

Identification of nine key genes by bioinformatics analysis for predicting poor prognosis in smoking-induced lung adenocarcinoma

Chuanli Ren^{*,‡,1} , Weixiu Sun^{‡,2}, Xu Lian² & Chongxu Han¹

¹Department of Laboratory Medicine, Clinical Medical College, Yangzhou University, Yangzhou, PR China

²Clinical Laboratory Diagnostics, Clinical Medical College, Dalian Medical University, Dalian, PR China

*Author for correspondence: renchl@163.com

‡Chuanli Ren and Weixiu Sun contribute equally to this work

Aim: To screen and identify key genes related to the development of smoking-induced lung adenocarcinoma (LUAD). **Materials & methods:** We obtained data from the GEO chip dataset GSE31210. The differentially expressed genes were screened by GEO2R. The protein interaction network of differentially expressed genes was constructed by STRING and Cytoscape. Finally, core genes were screened. The overall survival time of patients with the core genes was analyzed by Kaplan–Meier method. Gene ontology and Kyoto encyclopedia of genes and genomes bioaccumulation was calculated by DAVID. **Results:** Functional enrichment analysis indicated that nine key genes were actively involved in the biological process of smoking-related LUAD. **Conclusion:** 23 core genes and nine key genes among them were correlated with adverse prognosis of LUAD induced by smoking.

First draft submitted: 23 November 2019; Accepted for publication: 17 March 2020; Published online: 27 April 2020

Keywords: bioinformatics analysis • core genes • GSE31210 • lung adenocarcinoma • smoking

Lung cancer is the most common cause of cancer mortality worldwide, accounting for 18.4% of all cancer deaths [1]. Lung adenocarcinoma (LUAD) is the most common type of lung cancer, accounting for about 40% of all lung cancers [2]. Studies have shown that smokers have a higher frequency of genomic alterations in lung cancer than nonsmokers have [3,4]. In addition, a genome analysis found that some somatic cell mutations in LUAD were associated with smoking phenotype [5]. The *KRAS* mutation in LUAD is associated with smoking, but little is known about the association between tobacco exposure and oncogenes [6]. Therefore, screening for abnormal genes related to smoking and development of LUAD is important to clarify the pathogenesis of LUAD and search for early warning markers and treatment targets.

In this study, we extracted data from the gene chip GSE31210 in the GEO database and compared the differentially expressed genes (DEGs) in LUAD and normal tissues in smoking and nonsmoking groups. To provide theoretical guidance for clarifying the pathogenesis of smoking-related LUAD and screening molecular markers, the core genes screened out by various bioinformatics methods were closely related to the occurrence of LUAD.

Materials & methods

Data acquisition & genetic screening

Gene expression profile data GSE31210 were downloaded from the GEO database [7], which is the Affymetrix human genome U1332.0 chip [HG133-Plus.2] (Agilent GPL570 platform). The GSE31210 dataset included 246 lung tissues, 226 fresh frozen tissue specimens of LUAD primary lung adenocarcinoma ADCs of pathological stage I–II and 20 normal lung tissues. Expression profiles in of 226 LUADs (127 with *EGFR* mutation, 20 with *KRAS* mutation, 11 with *EML4-ALK* fusion and 68 triple negative cases). There were 111 LUAD tissues and 12 normal tissues in the smoking group (ever-smoker [ES]) and 115 LUAD tissues and eight normal tissues in

the nonsmoking group (never-smoker [NS]). Gene expression data from the ES and NS groups were analyzed by online tool GEO2R [8] with $p < 0.01$ and $|\logFC| > 2$. Venn software [9] was used to obtain upregulated and downregulated genes in each group. DEGs with $\logFC > 0$ were considered to be upregulated genes, while DEGs with $\logFC < 0$ were considered downregulated genes.

Protein–protein interaction network & module analysis

Data analysis software STRING [10] was used to analyze the interaction between differentially encoded proteins in the ES and NS groups (confidence > 0.4 , maximum value of interaction = 0) and a protein–protein interaction (PPI) network was established. Through the MCODE plug-in of Cytoscape 3.6.0 software, the most closely connected modules were selected from the constructed PPI network for further analysis (setting parameters as degree cut-off = 2, node score = 0.2, k-core = 2, maximum depth = 100) and the genes contained in the modules were core genes.

Survival analysis & verification of core genes

Kaplan–Meier plotter [11] was used to evaluate the effect of core genes on survival ($p < 0.05$). To evaluate smoking as an influencing factor, when analyzing ES data, we excluded subjects who had never smoked. When analyzing NS data, we selected only subjects who had never smoked. In GEPIA [12], the expression level of core genes was verified by $|\logFC| > 1$ and $p < 0.01$.

Functional enrichment of DEGs

GO function annotation and KEGG pathway enrichment of core genes were carried out through the online bioinformatics analysis website DAVID [13]. The biological process, cell composition, molecular function and tumor-related pathways of these genes were analyzed by p value was less than 0.05.

Construction of core gene–miRNA regulatory network

The network tool MiRDB [14] was used to select miRNAs that may regulate the core genes with a score of ≥ 80 . We constructed the core gene–miRNA regulatory network using Cytoscape version 3.6.0 software.

Results

Screening DEGs in LUAD & normal tissues

Through the online tools GEO2R and Venn analysis, 637 significant DEGs with $>$ fourfold difference in expression were detected in the ES group, including 249 upregulated and 388 downregulated genes (Figure 1A). We detected 336 significant DEGs in the NS group, including 131 upregulated and 205 downregulated genes (Figure 1B). There were 87 common differences between the two groups in the upregulated genes (Figure 1C) and 164 common differences in the downregulated genes (Figure 1D).

