters et al. 2013). More than 60% of LUAD patients were observed harboring

42   targetable gene alterations, which leading to remarkable responses in treating with tyrosine kinase

43   inhibitors (TKIs), and associated with improved survival rate (Kris et al. 2014). Despite the

44   substantial advance in combined therapies, the prognosis of LUAD is still dismal, the 5-year

45   survival rate is not over 20% (Chen et al. 2014b; Ettinger et al. 2013). Lacking sensitive and

46   specific early biomarkers, high possibility of drug resistance and metastasis, is considered to

47   contribute the high mortality of this disease. Therefore, there has a pressing need for identifying

48   the more sensitive and specific biomarkers or drug targets of LUAD for developing effective

49   diagnosis and treatment strategies.

50   Microarray technology provides an all-in-one system biology solution from hardware to

51   software systems. It can simultaneously scan the hybridization signals of tens of thousands of gene

52   probes in the chip and carry out quantitative analysis on the transcriptome profile of samples.

53   Recent advances especially in the algorithms of probe signal detection and analysis, such as the

54   introduction of artificial intelligence technologies, will make the results of microarray more

55   accurate and reliable(Gan et al. 2019a; Gan et al. 2019b; Peng 2006). The microarray technique

56   also provides a powerful tool for exploring the gene regulation pattern and molecular mechanisms

57 involved in oncogenesis and progression of LUAD. Recently, different types of biomarkers

58 including coding genes, miRNAs, long non-coding RNAs and circRNAs have been identified in

59 lung cancer. Dysregulation of these molecules is involved in the tumor progression or is associated

60 with the prognosis of patients (Di et al. 2019; Vargas & Harris 2016; Vencken et al. 2015; Wei &

61 Zhou 2016).   In view of the complexity of the molecular regulatory network of LUAD, current

62 studies on tumor biomarkers are not sufficient. Therefore, it is still necessary to identify novel

63 prognostic biomarkers, which will help us develop more sensitive and effective diagnostic and

64 therapeutic strategies. However, limited sample size and significant variability among different

65 projects make it hard to obtain credible results. In this study, four microarray datasets containing

66 mRNA expression data between LUAD and non-cancerous tissues were downloaded from GEO

67 and the DEGs were screened out. GO (Gene Ontology), KEGG (Kyoto Encyclopedia of Genes

68 and Genomes) and PPI network analyses were performed to explore the key modules and hub

69 genes involved in LUAD progression. In sum, 349 DEGs and 10 hub genes were screened out,

70 which may be candidate biomarkers for LUAD.

71

Materials & Methods

73 **Data download and pre-processing**

74 Four datasets GSE118370, GSE32863, GSE85841 and GSE43458, which contain the gene

75 expression data of LUAD and normal tissues, were downloaded from GEO

76 (https://www.ncbi.nlm.nih.gov/geo/) by getGEO function in package GEOquery (Davis & Meltzer

77 2007). The detail information of GEO datasets was listed in table 1. The raw expression files of

78 four microarray datasets were pre-processed according to the method described previously with

79 minor modifications (Giulietti et al. 2016). Briefly, the CEL format files were input and

80 background correction and normalization were conducted using the Robust Multichip Average

81 (RMA) function implemented in affy package in R environment (Bolstad et al. 2003; Irizarry et

82 al. 2003). Next, the array probes were converted into matched gene symbols according to

83 annotation information. In case of multiple probes corresponding to a single gene, the value of

84   gene expression was designated as the mean of the probes. Then, the batch effects among different

85   platforms were removed by ComBat function of sva package. Finally, the normalized microarray-

86   based data of four datasets were merged into a single global dataset which contained a total of

87   12,926 common genes in all four GEO datasets.

88   **Identification of DEGs**

89   Identifying differentially expressed genes in different disease states and investigating their

90   functions and interactions may help to unravel potential regulatory mechanisms for disease

91   occurrence and progression. In present study, the DEGs between LUAD and normal tissues were

92   identified by limma package (Ritchie et al. 2015). The Benjamini-Hochberg procedure was

93   introduced to reduce the false positive rate (FDR) in multiple comparisons (Benjamini & Hochberg

94   1995). Genes with | log2 Fold Change | $\geqslant$ 1 and FDR < 0.05 were considered as DEGs.

95   **GO and KEGG enrichment analysis**

96   GO and KEGG analyses were conducted using enrichGO and enrichKEGG functions of

97   clusterProfiler package, respectively (Yu et al. 2012). p.adjust (FDR)<0.05 was considered to be

98   statistically significant.

99   **Construction of PPI network and module analysis**

100   The network of proteins interaction provides valuable clues for understanding the molecular

101   mechanisms underlying the progress of carcinoma. The PPI network was constructed by Search

102   Tool for the Retrieval of Interacting Genes (STRING) (https://string-db.org/) with interaction

103   score of 0.9 as the threshold. Subsequently, the candidate modules were detected by Cytoscape

104   plugin molecular complex detection (MCODE) with default parameters: degree cut-off=2, node

105   score cut-off=0.2, k-core=2, and max depth=100.

106   **Hub gene analysis**

107   The seed genes in modules referred to hub genes. The overall survival analyses were performed

108   using online tool Kaplan-Meier Plotter (http://kmplot.com/)(Gyorffy et al. 2013). The logrank

109   P<0.05 was considered statistically significant. The association of expression level of hub genes

110   with clinical traits were analysis using Oncomine database(Rhodes et al. 2007).

111

Results

113   **Data preprocessing and DEG screening**

114 Four GEO datasets were downloaded, pre-processed and merged into a global dataset which

115 contained 112 LUAD and 102 normal samples. Totally, 349 DEGs were identified by limma

116 package (Ritchie et al. 2015), including 74 up-regulated genes and 275 down-regulated genes. The

117 most statistically significant up-regulated and down-regulated genes are listed in Table 2. The

118 distribution of DEGs was presented by volcano plot (Fig. 1).

119 **GO and KEGG analysis**

120 The biological functions and pathways analyses were conducted using R package clusterProfiler

121 (Yu et al. 2012). The GO categories of biological process (BP), cellular component (CC) and

122 molecular function (MF) were enriched respectively (Fig.2A-C) and the top 15 GO terms of the

123 up-regulated and down-regulated DEGs were listed in supplemental table 1 ,2 respectively. The

124 up-regulated DEGs were mainly associated with extracellular matrix processing, such as

125 extracellular matrix organization (BP, GO:0030198), extracellular matrix (CC, GO:0031012) and

126 serine-type endopeptidase activity (MF, GO:0004252). The down-regulated DEGs were most

127 significantly related to response to corticosteroid (BP, GO:0031960), extracellular matrix (CC,

128 GO:0031012) and growth factor binding (MF, GO:0019838). The DEGs were mainly enriched in

129 pathways of complement and coagulation cascades(hsa04610) , malaria (hsa05144), prion

130 diseases(hsa05020), fluid shear stress and atherosclerosis(hsa05418) , AGE-RAGE signaling

131 pathway in diabetic complications(hsa04933) , vascular smooth muscle contraction (hsa04270),

132 IL-17 signaling pathway (hsa04657), leukocyte transendothelial migration (hsa04670), protein

133 digestion and absorption (hsa04974) and drug metabolism - cytochrome P450(hsa00982) (Table 3

134 and Fig. 2D).

135 **Construction of PPI network and module analysis**

136 PPI network reflect the spatiotemporal relationship of macromolecules within the cell which will

137 provide valuable information about molecular mechanisms in physiological and pathological

138   process. To explore the molecular mechanisms underlying LUAD progression, online database

139   STRING was applied to construct the PPI network. The interaction score of 0.9 (highest

140   confidence) was set as threshold, and nodes without connections were removed from network.

