

# Identification of Molecular Markers Associated with Diagnosis and Prognosis in Lung Adenocarcinoma by Bioinformatics Analysis

Zimeng Wei

Wuhan University School of Basic Medical Sciences <https://orcid.org/0000-0002-6534-3982>

Linnan Zang

Wuhan University School of Basic Medical Sciences

Min Zhao (✉ [zhaomin28@163.com](mailto:zhaomin28@163.com))

Wuhan University <https://orcid.org/0000-0001-9207-4629>

---

## Research

**Keywords:** Lung Adenocarcinoma, bioinformatics, differentially expressed genes, prognosis

**Posted Date:** September 23rd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-79373/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background** Lung adenocarcinoma (LUAD) is the main histological subtype of lung cancer. However, the molecular mechanism underlying LUAD is not yet clearly defined, but elucidating this process in detail would be of great significance for clinical diagnosis and treatment. Our aim is to identify the candidate key genes and pathways associated with diagnosis and prognosis in LUAD.

**Methods** In this study, three gene expression profiles GSE118370, GSE32863 and GSE43458 were retrieved from Gene Expression Omnibus database (GEO), and the common differentially expressed genes (DEGs) were identified by online GEO2R analysis tool. Subsequently, the enrichment analysis of function and signaling pathways of DEGs in LUAD were performed by gene ontology (GO) and The Kyoto Encyclopedia of Genes and Genomics (KEGG) analysis. The protein-protein interaction (PPI) networks of the DEGs were established through the Search Tool for the Retrieval of Interacting Genes (STRING) database and hub genes were screened by plug-in CytoHubba in Cytoscape. Afterwards, the miRNAs and the hub genes network was constructed via miRWalk. Finally, receiver operating characteristic (ROC) curve and Kaplan-Meier plotter were performed to analyze the diagnosis and prognosis efficacy of hub genes.

**Results:** A total of 311 DEGs were identified, including 74 up-regulated and 238 down-regulated genes. GO analysis results showed that DEGs were mainly enriched in biological processes including composition of extracellular matrix, regulation of angiogenesis and so on. KEGG analysis results revealed DEGs were mainly enrolled in cell adhesion signaling pathway. Subsequently, 10 hub genes, CDC20, CENPF, TPX2, TOP2A, KIAA0101, CDCA7, ASPM, ECT2, UBE2T and COL1A1, were identified. And TOP2A, CDCA7, TPX2 and COL1A1 showed strong relationships with each other and the miRNAs nearby in miRNAs-mRNA network obtained by miRWalk website. Finally, all these 10 hub genes were found significantly related to the diagnosis and prognosis of LUAD ( $p < 0.05$ ).

**Conclusions:** The identification of hub genes in this study will help us to understand the pathogenesis of LUAD, especially the molecular mechanisms of its development. Our results suggested that TOP2A, CDCA7, TPX2 and COL1A1 might present predictive value for the development and prognosis in LUAD, and might be used as potential molecular markers for the diagnosis and treatment of LUAD.

## 1. Introduction

Lung cancer is the leading cause of cancer morbidity and mortality in the world, with 2.1 million new cases and 1.8 million deaths predicted in 2018[1]. Although the great advance in surgical and molecular target therapies, the prognosis of lung cancer is still dismal[2]. According to histopathology characteristics, lung cancer can be divided into two main types: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC), with NSCLC being accounted for 85% of lung cancer with lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC) and large cell carcinoma (LCC) three main subtypes[3]. LUAD has been ranked the first cancer mortality in china[4],and receiving an operation

remains a primary method, unfortunately the 5-year survival rate is around 10–15%[5]. Lacking sensitive and effective early diagnosis biomarkers and tending to metastasis and rapid dissemination are deemed to contribute to the poor prognosis. Therefore, looking for reliable biomarkers or drug targets is of great importance for improving the diagnosis and prognosis of LUAD.

High-throughput screening technique which is based on RNA sequencing or microarray gene expression profiles has been widely used for gene detection and diseases analysis. Academic can rapidly and precisely get the knowledge of key genes associated with serious diseases like cancer via a large number of biological information data mining[6]. For instance, Huang et al. used integrated bioinformatics analysis of RNA sequencing data and superenhancer catalogs to identify superenhancer-associated circRNAs and further found that loss of superenhancer-regulated circNfix promoted cardiac regenerative repair and functional recovery[7]. As for lung cancer, compound CA-5f was considered as a novel late-stage autophagy inhibitor with potent anti-tumor effect against NSLC via bioinformatics technique[8]. The discovery and the lucubration of these key genes obtained through bioinformatics analysis contribute to the discovery of effective biomarkers, which can be applied to the diagnosis, treatment and prognosis of tumors.