Construction of PPI network & screening of core genes

The 637 DEGs in the ES group and 336 DEGs in the NS group were imported into the STRING website to construct their respective PPI networks (Figure 2A & C). The most closely connected modules in the PPI network were extracted using the MCODE plug-in in Cytoscape software. The highest-scoring module in the ES group contained 38 genes and a total of 656 interactions were formed, which were the core genes of the ES group (Figure 2B). The module with the highest connectivity in the NS group contained 11 core genes (Figure 2D). The results showed that there were nine genes, including *UBE2T* and *EXO1*, in the core genes, which indicated that the genetic mechanism of LUAD may not be related to smoking status. *CDT1* and *ERCC6L* existed independently in the NS group, while 29 core genes, such as *TPX2*, *MYBL2* and *CDC6*, existed independently in the ES group (Table 1). These results indicate that smoking can induce significant molecular changes and contribute to the early development of lung cancer.

Survival analysis & GEPIA validation of core genes in ES group

To investigate further the effects of smoking on the development of LUAD, we used the Kaplan–Meier plotter to determine the survival data of the core genes in the ES group. Patients with 24 out of the 29 genes specific to ES had significantly low survival rates (Figure 3 & Table 2), except for nine DEGs in both groups and the 24 core genes were highly expressed in the ES group. GEPIA was used to verify the different gene expression levels between

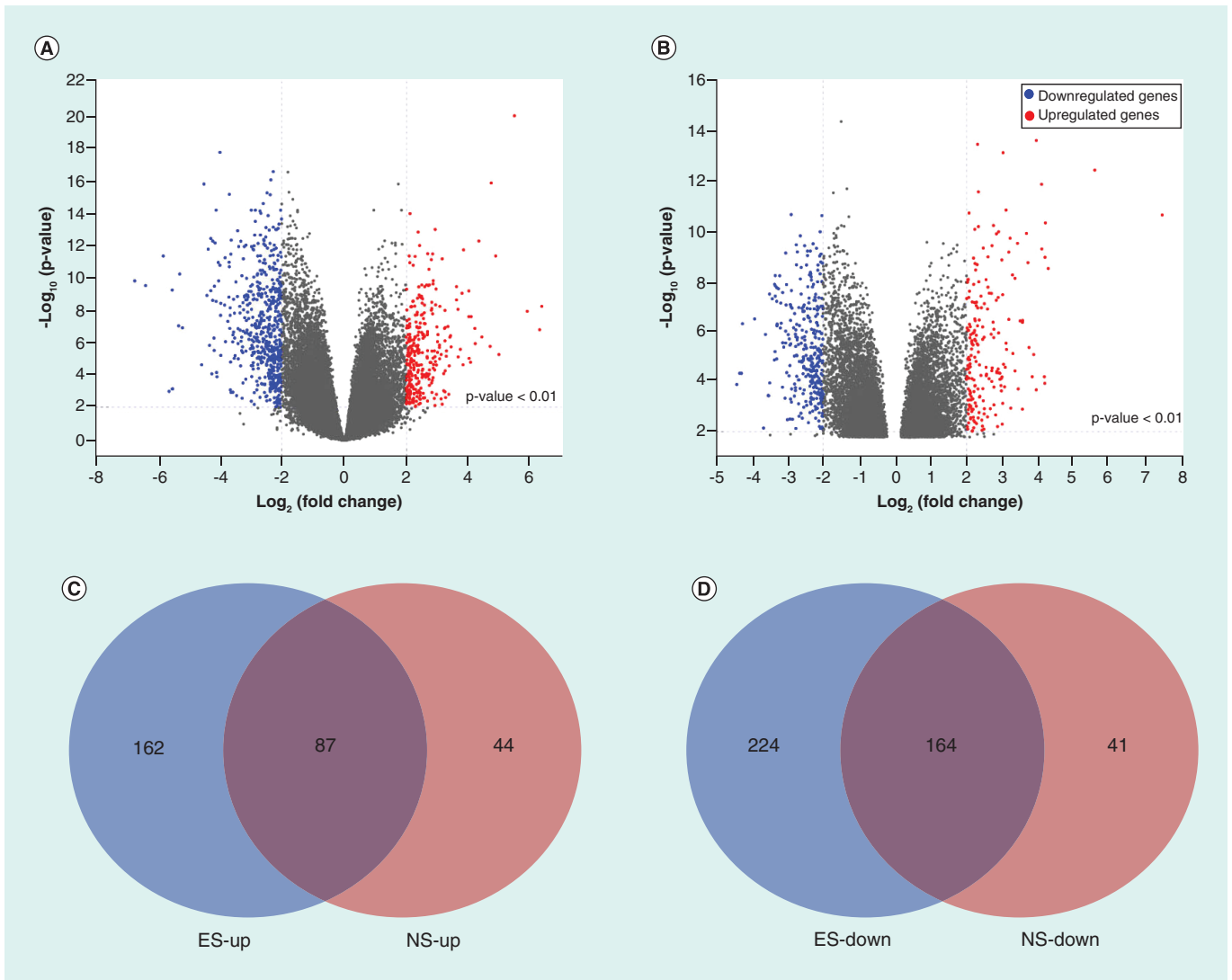


Figure 1. Identification of differentially expressed genes in lung adenocarcinoma and normal tissues. (A) 637 significant DEGs in ES group. **(B)** 336 significant DEGs in NS group. The red represents the selected upregulated genes and blue selected downregulated genes (P2). **(C)** 87 DEGs were upregulated in both the ES and NS groups (logFC > 0). **(D)** 164 DEGs were downregulated in both groups (logFC < 0). Different colors represent different groups. DEG: Differentially expressed gene; ES: Ever-smokers; NS: Never-smokers.

Table 1. Common and specific core genes in ever-smokers and never-smokers groups.

Category	Genes
Common to both groups	<i>UBE2T, EXO1, TOP2A, CDCA7, HMMR, ANLN, RAD54L, DEPDC1, CDCA8</i>
Specific to ES	<i>RRM2, BUB1, ESCO2, NUF2, E2F8, NCAPG, MYBL2, DEPDC1B, TPX2, CENPF, KIAA0101, NCAPH, DIAPH3, KIF4A, CDC6, MELK, NEIL3, DLGAP5, CENPM, BUB1B, MKI67, CEP55, CCNA2, CCNB2, CDC25C, CDC45, KIF2C, BIRC5, ASPM</i>
Specific to NS	<i>CDT1, ERCC6L</i>

ES: Ever-smokers; NS: Never-smoker.