141   Finally, the PPI network consisted of 349 nodes with 277 edges, and average local clustering

142   coefficient was 0.337(PPI enrichment p-value< 1.0E-16) (Fig. 3A). Then, the key modules were

143   identified via MCODE plugin. Eleven functional clusters of modules and related hub genes were

144   detected. The top 3 significant modules were presented in Fig. 3C-E. The KEGG analysis of

145   module genes revealed that the top 3 modules were mainly associated with chemokine signaling

146   pathway(hsa04062), complement and coagulation cascades(hsa04610), human cytomegalovirus

147   infection(hsa05163) and vascular smooth muscle contraction(hsa04270) (Fig. 3B).

148   **Hub genes analysis**

149   A total of 11 genes were identified as hub genes. The overall survival analysis of the hub genes

150   was performed using Kaplan-Meier curve. Except VSIG4, LUAD patients with downregulated

151   ADAMTS8, AOX1, EFEMP1, METTL7A, MYH10, PTGER4, TMOD1, CDH13 and upregulated

152   PRC1 showed worse overall survival (Fig. 4A-I). It is worth noting that FCER1G is downregulated

153   in LUAD patients, but the low expression level is associated with better overall survival (HR =

154   1.87) (Fig. 4J). Subsequently, the expression status of hub genes with HR<0.5 or HR>2 were

155   further validated using oncomine database. The result showed that ADAMTS8, METTL7A and

156   MYH10 were significantly downregulated and PRC1 was markedly overexpressed in LUAD in

157   the different datasets (Fig. 5). In the Okayama Lung dataset, the alternation of ADAMTS8,

158   METTL7A, MYH10 and PRC1 were associated with tumor grade (Fig. 6), implicating vital roles

159   of these genes in the carcinogenesis or progression of LUAD.

160

161   **Discussion**

162   In this study, four GEO datasets were analyzed and 349 DEGs were identified, including 74 up-

163   regulated and 275 down-regulated genes. The KEGG analysis revealed the top three enriched

164   pathways were complement and coagulation cascades, malaria and prion diseases. These

165 annotation results provided valuable clues to reveal molecular interactions in the development of

166 LUAD. Indeed, the complement system has been reported to play critical roles in tumor

167 progression (Zhao et al. 2019). The overexpressed C3 in blood and downregulated C3 in tumor

168 tissues were observed in lung cancer patients, and related to poorer prognosis (Ajona et al. 2013;

169 Lin et al. 2014; Mehan et al. 2012). The confusing experimental results suggest that the

170 complement system may be involved in complex tumor regulatory processes by reshaping tumor

171 microenvironment, which is worthy of further study. In recent years, malaria and prion disease,

172 which had not previously attracted much attention, have also been found to be associated with

173 tumors. Epidemiological study has shown that the incidence of malaria is negatively correlated

174 with the mortality of colorectal cancer, breast cancer and lung cancer (Qin et al. 2017). The

175 proposed anti-tumor mechanisms included systemly stimulating the immune responses (Chen et

176 al. 2011) and inhibiting key pathways in tumor progress (Deng et al. 2018). Previously, prion

177 protein (PrPc) was thought to play a role only in the central nervous system. However,

178 accumulating evidence shows that PrPc has wider biological functions that were not previously

179 expected (Mehrpour & Codogno 2010). PrPc may be associated with the biology of many cancers,

180 and overexpression of PrPc promotes the proliferation, invasion and metastasis of the gastric

181 cancer cell(Liang et al. 2009; Mehrpour & Codogno 2010; Pan et al. 2006). These data may

182 provide new ideas and directions for the mechanism research and therapeutic strategy of lung

183 adenocarcinoma. The GO analyses indicated that DEGs were significantly related to biological

184 process of extracellular matrix organization and process. Previous studies have reported that

185 extracellular matrix (ECM) remodeling promotes cancer progression and is associated with a poor

186 prognosis in lung cancer patients (Kopparam et al. 2017; Xia et al. 2012). The up- and down-

187 regulated DEGs were simultaneously enriched into extracellular matrix process, which is

188 consistent with the propensity for metastasis and highly invasive characteristics of LUAD.

189 In line with the GO analysis, among 11 hub genes identified by PPI network and modules

190 analysis, EFEMP1, PTGER4, ADAMTS8, CDH13, MYH10 and METTL7A are associated with

191 extracellular matrix process, and survival analysis indicated that down-regulation of these hub

192   genes was associated with worse overall survival. EFEMP1 plays distinct biological functions in

193   different tumors. In osteosarcoma and gliomas, EFEMP1 is overexpressed and promotes the

194   invasion and metastasis of tumor cells in vitro and in vivo by activating the expression of MMP-2

195   and notch signaling, respectively(Hu et al. 2012; Hu et al. 2009; Wang et al. 2015), while in gastric

196   cancer, endometrial carcinoma, hepatocellular carcinoma and lung cancer, EFEMP1 is

197   downregulated, and is proposed as a prognosis biomarker(Chen et al. 2014a; Kim et al. 2014;

198   Nomoto et al. 2010; Yang et al. 2013; Yue et al. 2007; Zhu et al. 2014). Overexpression of

199   EFEMP1 has been reported to suppressed invasion and migration of LUAD cells via inhibiting the

200   epithelial-to-mesenchymal transition (EMT) pathway(Kim et al. 2014).PTGER4 is overexpressed

201   and proposed as a therapeutic target for LUAD and other cancers(Doherty et al. 2009; Fulton et

202   al. 2006; Heinrichs et al. 2018; Kim et al. 2010; Ma et al. 2006; Xin et al. 2012). The above report

203   is inconsistent with our results, which may be due to the heterogeneity of the tumor and the limited

204   number of samples. Therefore, subsequent large sample functional verification is required.

205   ADAMTS8 encodes an inactive proenzyme and forms mature active enzyme by proteolysis(Apte

206   2009). ADAMTS8 is identified as a secretory angiogenesis inhibitor that inhibits VEGF-mediated

207   angiogenesis by blocking the EGFR signaling pathway(Choi et al. 2014; Dunn et al. 2006;

208   Vazquez et al. 1999). Downregulated ADAMTS8 in some cancers including NSCLC (Heighway

209   et al. 2002; Huang et al. 2019; Masui et al. 2001; Porter et al. 2004; Rodriguez-Rodero et al. 2013;

210   Zhao et al. 2018) is associated with poorer prognosis (Drilon et al. 2014; Li et al. 2015; Porter et

211   al. 2006). CDH13 encodes a member of cadherin superfamily. The hypermethylation in promoter

212   region of CDH13 was frequently observed in lung cancer, and proposed to correlate to drug

213   sensitivity and poorer prognosis (Kontic et al. 2012; Toyooka et al. 2006; Zhai & Li 2014; Zhong

214   et al. 2015). As the only upregulated hub gene, PRC1 is involved in the process of cytokinesis and

215   is upregulated in breast cancer, hepatocellular carcinoma, and lung cancer(Jiang et al. 1998; Liu

216   et al. 2018; Shimo et al. 2007; Zhan et al. 2017a). Recent research has found that it promoted

217   proliferation and metastasis of lung adenocarcinoma cells via potentiating the Wnt/β-catenin

218 pathway(Zhan et al. 2017b), and inhibiting the expression of PRC1 in LUAD cells by miR-1-3p

219 was reported to suppress tumorigenesis (Li et al. 2019).

220   Literature retrieval showed that the relation of LUAD and hub genes MYH10, METTL7A,

221 FCER1G and TMOD1 has not been reported. MYH10 belongs to the myosin superfamily, which

222 can regulate ECM remodeling (Kim et al. 2018; Kim et al. 2015). Methyltransferase METTL7A

223 was found to facilitate Hepatitis C Virus (HCV) propagation via recruitment of NS4B (Park et al.

224 2015). FCER1G encodes a high affinity IgE receptor which is associated with prognosis of renal

225 clear cell carcinoma (Chen et al. 2017). As an aldehyde oxidase, TMOD1 is an actin-capping

226 protein involved in regulation of the length and depolymerization of actin filaments.