The aim of our study was to perform multiple analyses towards key genes of LUAD through integrated bioinformatics methods. First of all, we downloaded three microarray datasets containing mRNA expression data from Gene Expression Omnibus (GEO) and the differentially expressed genes (DEGs) were screened out via GEO2R online tool. Next, by virtue of Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), functional annotation and enrichment analysis about these DEGs were performed, and then we established protein-protein interaction (PPI) network and obtained the hub genes in LUAD. Then, the miRNAs and the hub genes network was constructed to select and identify the candidate biomarkers. Last but not least, ROC diagnosis and Kaplan-Meier survival analysis were carried out to provide support for the diagnosis, targeted therapy and prognosis of LUAD.

## 2. Materials And Methods

### 2.1. Acquisition and processing of GEO data

3 LUAD datasets were obtained from GEO database (available online: <http://www.ncbi.nlm.nih.gov/geo>) [9], including the gene expression profiles GSE118370 (6 LUAD samples, 6 normal samples), GSE32863 (58 LUAD samples, 58 normal samples) and GSE43458 (80 LUAD samples, 30 normal samples). The GSE118370 dataset (Submission date: Jan 23, 2019) was based on the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array, Affymetrix, Inc). The GSE32863 dataset (Submission date: Mar 21, 2012) was based on the GPL6884 platform (Illumina HumanWG-6 v3.0 expression beadchip). The dataset GSE43458 (Submission date: Aug 06, 2013) was based on the GPL6244 platform ([HuGene-1\_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]). DEGs were screened and identified via GEO2R online tools (available online: <https://www.ncbi.nlm.nih.gov/geo/geo2r/>) with the cut-off  $|\log_2FC| \geq 1$  and adj.  $p < 0.05$ . The numbers of DEGs we got from GSE118370, GSE32863 and GSE43458

were 3302, 1584 and 1025, respectively. DEGs co-occurred in all 3 datasets were integrated together and showed in a Venn diagram (available online: <http://bioinformatics.psb.ugent.be/webtools/Venn/>).

## **2.2. GO and KEGG Pathway enrichment analyses and visualization**

Functional annotations and pathway enrichment analyses on DEGs were carried out through Enricher website (available online: <https://amp.pharm.mssm.edu/Enrichr/>)[10], which included the visual results of GO and KEGG enrichment analyses. GO can be used to study gene-related biological process (BP), molecular function (MF) and cellular component (CC), while KEGG can simulate and analyze the pathways that DEGs were enriched in.

## **2.3. PPI network and acquisition of hub genes**

PPI network can explain the possible mechanisms of the genes involved in intracellular processing. So the STRING (available online: <https://stringdb.org/>, Version11.0)[11], an online tool was first applied to establish a PPI network of DEGs, which was then imported into Cytoscape software (version 3.6.1)[12]. In addition, the plug-in CytoHubba in Cytoscape was used to search and analyze the hub genes that were generated through Maximal Clique Centrality (MCC) score.

## **2.4. The construction of miRNA interaction network**

MiRWalk website (available online: <http://mirwalk.umm.uni-heidelberg.de/>)[13] was used to establish a miRNA-mRNA regulatory network of the hub genes we obtained from CytoHubba plug-in.

## **2.5. Diagnosis and prognosis analyses of hub genes**

Receiver operating characteristic (ROC) can evaluate the diagnostic efficacy of each index and the larger the area under curve (AUC), the higher the diagnostic efficacy[14]. We used SPSS (version 24) to gather statistics and analyze the case samples in GSE43458 dataset. In addition, the Kaplan-Meier plotter (available online: <http://kmplot.com/>)[15], a online database for survival analysis, was used to analyze the prognosis value of the hub genes. The hazard ratio (HR) with 95% confidence intervals and logrank P value were calculated and displayed on the diagram.

# **3. Results**

## **3.1. Identification and screening of DEGs**

Three microarray datasets: GSE118370, GSE32863 and GSE43458 were obtained from GEO database, including 144 samples from LUAD patients and 94 samples of normal lung tissues. The differentially expressed genes (DEGs) were identified by online GEO2R analysis tool with the cut-off criterion of  $|\log_2FC| \geq 1$  and  $\text{adj. } p < 0.05$ , and got 3302, 1584 and 1025 DEGs from each dataset, respectively. Afterwards, a Venn diagram (Fig. 1) was used to get an intersection of DEGs from the 3 datasets above,

acquiring a total of 311 DEGs with a consistent expression, in which there were 74 up-regulated genes and 144 down-regulated genes (Table 1) in the LUAD samples compared to the normal samples.

## 3.2. GO analysis and KEGG pathway analysis of DEGs

To evaluate the function of DEGs, we performed functional annotations and pathway enrichment analyses via GO and KEGG. According to the result of KEGG, DEGs emerged a high enrichment in the pathways of cell adhesion, protein digestion and absorption, tyrosine metabolism, ECM receptor interaction and so on (Fig. 2A). GO analysis showed that DEGs mainly participated in the biological processes (BP) of the extracellular matrix organization, regulation of angiogenesis, negative regulation of blood vessel morphogenesis and so on (Fig. 2B). The molecular function (MF) included amyloid-beta binding, the activation of protein homodimerization, the activation of metalloendopeptidase inhibitors, the binding of low-density lipoprotein and so on (Fig. 2C). The analysis of cellular component (CC) emerged that these genes mainly encoded the components on the plasma membrane (Fig. 2D).