Table 2. Prognosis of 38 core genes in ever-smokers group.

Category	Genes
Genes with significantly worse survival (p < 0.05)	<i>UBE2T, EXO1, TOP2A, CDCA7, HMMR, RAD54L, DEPDC1, ANLN, CDCA8, RRM2, TPX2, BUB1, E2F8, NCAPG, CENPF, CDC6, KIAA0101, NCAPH, MELK, NEIL3, ASPM, BUB1B, MKI67, CEP55, CCNA2, DLGAP5, KIF4A, CCNB2, MYBL2, CDC25C, KIF2C, BIRC5, CENPM</i>
Genes without significantly worse survival (p > 0.05)	<i>ESCO2, NUF2, DIAPH3, CDC45, DEPDC1B</i>

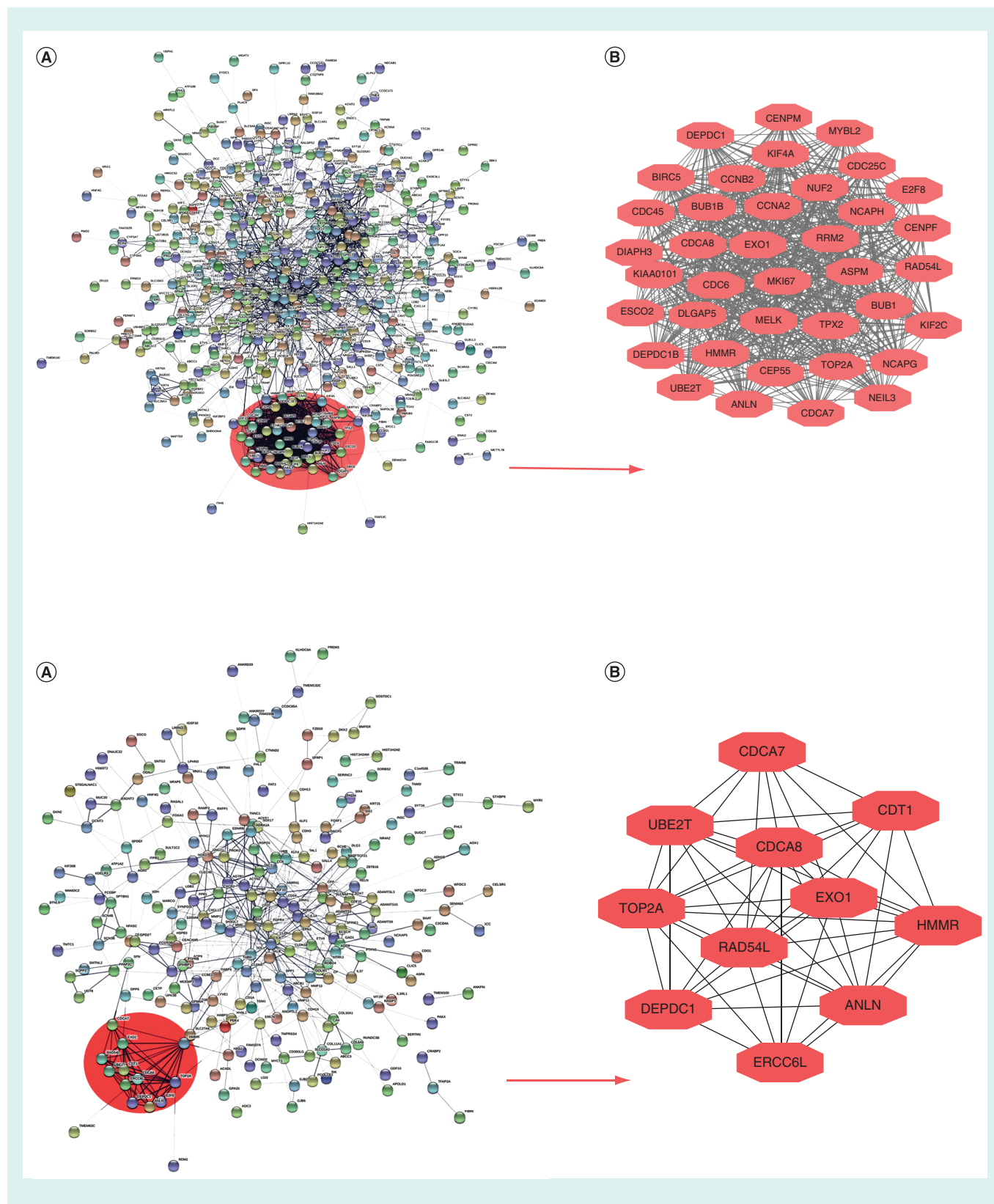


Figure 2. Protein–protein interaction network constructed by STRING database and module analysis in ever-smokers and never-smokers groups. (A) There were 637 differentially expressed genes in the ES group. **(B)** The top module selected from the ES group PPI network included 38 core genes. **(C)** There were 336 differentially expressed genes in the NS group. **(D)** The top module selected from the NS group PPI network included 11 core genes. The network was formed after removing independent nodes. ES: Ever-smokers; NS: Never-smokers; PPI: Protein protein interaction.