227 Overexpression of TMOD1 promotes cell proliferation of breast cancers and metastasis of oral

228 squamous cell carcinoma (Ito-Kureha et al. 2015; Suzuki et al. 2016). However, the molecular

229 mechanism of these newly identified hub genes in LUAD remains largely unknown, and further

230 functional studies were warranted.

231   In present study, we have identified a set of candidate biomarkers that may play an important

232 role in the progression of LUAD. These newly identified hub genes could be used as research

233 subjects for exploring their roles in the disease process, so as to further deepen our understanding

234 of the molecular mechanism of LUAD, and also as potential prognostic biomarkers for clinical

235 validation studies to clarify their prognostic effects. However, there are still some limitations in

236 this study. First, this study is based on bioinformatics analysis of published data and lacks

237 experimental verification, and the study cannot determine whether there is a causal relationship

238 between the differential expression of hub genes and disease progression. Finally, although we

239 combined four GEO datasets, the number of samples is still relatively small, which may lead to

240 potential unreliable results. Therefore, subsequent bioinformatics analysis and experimental

241 verification with larger samples are necessary.

242

243 **Conclusions**

244 In summary, this study identified several differentially expressed genes by integrating four GEO

245 datasets and extracted 11 hub genes from PPI network, among which 10 hub genes were shown to

246 be related to the occurrence and development of lung adenocarcinoma and 4 hub genes have not

247 been previously reported but may play an important role in LUAD. The molecular mechanism of

248 these novel hub genes in LUAD is worthy of further study, and relevant prognostic model can also

249 be constructed based on these genes for risk assessment, classification and prognostic judgment of

250 patients with LUAD.

251

References

253 Ajona D, Pajares MJ, Corrales L, Perez-Gracia JL, Agorreta J, Lozano MD, Torre W, Massion PP, de-Torres JP,
254   Jantus-Lewintre E, Camps C, Zulueta JJ, Montuenga LM, and Pio R. 2013. Investigation of complement
255   activation product c4d as a diagnostic and prognostic biomarker for lung cancer. *J Natl Cancer Inst* 105:1385-
256   1393. 10.1093/jnci/djt205

257 Apte SS. 2009. A disintegrin-like and metalloprotease (reprolysin-type) with thrombospondin type 1 motif
258   (ADAMTS) superfamily: functions and mechanisms. *J Biol Chem* 284:31493-31497.
259   10.1074/jbc.R109.052340

260 Benjamini Y, and Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to
261   multiple testing. *Journal of the Royal statistical society: series B* 57:289-300.

262 Bolstad BM, Irizarry RA, Astrand M, and Speed TP. 2003. A comparison of normalization methods for high density
263   oligonucleotide array data based on variance and bias. *Bioinformatics* 19:185-193.
264   10.1093/bioinformatics/19.2.185

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

268 Chen L, He Z, Qin L, Li Q, Shi X, Zhao S, Chen L, Zhong N, and Chen X. 2011. Antitumor effect of malaria parasite
269   infection in a murine Lewis lung cancer model through induction of innate and adaptive immunity. *PLoS
270   One* 6:e24407. 10.1371/journal.pone.0024407

271 Chen L, Yuan L, Wang Y, Wang G, Zhu Y, Cao R, Qian G, Xie C, Liu X, Xiao Y, and Wang X. 2017. Co-expression
272   network analysis identified FCER1G in association with progression and prognosis in human clear cell renal
273   cell carcinoma. *Int J Biol Sci* 13:1361-1372. 10.7150/ijbs.21657

274 Chen X, Meng J, Yue W, Yu J, Yang J, Yao Z, and Zhang L. 2014a. Fibulin-3 suppresses Wnt/beta-catenin signaling
275   and lung cancer invasion. *Carcinogenesis* 35:1707-1716. 10.1093/carcin/bgu023

276 Chen Z, Fillmore CM, Hammerman PS, Kim CF, and Wong KK. 2014b. Non-small-cell lung cancers: a heterogeneous
277   set of diseases. *Nat Rev Cancer* 14:535-546. 10.1038/nrc3775

278 Choi GC, Li J, Wang Y, Li L, Zhong L, Ma B, Su X, Ying J, Xiang T, Rha SY, Yu J, Sung JJ, Tsao SW, Chan AT,
279   and Tao Q. 2014. The metalloprotease ADAMTS8 displays antitumor properties through antagonizing
280   EGFR-MEK-ERK signaling and is silenced in carcinomas by CpG methylation. *Mol Cancer Res* 12:228-

281   238. 10.1158/1541-7786.Mcr-13-0195

282   Davis S, and Meltzer PS. 2007. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor.
283   *Bioinformatics* 23:1846-1847. 10.1093/bioinformatics/btm254

284   Deng XF, Zhou D, Liu QX, Zheng H, Ding Y, Xu WY, Min JX, and Dai JG. 2018. Plasmodium circumsporozoite
285   protein suppresses the growth of A549 cells via inhibiting nuclear transcription factor kappaB. *Oncol Lett*
286   15:6585-6591. 10.3892/ol.2018.8115

287   Di X, Jin X, Li R, Zhao M, and Wang K. 2019. CircRNAs and lung cancer: Biomarkers and master regulators. *Life*
288   *Sci* 220:177-185. 10.1016/j.lfs.2019.01.055

289   Doherty GA, Byrne SM, Molloy ES, Malhotra V, Austin SC, Kay EW, Murray FE, and Fitzgerald DJ. 2009.
290   Proneoplastic effects of PGE2 mediated by EP4 receptor in colorectal cancer. *BMC Cancer* 9:207.
291   10.1186/1471-2407-9-207

292   Drilon A, Sugita H, Sima CS, Zauderer M, Rudin CM, Kris MG, Rusch VW, and Azzoli CG. 2014. A prospective
293   study of tumor suppressor gene methylation as a prognostic biomarker in surgically resected stage I to IIIA
294   non-small-cell lung cancers. *J Thorac Oncol* 9:1272-1277. 10.1097/jto.0000000000000256

295   Dunn JR, Reed JE, du Plessis DG, Shaw EJ, Reeves P, Gee AL, Warnke P, and Walker C. 2006. Expression of
296   ADAMTS-8, a secreted protease with antiangiogenic properties, is downregulated in brain tumours. *Br J*
297   *Cancer* 94:1186-1193. 10.1038/sj.bjc.6603006

298   Ettinger DS, Akerley W, Borghaei H, Chang AC, Cheney RT, Chirieac LR, D'Amico TA, Demmy TL, Govindan R,
299   Grannis FW, Jr., Grant SC, Horn L, Jahan TM, Komaki R, Kong FM, Kris MG, Krug LM, Lackner RP,
300   Lennes IT, Loo BW, Jr., Martins R, Otterson GA, Patel JD, Pinder-Schenck MC, Pisters KM, Reckamp K,
301   Riely GJ, Rohren E, Shapiro TA, Swanson SJ, Tauer K, Wood DE, Yang SC, Gregory K, and Hughes M.
302   2013. Non-small cell lung cancer, version 2.2013. *J Natl Compr Canc Netw* 11:645-653; quiz 653.

303   Fulton AM, Ma X, and Kundu N. 2006. Targeting prostaglandin E EP receptors to inhibit metastasis. *Cancer Res*
304   66:9794-9797. 10.1158/0008-5472.Can-06-2067

305   Gan Z, Zeng N, Zou F, Chen J, Du M, Liao L-C, Li H, and Zhang Y. 2019a. Multilevel Segmentation Optimized by
306   Physical Information for Gridding of Microarray Images. *Ieee Access* 7:32146-32153.