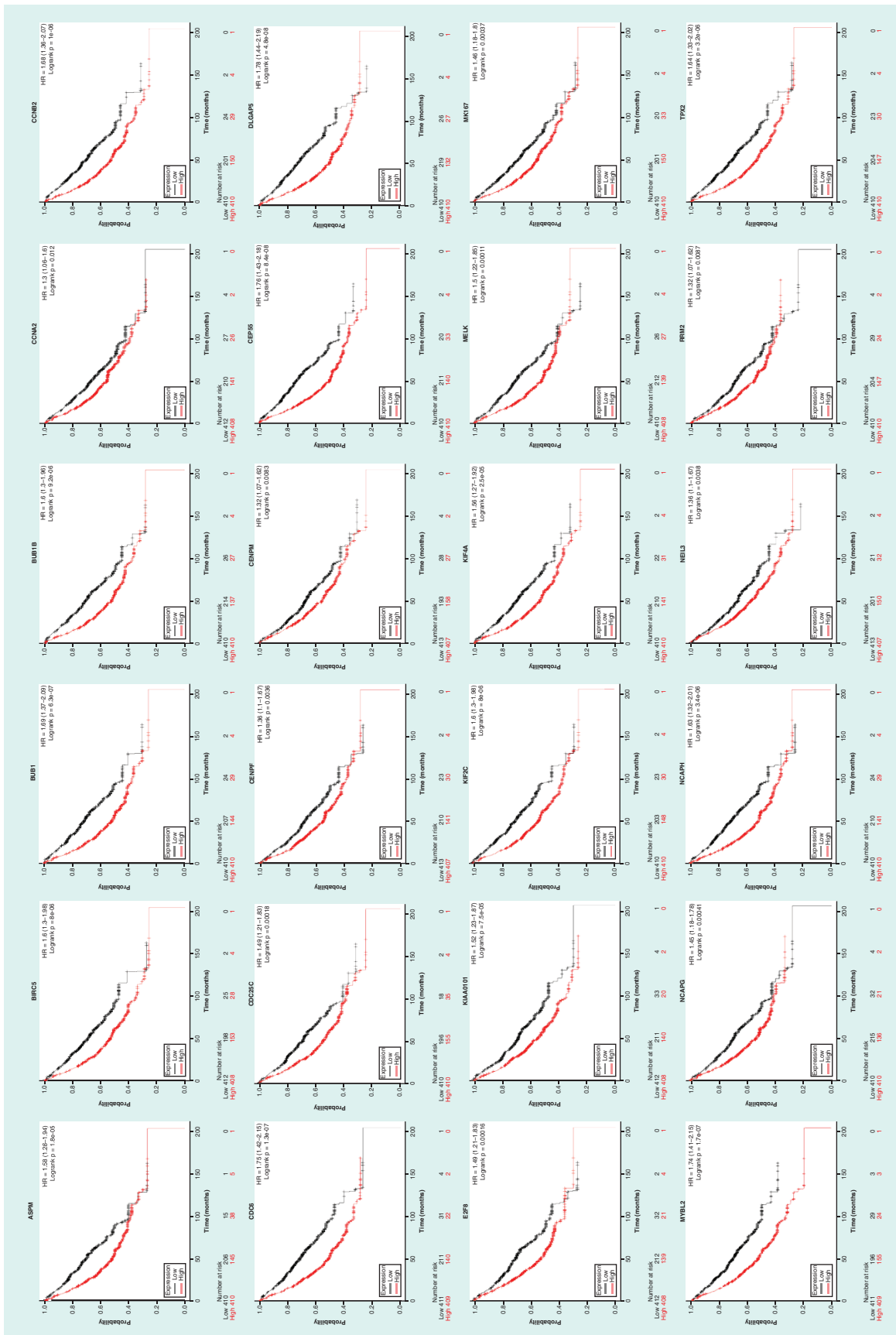


Figure 3. Kaplan–Meier survival analysis of the core genes specific to ever-smokers group. Patients with 24 out of the 29 genes had significantly worse prognosis ($p < 0.05$).

Table 3. GO analysis of 23 core genes.

ID & term	Count	p-value	Genes
GO_BP.ID:0090307 mitotic spindle assembly	3	5.28 E ⁻⁰⁴	<i>TPX2, BIRC5, MYBL2</i>
GO_BP.ID:0051301 cell division	3	1.79 E ⁻⁰³	<i>CDC6, NCAPH, BIRC5</i>
GO_BP.ID:0007076 Mitotic chromosome condensation	2	1.51 E ⁻⁰²	<i>NCAPH, NCAPG</i>
GO_BP.ID:0007094 mitotic spindle assemblycheckpoint	2	1.64 E ⁻⁰²	<i>BUB1, BUB1B</i>
GO_BP.ID:0000086 G2/M transition of mitotic cell cycle	2	3.53 E ⁻⁰²	<i>BIRC5, CDC25C</i>
GO_CC.ID: 0005634 nucleus	10	2.10 E ⁻⁰³	<i>CCNB2, RRM2, E2F8, DLGAP5, KIAA0101, ASPM, MELK, BIRC5, CDC25C, MYBL2</i>
GO_CC.ID: 0000796 condensin complex	2	5.45 E ⁻⁰³	<i>NCAPH, NCAPG</i>
GO_CC.ID: 0016020 membrane	5	2.44 E ⁻⁰²	<i>KIF2C, NCAPH, MKI67, NCAPG, BUB1</i>
GO_MF.ID:0005524 ATP binding	5	3.26 E ⁻⁰²	<i>KIF2C, CDC6, KIF4A, BUB1, MELK</i>

BP: Biological processes; CC: Cell composition; MF: Molecular function.

Table 4. Enrichment of 23 core genes KEGG pathways.

Pathway ID	Term	Count	p-value	Genes
cfa04110	Cell cycle	6	3.78 E ⁻⁰⁸	<i>CDC6, CCNB2, BUB1, BUB1B, CDC25C, CCNA2</i>
cfa04914	Progesterone-mediated oocyte maturation	4	6.74 E ⁻⁰⁵	<i>CCNB2, BUB1, CDC25C, CCNA2</i>
cfa04114	Oocyte meiosis	3	5.05 E ⁻⁰³	<i>CCNB2, BUB1, CDC25C</i>

LUAD and normal lung tissues in the ES group. 23 of the 24 genes were highly expressed in LUAD samples, in contrast to normal lung tissue (Figure 4).

Functional enrichment of core genes in ES group

The results of GO functional enrichment analysis showed that the biological processes of core genes related to smoking were cell division and mitosis; cell composition was mainly nucleus, condensin complex and membrane and main molecular function was ATP binding (Table 3). There were nine key genes involved in the biological processes of LUAD, including *TPX2*, *BIRC5* and *MYBL2*, the main participant of mitosis; *BUB1* and *BUB1B*, mitotic spindle assembly checkpoints and *NCAPH* and *NCAPG* involved in mitotic chromosome condensation. *CDC6* was mainly involved in cell division, while *CDC25C* was related to G2/M transition of the mitotic cell cycle.