307   Gan Z, Zou F, Zeng N, Xiong B, Liao L, Li H, Luo X, and Du M. 2019b. Wavelet Denoising Algorithm based on
308   NDOA Compressed Sensing for Fluorescence Image of Microarray. *Ieee Access* 7:13338-13346.
309   http://10.0.4.85/ACCESS.2019.2891759

310   Giulietti M, Occhipinti G, Principato G, and Piva F. 2016. Weighted gene co-expression network analysis reveals key
311   genes involved in pancreatic ductal adenocarcinoma development. *Cell Oncol (Dordr)* 39:379-388.
312   10.1007/s13402-016-0283-7

313   Gyorffy B, Surowiak P, Budczies J, and Lanczky A. 2013. Online survival analysis software to assess the prognostic
314   value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One* 8:e82241.
315   10.1371/journal.pone.0082241

316   Heighway J, Knapp T, Boyce L, Brennand S, Field JK, Betticher DC, Ratschiller D, Gugger M, Donovan M, Lasek
317   A, and Rickert P. 2002. Expression profiling of primary non-small cell lung cancer for target identification.
318   *Oncogene* 21:7749-7763. 10.1038/sj.onc.1205979

319   Heinrichs SKM, Hess T, Becker J, Hamann L, Vashist YK, Butterbach K, Schmidt T, Alakus H, Krasniuk I, Hoblinger
320   A, Lingohr P, Ludwig M, Hagel AF, Schildberg CW, Veits L, Gyvyte U, Weise K, Schuller V, Bohmer AC,
321   Schroder J, Gehlen J, Kreuser N, Hofer S, Lang H, Lordick F, Malfertheiner P, Moehler M, Pech O, Vassos

322 N, Rodermann E, Izbicki JR, Kruschewski M, Ott K, Schumann RR, Vieth M, Mangold E, Gasenko E,
323 Kupcinskas L, Brenner H, Grimminger P, Bujanda L, Sopena F, Espinel J, Thomson C, Perez-Aisa A, Campo
324 R, Geijo F, Collette D, Bruns C, Messerle K, Gockel I, Nothen MM, Lippert H, Ridwelski K, Lanas A, Keller
325 G, Knapp M, Leja M, Kupcinskas J, Garcia-Gonzalez MA, Venerito M, and Schumacher J. 2018. Evidence
326 for PTGER4, PSCA, and MBOAT7 as risk genes for gastric cancer on the genome and transcriptome level.
327 *Cancer Med* 7:5057-5065. 10.1002/cam4.1719
328 Hu B, Nandhu MS, Sim H, Agudelo-Garcia PA, Saldivar JC, Dolan CE, Mora ME, Nuovo GJ, Cole SE, and Viapiano
329 MS. 2012. Fibulin-3 promotes glioma growth and resistance through a novel paracrine regulation of Notch
330 signaling. *Cancer Res* 72:3873-3885. 10.1158/0008-5472.Can-12-1060
331 Hu B, Thirtamara-Rajamani KK, Sim H, and Viapiano MS. 2009. Fibulin-3 is uniquely upregulated in malignant
332 gliomas and promotes tumor cell motility and invasion. *Mol Cancer Res* 7:1756-1770. 10.1158/1541-
333 7786.Mcr-09-0207
334 Huang J, Sun Y, Chen H, Liao Y, Li S, Chen C, and Yang Z. 2019. ADAMTS5 acts as a tumor suppressor by inhibiting
335 migration, invasion and angiogenesis in human gastric cancer. *Gastric Cancer* 22:287-301. 10.1007/s10120-
336 018-0866-2
337 Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, and Speed TP. 2003. Exploration,
338 normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249-264.
339 10.1093/biostatistics/4.2.249
340 Ito-Kureha T, Koshikawa N, Yamamoto M, Semba K, Yamaguchi N, Yamamoto T, Seiki M, and Inoue J. 2015.
341 Tropomodulin 1 expression driven by NF-kappaB enhances breast cancer growth. *Cancer Res* 75:62-72.
342 10.1158/0008-5472.Can-13-3455
343 Jiang W, Jimenez G, Wells NJ, Hope TJ, Wahl GM, Hunter T, and Fukunaga R. 1998. PRC1: a human mitotic spindle-
344 associated CDK substrate protein required for cytokinesis. *Mol Cell* 2:877-885.
345 Kim HT, Yin W, Jin YJ, Panza P, Gunawan F, Grohmann B, Buettner C, Sokol AM, Preussner J, Guenther S, Kostin
346 S, Ruppert C, Bhagwat AM, Ma X, Graumann J, Looso M, Guenther A, Adelstein RS, Offermanns S, and
347 Stainier DYR. 2018. Myh10 deficiency leads to defective extracellular matrix remodeling and pulmonary
348 disease. *Nat Commun* 9:4600. 10.1038/s41467-018-06833-7
349 Kim IG, Kim SY, Choi SI, Lee JH, Kim KC, and Cho EW. 2014. Fibulin-3-mediated inhibition of epithelial-to-
350 mesenchymal transition and self-renewal of ALDH+ lung cancer stem cells through IGF1R signaling.
351 *Oncogene* 33:3908-3917. 10.1038/onc.2013.373
352 Kim JI, Lakshmikanthan V, Frilot N, and Daaka Y. 2010. Prostaglandin E2 promotes lung cancer cell migration via
353 EP4-betaArrestin1-c-Src signalsome. *Mol Cancer Res* 8:569-577. 10.1158/1541-7786.Mcr-09-0511
354 Kim JS, Kurie JM, and Ahn YH. 2015. BMP4 depletion by miR-200 inhibits tumorigenesis and metastasis of lung
355 adenocarcinoma cells. *Mol Cancer* 14:173. 10.1186/s12943-015-0441-y
356 Kontic M, Stojsic J, Jovanovic D, Bunjevacki V, Ognjanovic S, Kuriger J, Puumala S, and Nelson HH. 2012. Aberrant
357 promoter methylation of CDH13 and MGMT genes is associated with clinicopathologic characteristics of
358 primary non-small-cell lung carcinoma. *Clin Lung Cancer* 13:297-303. 10.1016/j.cllc.2011.11.003
359 Kopparam J, Chiffelle J, Angelino P, Piersigilli A, Zangger N, Delorenzi M, and Meylan E. 2017. RIP4 inhibits
360 STAT3 signaling to sustain lung adenocarcinoma differentiation. *Cell Death Differ* 24:1761-1771.
361 10.1038/cdd.2017.81
362 Kris MG, Johnson BE, Berry LD, Kwiatkowski DJ, Iafrate AJ, Wistuba, II, Varella-Garcia M, Franklin WA, Aronson

363   SL, Su PF, Shyr Y, Camidge DR, Sequist LV, Glisson BS, Khuri FR, Garon EB, Pao W, Rudin C, Schiller
364   J, Haura EB, Socinski M, Shirai K, Chen H, Giaccone G, Ladanyi M, Kugler K, Minna JD, and Bunn PA.
365   2014. Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. *JAMA*
366   311:1998-2006. 10.1001/jama.2014.3741

367 Leek JT, Johnson WE, Parker HS, Jaffe AE, and Storey JD. 2012. The sva package for removing batch effects and
368   other unwanted variation in high-throughput experiments. *Bioinformatics* 28:882-883.
369   10.1093/bioinformatics/bts034

370 Li C, Xiong Y, Yang X, Wang L, Zhang S, Dai N, Li M, Ren T, Yang Y, Zhou SF, Gan L, and Wang D. 2015. Lost
371   expression of ADAMTS5 protein associates with progression and poor prognosis of hepatocellular
372   carcinoma. *Drug Des Devel Ther* 9:1773-1783. 10.2147/dddt.S77069

373 Li T, Wang X, Jing L, and Li Y. 2019. MiR-1-3p Inhibits Lung Adenocarcinoma Cell Tumorigenesis via Targeting
374   Protein Regulator of Cytokinesis 1. *Front Oncol* 9:120. 10.3389/fonc.2019.00120

375 Liang J, Wang J, Luo G, Pan Y, Wang X, Guo C, Zhang D, Yin F, Zhang X, Liu J, Wang J, Guo X, Wu K, and Fan
376   D. 2009. Function of PrPC (1-OPRD) in biological activities of gastric cancer cell lines. *J Cell Mol Med*
377   13:4453-4464. 10.1111/j.1582-4934.2009.00687.x