KEGG pathway analysis showed that six core genes were enriched in cell-cycle-related pathways, four core genes were enriched in pathways related to progesterone-mediated oocyte maturation and three core genes were enriched in oocyte-meiosis-related pathways (Table 4). KEGG results were used to analyze nine key genes involved in the biological processes of LUAD. Four genes (*BUB1*, *BUB1B*, *CDC6* and *CDC25C*) were enriched in cell-cycle-related pathways; *BUB1*, *BUB1B* and *CDC25C* were concentrated in G2/M phase.

Construction of a core gene-miRNA network

The miRNA data that may regulate smoking-related core genes were downloaded by MiRDB. The core gene-miRNA network map constructed after MCODE input showed that there was an interaction between 80 miRNA and 18 target genes. In this network, *CEP55* (connection = 18), *RRM2* (connection = 12), *NCAPG* (connection = 9) and *MYBL2* (connection 9) were highly controlled by the miRNA network. In addition, miR-19a-3p/miR-19b-3p regulated expression of *CEP55*, *CCNA2* and *E2F8* (Figure 5).

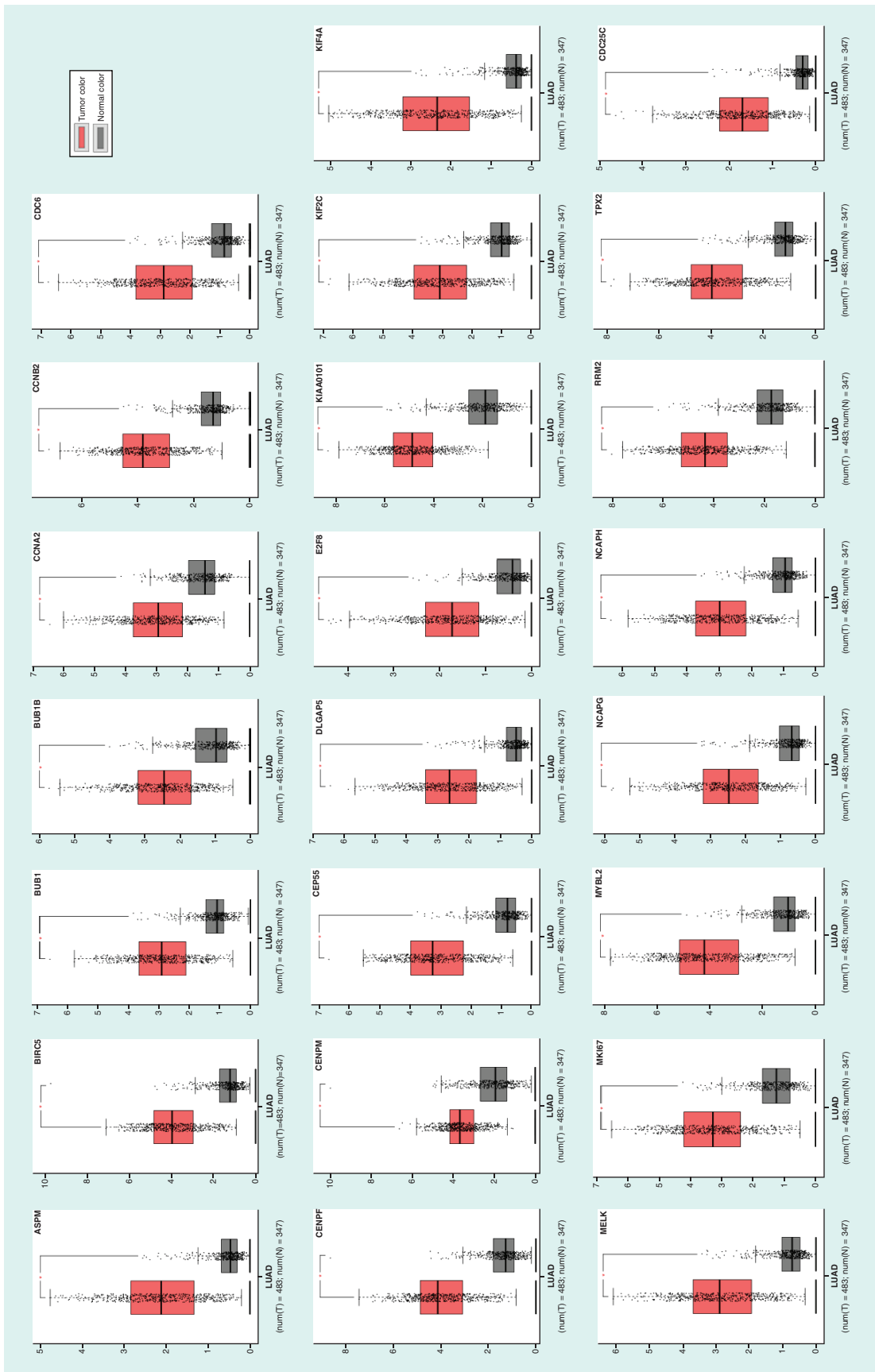


Figure 4. Differences in expression of core genes specific to ever-smokers group in lung adenocarcinoma tissues and normal specimens. 23 out of 24 core genes showed high expression in lung adenocarcinoma specimens ($|\text{LogFC}| > 1$ and $p < 0.01$).