378 Lin K, He S, He L, Chen J, Cheng X, Zhang G, and Zhu B. 2014. Complement component 3 is a prognostic factor of
379   nonsmall cell lung cancer. *Mol Med Rep* 10:811-817. 10.3892/mmr.2014.2230

380 Liu X, Li Y, Meng L, Liu XY, Peng A, Chen Y, Liu C, Chen H, Sun S, Miao X, Zhang Y, Zheng L, and Huang K.
381   2018. Reducing protein regulator of cytokinesis 1 as a prospective therapy for hepatocellular carcinoma. *Cell*
382   *Death Dis* 9:534. 10.1038/s41419-018-0555-4

383 Ma X, Kundu N, Rifat S, Walser T, and Fulton AM. 2006. Prostaglandin E receptor EP4 antagonism inhibits breast
384   cancer metastasis. *Cancer Res* 66:2923-2927. 10.1158/0008-5472.Can-05-4348

385 Maemura K, Watanabe K, Ando T, Hiyama N, Sakatani T, Amano Y, Kage H, Nakajima J, Yatomi Y, Nagase T, and
386   Takai D. 2018. Altered editing level of microRNAs is a potential biomarker in lung adenocarcinoma. *Cancer*
387   *Sci* 109:3326-3335. 10.1111/cas.13742

388 Masui T, Hosotani R, Tsuji S, Miyamoto Y, Yasuda S, Ida J, Nakajima S, Kawaguchi M, Kobayashi H, Koizumi M,
389   Toyoda E, Tulachan S, Arii S, Doi R, and Imamura M. 2001. Expression of METH-1 and METH-2 in
390   pancreatic cancer. *Clin Cancer Res* 7:3437-3443.

391 Mehan MR, Ayers D, Thirstrup D, Xiong W, Ostroff RM, Brody EN, Walker JJ, Gold L, Jarvis TC, Janjic N, Baird
392   GS, and Wilcox SK. 2012. Protein signature of lung cancer tissues. *PLoS One* 7:e35157.
393   10.1371/journal.pone.0035157

394 Mehrpour M, and Codogno P. 2010. Prion protein: From physiology to cancer biology. *Cancer Lett* 290:1-23.
395   10.1016/j.canlet.2009.07.009

396 Nomoto S, Kanda M, Okamura Y, Nishikawa Y, Qiyong L, Fujii T, Sugimoto H, Takeda S, and Nakao A. 2010.
397   Epidermal growth factor-containing fibulin-like extracellular matrix protein 1, EFEMP1, a novel tumor-
398   suppressor gene detected in hepatocellular carcinoma using double combination array analysis. *Ann Surg*
399   *Oncol* 17:923-932. 10.1245/s10434-009-0790-0

400 Pan Y, Zhao L, Liang J, Liu J, Shi Y, Liu N, Zhang G, Jin H, Gao J, Xie H, Wang J, Liu Z, and Fan D. 2006. Cellular
401   prion protein promotes invasion and metastasis of gastric cancer. *FASEB J* 20:1886-1888. 10.1096/fj.06-
402   6138fje

403 Park EM, Lim YS, Ahn BY, and Hwang SB. 2015. AAM-B Interacts with Nonstructural 4B and Regulates Hepatitis

404   C Virus Propagation. *PLoS One* 10:e0132839. 10.1371/journal.pone.0132839

405   Peng Y. 2006. A novel ensemble machine learning for robust microarray data classification. *Comput Biol Med* 36:553-
406   573. 10.1016/j.compbiomed.2005.04.001

407   Porter S, Scott SD, Sassoon EM, Williams MR, Jones JL, Girling AC, Ball RY, and Edwards DR. 2004. Dysregulated
408   expression of adamalysin-thrombospondin genes in human breast carcinoma. *Clin Cancer Res* 10:2429-2440.

409   Porter S, Span PN, Sweep FC, Tjan-Heijnen VC, Pennington CJ, Pedersen TX, Johnsen M, Lund LR, Romer J, and
410   Edwards DR. 2006. ADAMTS8 and ADAMTS15 expression predicts survival in human breast carcinoma.
411   *Int J Cancer* 118:1241-1247. 10.1002/ijc.21476

412   Qin L, Chen C, Chen L, Xue R, Ou-Yang M, Zhou C, Zhao S, He Z, Xia Y, He J, Liu P, Zhong N, and Chen X. 2017.
413   Worldwide malaria incidence and cancer mortality are inversely associated. *Infect Agent Cancer* 12:14.
414   10.1186/s13027-017-0117-x

415   Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-
416   Beal C, Kulkarni P, Varambally S, Ghosh D, and Chinnaiyan AM. 2007. Oncomine 3.0: genes, pathways,
417   and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* 9:166-180.

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

420   Rodriguez-Rodero S, Fernandez AF, Fernandez-Morera JL, Castro-Santos P, Bayon GF, Ferrero C, Urdinguio RG,
421   Gonzalez-Marquez R, Suarez C, Fernandez-Vega I, Fresno Forcelledo MF, Martinez-Camblor P, Mancikova
422   V, Castelblanco E, Perez M, Marron PI, Mendiola M, Hardisson D, Santisteban P, Riesco-Eizaguirre G,
423   Matias-Guiu X, Carnero A, Robledo M, Delgado-Alvarez E, Menendez-Torre E, and Fraga MF. 2013. DNA
424   methylation signatures identify biologically distinct thyroid cancer subtypes. *J Clin Endocrinol Metab*
425   98:2811-2821. 10.1210/jc.2012-3566

426   Shimo A, Nishidate T, Ohta T, Fukuda M, Nakamura Y, and Katagiri T. 2007. Elevated expression of protein regulator
427   of cytokinesis 1, involved in the growth of breast cancer cells. *Cancer Sci* 98:174-181. 10.1111/j.1349-
428   7006.2006.00381.x

429   Sun KX, Zheng RS, Zeng HM, Zhang SW, Zou XN, Gu XY, Xia CF, Yang ZX, Li H, Chen WQ, and He J. 2018.
430   [The incidence and mortality of lung cancer in China, 2014]. *Zhonghua Zhong Liu Za Zhi* 40:805-811.
431   10.3760/cma.j.issn.0253-3766.2018.11.002

432   Suzuki T, Kasamatsu A, Miyamoto I, Saito T, Higo M, Endo-Sakamoto Y, Shiiba M, Tanzawa H, and Uzawa K.
433   2016. Overexpression of TMOD1 is associated with enhanced regional lymph node metastasis in human oral
434   cancer. *Int J Oncol* 48:607-612. 10.3892/ijo.2015.3305

435   Toyooka S, Tokumo M, Shigematsu H, Matsuo K, Asano H, Tomii K, Ichihara S, Suzuki M, Aoe M, Date H, Gazdar
436   AF, and Shimizu N. 2006. Mutational and epigenetic evidence for independent pathways for lung
437   adenocarcinomas arising in smokers and never smokers. *Cancer Res* 66:1371-1375. 10.1158/0008-5472.Can-
438   05-2625

439   Vargas AJ, and Harris CC. 2016. Biomarker development in the precision medicine era: lung cancer as a case study.
440   *Nat Rev Cancer* 16:525-537. 10.1038/nrc.2016.56

441   Vazquez F, Hastings G, Ortega MA, Lane TF, Oikemus S, Lombardo M, and Iruela-Arispe ML. 1999. METH-1, a
442   human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory
443   activity. *J Biol Chem* 274:23349-23357.