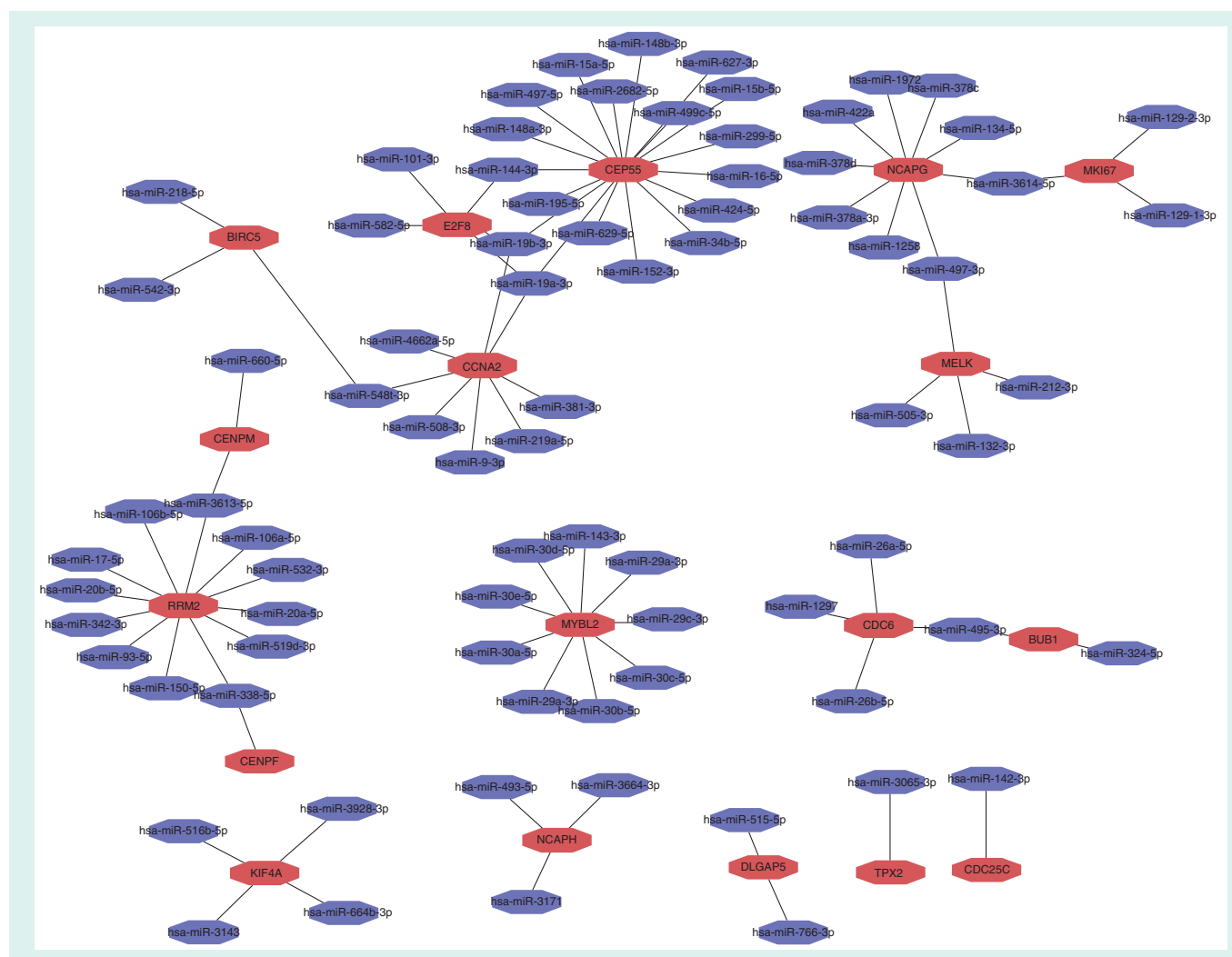


Figure 5. Core gene-miRNA networks in ever-smokers group. Red represents the core genes, blue the miRNAs that interacted with the core genes.

Discussion

Tobacco smoke contains >40 carcinogens that can cause direct or indirect damage to respiratory epithelium and its genes [15]. So far, many studies have been conducted to explore the effects of smoking on gene expression in lung cancer, but our understanding is still limited.

In this study, we identified the DEGs between the ES and NS groups and established PPI networks. We extracted the most closely related module gene in the PPI network as the key gene. Then, the key genes of the two groups were crossed; nine DEGs were identified as smoking independent and there were 29 ES-specific and two NS-specific genes. To explore the genomic changes caused by smoking, the survival of patients with the 29 genes unique to the ES group was shown by Kaplan-Meier curve. The higher the expression of 24 of the genes, the shorter was the overall survival time. At the same time, these 24 genes were highly expressed in the ES group. Subsequently, GEPIA confirmed that expression of 23 of these 24 genes in LUAD was significantly higher than in normal lung tissue.

The changes in the expression levels of these 23 characteristic genes may play a key role in the development of smoking-related LUAD. GO and KEGG help us to gain a better understanding of gene function. Nine key genes were found to be closely related to the biological process of smoking-related LUAD, including *TPX2*, *MYBL2*, *CDC6*, *NCAPH*, *NCAPG*, *BIRC5*, *BUB1*, *BUB1B* and *CDC25C*. Among them, *TPX2* is a multidomain protein that plays an important role in spindle formation, microtubule nucleation, and AURKA regulation [16]. It has been reported that *TPX2* is overexpressed in smoking-related LUAD and the high level of mRNA expression is related