444   Vencken SF, Greene CM, and McKiernan PJ. 2015. Non-coding RNA as lung disease biomarkers. *Thorax* 70:501-

445   503. 10.1136/thoraxjnl-2014-206193

446   Walters S, Maringe C, Coleman MP, Peake MD, Butler J, Young N, Bergstrom S, Hanna L, Jakobsen E, Kolbeck K,
447   Sundstrom S, Engholm G, Gavin A, Gjerstorff ML, Hatcher J, Johannesen TB, Linklater KM, McGahan CE,
448   Steward J, Tracey E, Turner D, Richards MA, and Rachet B. 2013. Lung cancer survival and stage at
449   diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: a population-based study, 2004-
450   2007. *Thorax* 68:551-564. 10.1136/thoraxjnl-2012-202297

451   Wang Z, Cao CJ, Huang LL, Ke ZF, Luo CJ, Lin ZW, Wang F, Zhang YQ, and Wang LT. 2015. EFEMP1 promotes
452   the migration and invasion of osteosarcoma via MMP-2 with induction by AEG-1 via NF-kappaB signaling
453   pathway. *Oncotarget* 6:14191-14208. 10.18632/oncotarget.3691

454   Wei MM, and Zhou GB. 2016. Long Non-coding RNAs and Their Roles in Non-small-cell Lung Cancer. *Genomics*
455   *Proteomics Bioinformatics* 14:280-288. 10.1016/j.gpb.2016.03.007

456   Xia Y, Yeddula N, Leblanc M, Ke E, Zhang Y, Oldfield E, Shaw RJ, and Verma IM. 2012. Reduced cell proliferation
457   by IKK2 depletion in a mouse lung-cancer model. *Nat Cell Biol* 14:257-265. 10.1038/ncb2428

458   Xin X, Majumder M, Girish GV, Mohindra V, Maruyama T, and Lala PK. 2012. Targeting COX-2 and EP4 to control
459   tumor growth, angiogenesis, lymphangiogenesis and metastasis to the lungs and lymph nodes in a breast
460   cancer model. *Lab Invest* 92:1115-1128. 10.1038/labinvest.2012.90

461   Yang T, Qiu H, Bao W, Li B, Lu C, Du G, Luo X, Wang L, and Wan X. 2013. Epigenetic inactivation of EFEMP1 is
462   associated with tumor suppressive function in endometrial carcinoma. *PLoS One* 8:e67458.
463   10.1371/journal.pone.0067458

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

466   Yue W, Dacic S, Sun Q, Landreneau R, Guo M, Zhou W, Siegfried JM, Yu J, and Zhang L. 2007. Frequent inactivation
467   of RAMP2, EFEMP1 and Dutt1 in lung cancer by promoter hypermethylation. *Clin Cancer Res* 13:4336-
468   4344. 10.1158/1078-0432.Ccr-07-0015

469   Zhai X, and Li SJ. 2014. Methylation of RASSF1A and CDH13 genes in individualized chemotherapy for patients
470   with non-small cell lung cancer. *Asian Pac J Cancer Prev* 15:4925-4928.

471   Zhan P, Xi GM, Liu HB, Liu YF, Xu WJ, Zhu Q, Zhou ZJ, Miao YY, Wang XX, Jin JJ, Lv TF, and Song Y. 2017a.
472   Protein regulator of cytokinesis-1 expression: prognostic value in lung squamous cell carcinoma patients. *J*
473   *Thorac Dis* 9:2054-2060. 10.21037/jtd.2017.06.91

474   Zhan P, Zhang B, Xi GM, Wu Y, Liu HB, Liu YF, Xu WJ, Zhu QQ, Cai F, Zhou ZJ, Miu YY, Wang XX, Jin JJ, Li
475   Q, Qian LP, Lv TF, and Song Y. 2017b. PRC1 contributes to tumorigenesis of lung adenocarcinoma in
476   association with the Wnt/beta-catenin signaling pathway. *Mol Cancer* 16:108. 10.1186/s12943-017-0682-z

477   Zhao P, Wu J, Lu F, Peng X, Liu C, Zhou N, and Ying M. 2019. The imbalance in the complement system and its
478   possible physiological mechanisms in patients with lung cancer. *BMC Cancer* 19:201. 10.1186/s12885-019-
479   5422-x

480   Zhao X, Yang C, Wu J, and Nan Y. 2018. ADAMTS8 targets ERK to suppress cell proliferation, invasion, and
481   metastasis of hepatocellular carcinoma. *Onco Targets Ther* 11:7569-7578. 10.2147/ott.S173360

482   Zhong YH, Peng H, Cheng HZ, and Wang P. 2015. Quantitative assessment of the diagnostic role of CDH13 promoter
483   methylation in lung cancer. *Asian Pac J Cancer Prev* 16:1139-1143.

484   Zhu XJ, Liu J, Xu XY, Zhang CD, and Dai DQ. 2014. Novel tumor-suppressor gene epidermal growth factor-
485   containing fibulin-like extracellular matrix protein 1 is epigenetically silenced and associated with invasion

486   and metastasis in human gastric cancer. *Mol Med Rep* 9:2283-2292. 10.3892/mmr.2014.2135
487
488

The detail information of four GEO datasets

Table 1 The detail information of four GEO datasets

ID	Tissue	Platform	Normal	Tumor
GSE118370	LUAD	GPL570	6	6
GSE85841	LUAD	GPL20115	8	8
GSE43458	LUAD	GPL6244	30 (never-smoker)	40 (never-smoker)
GSE32863	LUAD	GPL6884	58	58

GEO: Gene Expression Omnibus; LUAD: Lung Adenocarcinoma

Top ten up- and down-regulated DEGs

Table 2 Top ten up- and down-regulated DEGs

Up-regulated DEGs				Down-regulated DEGs		
Gene symbol	**Log2FC**	**FDR**		**Gene symbol**	**Log2FC**	**FDR**
SPP1	3.08474	1.74E-33		FABP4	-3.37555	1.25E-47
OCIAD2	1.477746	1.13E-29		STX11	-2.01766	1.51E-44
ETV4	1.366623	5.38E-27		CAV1	-2.70539	6.04E-43
TOP2A	1.777989	3.24E-26		FHL1	-2.3099	5.77E-42
COL10A1	1.364488	3.74E-26		TEK	-2.30698	1.09E-40
PROM2	1.570558	7.77E-26		AGER	-2.69928	6.05E-40
MMP11	1.929847	1.03E-23		FMO2	-2.44791	1.42E-39
UBE2T	1.154061	5.78E-23		CRYAB	-1.98382	1.77E-39
ABCC3	1.444218	3.71E-22		GRK5	-1.4854	1.53E-38
BAIAP2L1	1.067336	4.15E-22		TMEM100	-2.928	4.79E-38

1   DEGs: Differentially Expressed Genes; FDR: False Discovery Rate; Log2FC: Log2(Fold Change)

KEGG enriched pathways of DEGs

Table 3 KEGG enriched pathways of DEGs

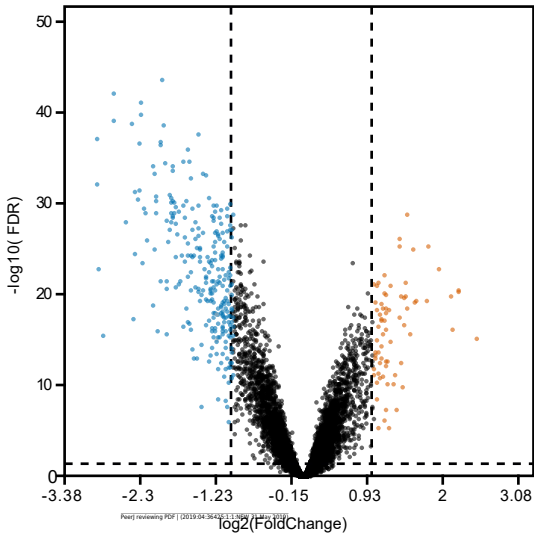
ID	Description	Count	p.adjust(FDR)
hsa04610	Complement and coagulation cascades	13	1.12E-05
hsa05144	Malaria	9	0.000271
hsa05020	Prion diseases	7	0.001406
hsa05418	Fluid shear stress and atherosclerosis	13	0.001921
hsa04933	AGE-RAGE signaling pathway in diabetic complications	10	0.007362
hsa04270	Vascular smooth muscle contraction	11	0.014433
hsa04657	IL-17 signaling pathway	9	0.014433
hsa04974	Protein digestion and absorption	8	0.039299
hsa04670	Leukocyte transendothelial migration	9	0.039299
hsa00982	Drug metabolism - cytochrome P450	7	0.039299

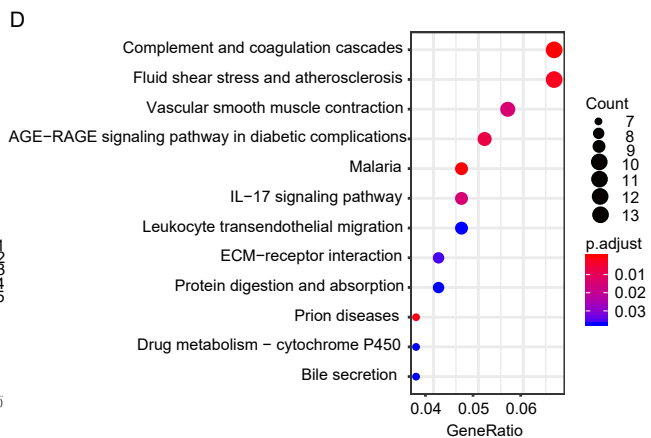
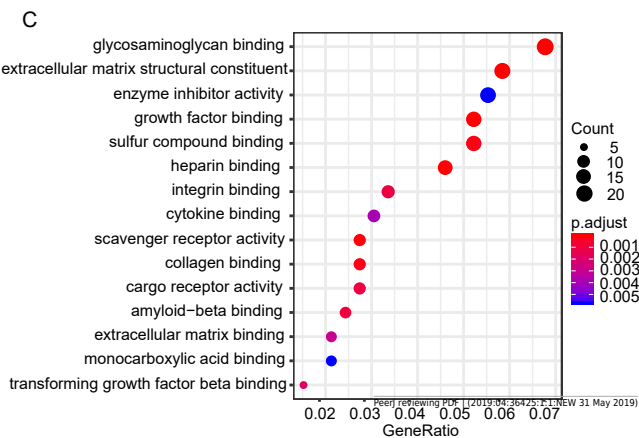
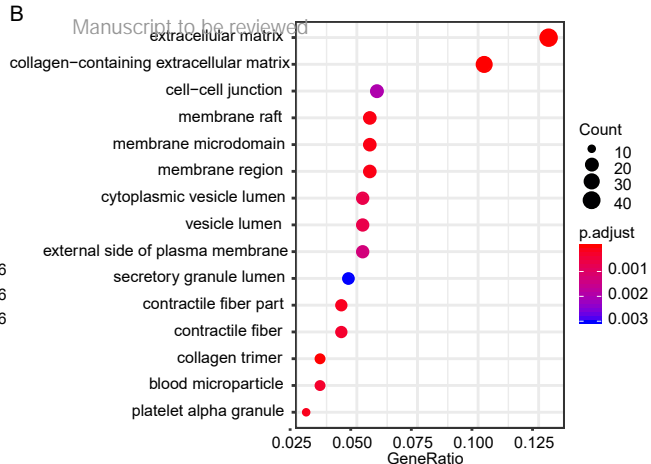
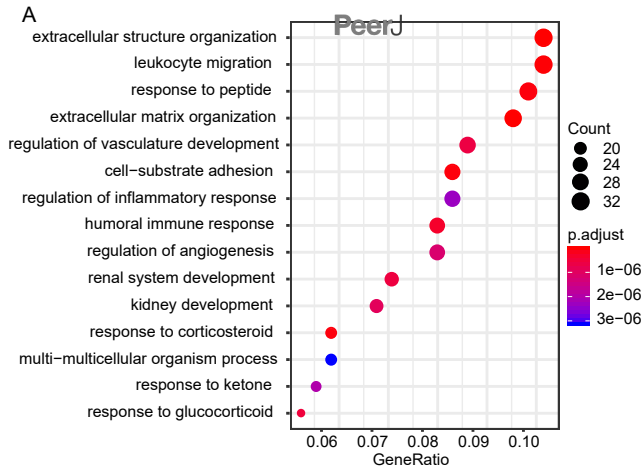
1 KEGG: Kyoto Encyclopedia of Genes and Genomes; DEGs: Differentially Expressed Genes; FDR: False

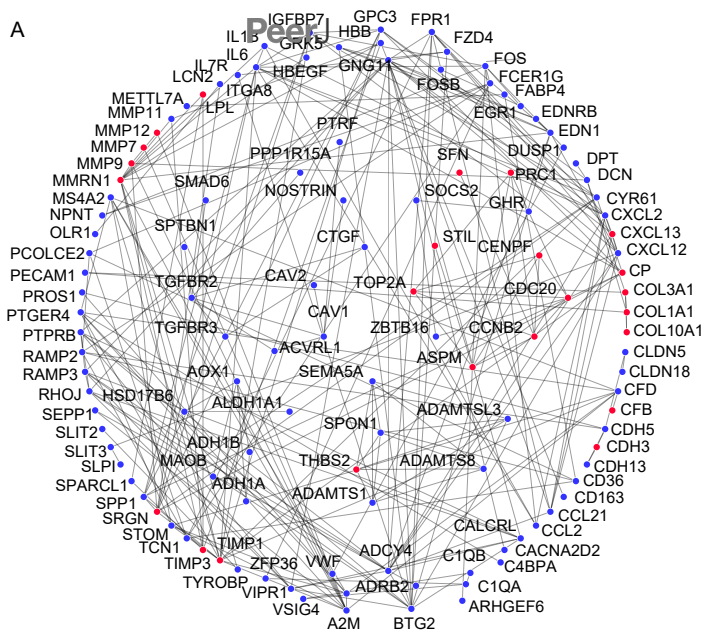
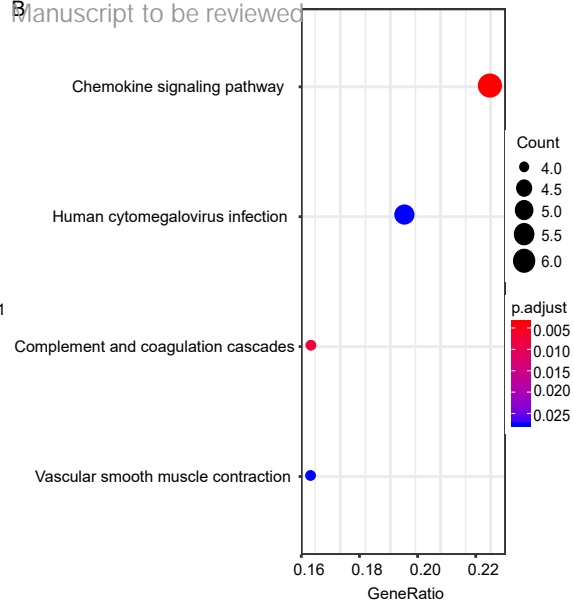
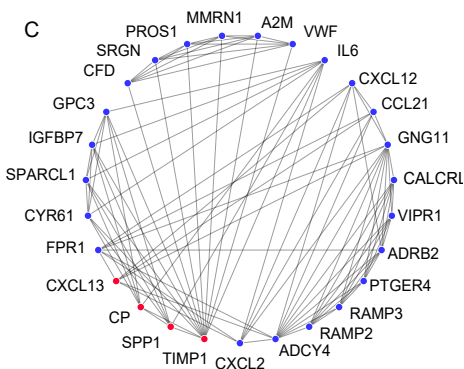
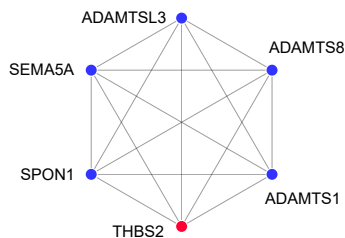
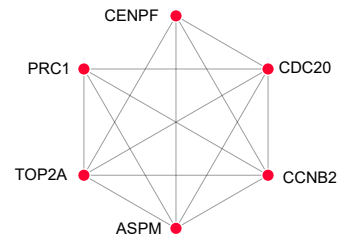
2 Discovery Rate