to poor prognosis of LUAD, which confirms the conclusion of our study [17]. *MYBL2* is a transcription factor of the Myb proto-oncogene family, which is widely expressed in a variety of cancers and upregulated to facilitate proliferation of cancer cells [18–20]. A study of human embryonic lung fibroblasts has shown that *MYBL2* can delay cell aging by repressing transcription of p16^{INK4a} (CDK4/6 inhibitor, which plays an important role in premature senescence), while knockout of *MYBL2* increases G1 phase arrest. That is, *MYBL2* weakens the replicative ability of cells [21]. In the replication of chromosomal DNA in eukaryotic cells, *CDC6* is essential for the establishment and maintenance of precomplexes [22]. In addition, *CDC6* can promote apoptosis. Lung function continues to decline in patients with chronic obstructive pulmonary disease after quitting smoking, which is related to the apoptosis-promoting characteristics of mutated *CDC6* [23]. It is worth noting that another Brazilian researcher detected that missense mutations in *CDC6* were found only in nonsmokers compared with smokers with non-small-cell lung cancer [24]. Whether this is consistent with the high expression of *CDC6* in the ES group in our study needs further study. In a recent study, microarray data sets identified *NCAPH* as a key gene in non-small-cell lung cancer patients in western countries [25]. *NCAPG*, also known as *CAPG*, is an actin-binding protein that plays a key role in regulating cell migration and invasion [26]. Some researchers have found that hypoxia can increase expression of *CapG* in lung cancer cell lines A549 and H358; *CapG* contributes to the migratory ability of lung cancer cells [27]. Compared with nonsmoking patients, patients with a history of smoking are more likely to develop hypoxia and smoking lung cancer patients with high expression of *NCAPG* have poor prognosis. *BIRC5* is a gene for physical examination of mitotic spindles, which not only regulates mitosis and apoptosis, but also participates in the pathological process [28]. It has been shown that expression of *BIRC5* is upregulated in breast, pancreatic, liver and esophageal cancers and others [29–32]; its high expression indicates poor clinical prognosis. In addition, Cao *et al.* have found that high level of *BIRC5* expression is related to poor prognosis of LUAD [33]. *BUB1* can be used as a spindle assembly checkpoint to control chromosome separation [34]. Pinto *et al.* have proved that overexpression of *BUB1* is closely related to Furhman grading and genomic replication of tumors [35]. The role of *BUB1* in cancer cells is still controversial. It was recently reported that overexpression of *BUB1B* can accelerate progression and recurrence of several cancers, but low expression leads to poor survival and metastasis of colon adenocarcinoma and some lung cancer cell lines [36]. In the present study, *BUB1B* was highly expressed in smokers, who had poor survival compared with nonsmokers. Although *BUB1B* needs further study, it is suggested that it may not play a major role in smoking-related LUAD, which shows that smoking-related LUAD development is a complex process involving multiple factors. Cigarette smoke is toxic to *Schizosaccharomyces pombe* through the *MAPK* pathway and S phase checkpoint and the common effect molecule of the two pathways is *CDC25*, cell cycle regulatory phosphatase [37]. Similarly, we showed that *CDC25C* is involved in the biological process of smoking-related LUAD through G2/M transition of the mitotic cell cycle.

Our analysis shows that most of the nine key genes involved in the biological pathway of smoking-related LUAD affect the occurrence and development of LUAD through the cell cycle. In fact, previous studies have shown that nicotine and its derivatives can mediate mitosis through a variety of molecular mechanisms [38]. The other 14 core genes may affect the progress of LUAD through some link or auxiliary action with these nine genes. In addition, in the constructed miRNA network, miR-19a-3p/miR-19b-3p regulated expression of three core genes, *CEP55*, *CCNA2* and *E2F8*, which could provide more options for early diagnosis and treatment of smoking-related LUAD.

Unlike previous studies on lung cancer patients with or without a history of smoking, we used an existing public database to select a large sample (including 246 samples of lung tissue), combined with analytical tools. The molecular changes related to the pathogenesis of LUAD in patients with a history of smoking were explored at the genetic level. The purpose of our analysis was to reveal the key gene changes induced by smoking in the development of LUAD and the possible mechanism, so as to provide new ideas and experimental entry points for treatment of lung and other cancers. Although our data were grouped according to smoking or not, we cannot completely exclude the influence of other variables. Therefore, it is necessary to find more samples and carry out functional experiments *in vivo* and *in vitro* to verify our results.

Conclusion

The 23 core genes screened by bioinformatics analysis – especially the nine key genes – are associated with poor prognosis of smoking-related LUAD, but the specific pathway and regulatory mechanism still need to be clarified.

Future perspective

In recent years, bioinformatics research can analyze huge and complex genomic information and it has a wide range of applications. In this article, a series of bioinformatics tools are used to analyze the tissue information from GEO database, which will help us to understand the occurrence and development of smoking-related LUAD. In addition, more and more studies combine the results of bioinformatics analysis with clinical pathology and apply experimental technology to verify. Finally open up a new path for the diagnosis and treatment of human diseases.

Author contributions

C Ren contributed towards topic selection, writing and revising the manuscript; W Sun contributed towards experiment design and writing; X Lian contributed towards checking information and drawing and C Han contributed towards revising the manuscript.

Financial & competing interests disclosure

This study was supported by the National Nature Science Foundation of China (81573220), the Foundation for Young Medical Scholar in Jiangsu Province (QNRC2016325; QNRC2016333) and Foundation for Young Medical Scholars in Yangzhou (RCC201835). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

We thank C Kerr from Liwen Bianji, Edanz Editing China (www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68(6), 394–424 (2018).
2. Zappa C, Mousa SA. Non-small-cell lung cancer: current treatment and future advances. *Transl. Lung Cancer Res.* 5(3), 288–300 (2016).
3. Govindan R, Ding L, Griffith M *et al.* Genomic landscape of non-small-cell lung cancer in smokers and never-smokers. *Cell* 150(6), 1121–1134 (2012).
4. Imielinski M, Berger AH, Hammerman PS *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 150(6), 1107–1120 (2012).
5. Shi J, Hua X, Zhu B *et al.* Somatic genomics and clinical features of lung adenocarcinoma: a retrospective study. *PLoS medicine* 13(12), e1002162 (2016).
6. Castelletti N, Kaiser JC, Simonetto C. Risk of lung adenocarcinoma from smoking and radiation arises in distinct molecular pathways. *Carcinogenesis* 40(10), 1240–1250 (2019).
7. NCBI. Gene Express Omnibus. www.ncbi.nlm.nih.gov/geo
8. NCBI. GEO2R. www.ncbi.nlm.nih.gov/geo/geo2r/
9. Bioinformatics and Evolutionary Genomics. <http://bioinformatics.psb.ugent.be/webtools/Venn/>
10. STRING. <https://string-db.org/cgi/input.pl>
11. Kaplan Meier plotter. <http://kmplot.com/analysis>
12. Gene Expression Profiling Interactive Analysis. <http://gepia.cancer-pku.cn/>
13. DAVID Bioinformatics Database. Welcome to DAVID 6.8 (2020). <https://david.ncifcrf.gov/>
14. miRDB. <http://mirdb.org>
15. Alavanja MC. Biologic damage resulting from exposure to tobacco smoke and from radon: implication for preventive interventions. *Oncogene* 21(48), 7365–7375 (2002).
16. Petry S, Groen AC, Ishihara K, Mitchison TJ, Vale RD. Branching microtubule nucleation in *Xenopus* egg extracts mediated by augmin and TPX2. *Cell* 152(4), 768–777 (2013).
17. Zhang MY, Liu XX, Li H, Li R, Liu X, Qu YQ. Elevated mRNA levels of AURKA, CDC20 and TPX2 are associated with poor prognosis of smoking related lung adenocarcinoma using bioinformatics analysis. *Int. J. Med. Sci.* 15(14), 1676–1685 (2018).
18. Yu R, Li C, Lin X *et al.* Clinicopathologic features and prognostic implications of MYBL2 protein expression in pancreatic ductal adenocarcinoma. *Pathol. Res. Pract.* 213(8), 964–968 (2017).

19. Zhang X, Lv QL, Huang YT, Zhang LH, Zhou HH. Akt/FoxM1 signaling pathway-mediated upregulation of MYBL2 promotes progression of human glioma. *J. Exp. Clin. Cancer Res.* 36(1), 105 (2017).
20. Calvisi DF, Simile MM, Ladu S *et al.* Activation of v-Myb avian myeloblastosis viral oncogene homolog-like2 (MYBL2)-LIN9 complex contributes to human hepatocarcinogenesis and identifies a subset of hepatocellular carcinoma with mutant p53. *Hepatology* 53(4), 1226–1236 (2011).
21. Huang Y, Wu J, Li R *et al.* B-MYB delays cell aging by repressing p16 (INK4alpha) transcription. *Cell. Mol. Life Sci.* 68(5), 893–901 (2011).
22. Cocker JH, Piatti S, Santocanale C, Nasmyth K, Diffley JF. An essential role for the Cdc6 protein in forming the pre-replicative complexes of budding yeast. *Nature* 379(6561), 180–182 (1996).
23. Takabatake N, Toriyama S, Igarashi A *et al.* A novel polymorphism in CDC6 is associated with the decline in lung function of ex-smokers in COPD. *Biochem. Biophys. Res. Commun.* 381(4), 554–559 (2009).
24. Couto PP, Bastos-Rodrigues L, Schayek H *et al.* Spectrum of germline mutations in smokers and non-smokers in Brazilian non-small-cell lung cancer (NSCLC) patients. *Carcinogenesis* 38(11), 1112–1118 (2017).
25. Ma Q, Xu Y, Liao H *et al.* Identification and validation of key genes associated with non-small-cell lung cancer. *J. Cell. Physiol.* 234(12), 22742–22752 (2019).
26. Van Den Abbeele A, De Corte V, Van Impe K *et al.* Downregulation of gelsolin family proteins counteracts cancer cell invasion *in vitro*. *Cancer Lett.* 255(1), 57–70 (2007).
27. Shao F, Zhang R, Don L, Ying K. Overexpression of gelsolin-like actin-capping protein is associated with progression of lung adenocarcinoma. *Tohoku J. Exp. Med.* 225(2), 95–101 (2011).
28. Altieri DC. Survivin, versatile modulation of cell division and apoptosis in cancer. *Oncogene* 22(53), 8581–8589 (2003).
29. Hamy AS, Bieche I, Lehmann-Che J *et al.* BIRC5 (survivin): a pejorative prognostic marker in stage II/III breast cancer with no response to neoadjuvant chemotherapy. *Breast Cancer Res. Treat.* 159(3), 499–511 (2016).
30. Zhou L, Lu J, Liang ZY *et al.* High nuclear Survivin expression as a poor prognostic marker in pancreatic ductal adenocarcinoma. *J. Surg. Oncol.* 118(7), 1115–1121 (2018).
31. Tian QG, Wu YT, Liu Y *et al.* Expressions and correlation analysis of HIF-1alpha, survivin and VEGF in patients with hepatocarcinoma. *Eur. Rev. Med. Pharmacol. Sci.* 22(11), 3378–3385 (2018).
32. Shang X, Liu G, Zhang Y *et al.* Downregulation of BIRC5 inhibits the migration and invasion of esophageal cancer cells by interacting with the PI3K/Akt signaling pathway. *Oncol. Lett.* 16(3), 3373–3379 (2018).
33. Cao Y, Zhu W, Chen W, Wu J, Hou G, Li Y. Prognostic value of BIRC5 in lung adenocarcinoma lacking *EGFR*, *KRAS*, and *ALK* mutations by integrated bioinformatics analysis. *Dis. Markers* 2019, 5451290 (2019).
34. Cahill DP, Lengauer C, Yu J *et al.* Mutations of mitotic checkpoint genes in human cancers. *Nature* 392(6673), 300–303 (1998).
35. Pinto M, Vieira J, Ribeiro FR *et al.* Overexpression of the mitotic checkpoint genes BUB1 and BUBR1 is associated with genomic complexity in clear cell kidney carcinomas. *Cell. Oncol.* 30(5), 389–395 (2008).
36. Dong S, Huang F, Zhang H, Chen Q. Overexpression of BUB1B, CCNA2, CDC20, and CDK1 in tumor tissues predicts poor survival in pancreatic ductal adenocarcinoma. *Biosci. Rep.* 39(2), pii: BSR20182306 (2019).
37. Sundaram G, Palchadhuri S, Dixit S, Chattopadhyay D. MAPK mediated cell cycle regulation is associated with Cdc25 turnover in *S. pombe* after exposure to genotoxic stress. *Cell Cycle* 7(3), 365–372 (2008).
38. Schaal C, Chellappan SP. Nicotine-mediated cell proliferation and tumor progression in smoking-related cancers. *Mol. Cancer Res.* 12(1), 14–23 (2014).