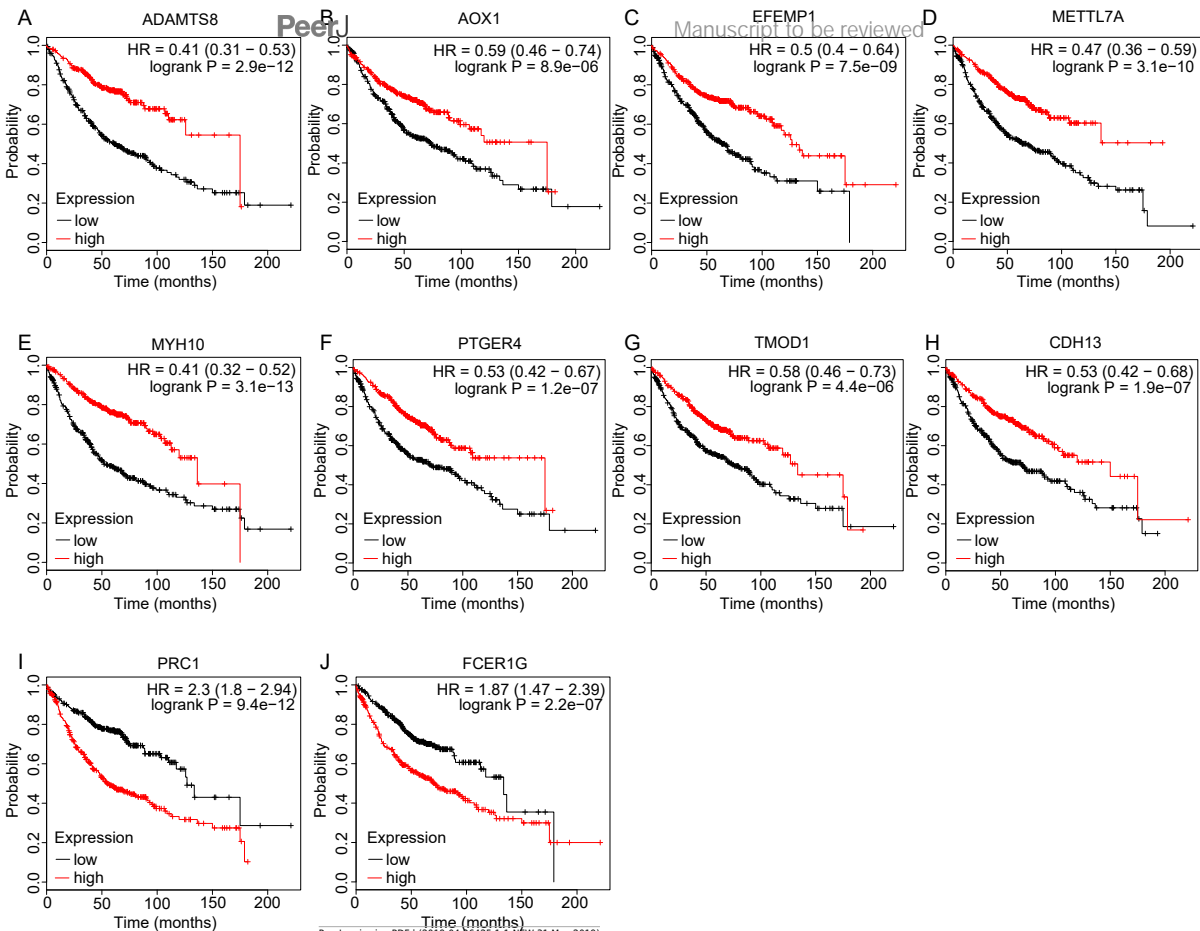
Volcano plot of the DEGs.

The vermilion and blue dots represent DEGs filtered based on the cut-off values of |log2FoldChange| > 1.0 and adjusted P value(FDR) < 0.05, while the black dots represent genes that are not satisfied the cut-off values of differential expression. The horizon dotted line indicates the position of -log10(FDR) = 0.05, and the vertical dotted lines indicate the positions of |log2FoldChange|=1.0. Vermilion : upregulation; Blue: downregulation. DEGs: differentially expressed genes. FDR:False Discovery Rate.

Volcano Plot of DEGS

GO and KEGG analysis of DEGs.

(A-C) The top 15 terms of GO categories of biological process (BP), cellular component (CC) and molecular function MF, respectively. (D) KEGG pathway analysis of DEGs, p.adjust(FDR) <0.05 was considered significantly. GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes;FDR:False Discovery Rate.

Protein-protein interaction network of DEGs and modules analysis.

(A) PPI network of DEGs; (B) The KEGG enrichment analysis of the genes in the top 3 modules. (C-E) the top 3 modules of PPI network. The red nodes represent the upregulated DEGs. The green nodes represent the downregulated DEGs. PPI: protein-protein interaction; DEG: differentially expressed gene; KEGG: Kyoto Encyclopedia of Genes and Genomes.

Overall survival analyses of hub genes.

(A-J) The overall survival analyses of hub genes were performed using Kaplan- Meier Plotter online platform. Logrank P<0.05 was considered statistically significant.

Oncomine analysis of cancer vs. normal tissue of ADAMTS8 , METTL7A , MYH10 and PRC1.

Heat maps of ADAMTS8(A), METTL7A(B), MYH10(C) and PRC1(D) gene expression in lung adenocarcinoma samples vs. normal tissues. 1. Lung Adenocarcinoma vs. Normal, Hou Lung (Hou et al. 2010) ; 2. Lung Adenocarcinoma vs. Normal, Landi Lung (Landi et al. 2008) ; 3. Lung Adenocarcinoma vs. Normal, Selamat Lung (Selamat et al. 2012) ; 4. Lung Adenocarcinoma vs. Normal, Su Lung (Su et al. 2007) .

A

Median Rank	P- Value	Gene
79.5	7.34E-28	ADAMTS8

1 2 3 4

1. P-value=5.54E-27 FoldChange= -6.612
2. P-value=1.47E-27 FoldChange= -2.990
3. P-value=9.12E-38 FoldChange= -2.848
4. P-value=1.81E-11 FoldChange= -8.009

B

Median Rank	P- Value	Gene
547.5	1.13E-9	METTL7A

1 2 3 4

1. P-value=8.55E-13 FoldChange= -2.482
2. P-value=2.26E-28 FoldChange= -3.213
3. P-value=8.11E-17 FoldChange= -2.522
4. P-value=2.25E-9 FoldChange= -3.382

C

Median Rank	P- Value	Gene
235.5	1.07E-23	MYH10

1 2 3 4

1. P-value=4.97E-17 FoldChange= -2.673
2. P-value=2.15E-23 FoldChange= -2.649
3. P-value=1.79E-26 FoldChange= -2.797
4. P-value=1.72E-12 FoldChange= -3.491

D

Median Rank	P- Value	Gene
220.5	5.56E-7	PRC1

1 2 3 4

1. P-value=2.41E-17 FoldChange= 4.260
2. P-value=1.88E-14 FoldChange= 2.661
3. P-value=1.16E-19 FoldChange= 2.677
4. P-value=1.11E-6 FoldChange= 2.610

1 5 10 25   25 10 5 1   Not measured

%

The association between the expression of selected hub genes and tumor stage.

(A-D) The expressions of ADAMTS8, METTL7A, MYH10, PRC1 were correlated with tumor stage in the Okayama Lung dataset. 0: No value; 1: Stage I; 2: Stage II.