## 3.3. Analysis of PPI network and identification of hub genes

Exploring the protein-protein interaction in organisms can help to study molecular mechanisms of some diseases from a perspective of system. We built a PPI network via STRING website (Fig. 3A), then the interaction data were imported into Cytoscape, and then CytoHubba plug-in was applied to confirm 10 hub genes according to the scores of MCC. The results were shown in Fig. 3B. They were CDC20, CENPF, TPX2, TOP2A, KIAA0101, CDCA7, ASPM, ECT2, UBE2T and COL1A1, and their full names and functions were shown in Table 2.

## 3.4. miRNA regulatory network of hub genes

Through inputting 10 hub genes into miRWalk website, we obtained the miRNAs regulatory network related with the hub genes (Fig. 4). The blue spots in the center were the hub genes, and what surrounded them were the miRNA which might regulate the hub genes. As were shown in Fig. 4, these four genes: TPX2, TOP2A, CDCA7 and COL1A1 had stronger interactions and formed regulation networks with the miRNAs nearby, in which the central miRNAs has-miR-6129, has-miR-6759-5p, has-miR-6510-5p and so on had two more connections with the four genes above. miR-145[16], hsa-let-7d-5p[17] and so on also proved to have associations with the occurrence and the development of tumors.

## 3.5. Diagnosis and prognosis analyses of hub genes

Based on the GSE43458 dataset, SPSS software was used to analyze the diagnostic efficacy of these 10 hub genes. The results showed that the AUC of all the hub genes were above 0.800 (Fig. 5), CDC20 (AUC = 0.940), CENPF (AUC = 0.935), TPX2 (AUC = 0.935), TOP2A (AUC = 0.954), KIAA0101 (AUC = 0.925), CDCA7 (AUC = 0.939), ASPM (AUC = 0.909), ECT2 (AUC = 0.893), UBE2T (AUC = 0.982) and COL1A1 (AUC = 0.947). Among them, the AUC of TOP2A and UBE2T were above 0.950, indicating that as diagnosis

indexes, TOP2A and UBE2T would have better efficacy. Kaplan-Meier prognosis analysis has been performed, finding that all the hub genes especially UBE2T (HR = 2.23[1.87–2.66], logrank  $p < 1E-16$ ) and TPX2 (HR = 1.86[1.64–2.12], logrank  $p < 1E-16$ ), had significant relations with the prognosis of LUAD patients, as what were shown in Fig. 6.

## 4. Discussion

In this study, 311 DEGs were identified including 74 up-regulated and 238 down-regulated genes in five datasets from the GEO database. Furthermore, through Enricher online web tool, we visualized the outcomes derived from Gene Ontology and KEGG pathway enrichment analysis. As for the biological processes, these DEGs were enriched in extracellular matrix organization, regulation of angiogenesis, negative regulation of blood vessel morphogenesis, and negative regulation of angiogenesis and so on. As for the molecular functions, the DEGs were mainly enriched in amyloid-beta binding, protein homodimerization activity, metalloendopeptidase inhibitor activity, low-density lipoprotein particle binding and calcium ion binding. For the cellular components, the DEGs showed enrichment in the integral component of plasma membrane. Angiogenesis is closely related to the occurrence and progression of cancers<sup>[18]</sup> and the formation of extracellular matrix is also associated with tumor metastasis and invasion [19, 20]. According to KEGG pathway enrichment analysis, tyrosine metabolism was found to be significant in LUAD. Li et al. revealed that activation of tyrosine metabolism in CD13<sup>+</sup> cancer stem cells may drive relapse in hepatocellular carcinoma by means of generating nuclear acetyl-CoA to acetylate and stabilize Foxd3, and allowing CD13<sup>+</sup> cancer stem cells to sustain quiescence and resistance to chemotherapeutic agents[21]. Nitration of protein tyrosine has proved to be involved in a variety of biological processes, including signal transduction, protein degradation, energy metabolism, mitochondrial dysfunction, enzyme inactivation, immunogenic response, cell apoptosis and cell death, and plays an important role in the occurrence and metastasis of lung cancer[22]. Therefore, the signaling pathway of tyrosine metabolism was expected to be a potential drug therapy target for LUAD.

Next, DEGs PPI network was constructed via the STRING online database and Cytoscape software. By virtue of “CytoHubba” plug-in, the top ten hub genes, CDC20, CENPF, TPX2, TOP2A, KIAA0101, CDCA7, ASPM, ECT2, UBE2T, COL1A1 were identified and they were all found up-regulated in LUAD. In addition, miRWalk, an online analysis tool, was used to construct a network of miRNAs associated with the regulation of these genes, and finally four genes (TOP2A, CDCA7, TPX2 and COL1A1) were found to show strong associations with each other, which could be considered as new effective targets to improve the prognosis of LUAD patients.

Topoisomerase II (TOP2) has been clarified to have crucial functions, including DNA replication, transcription and chromosome segregation, and more and more active anticancer drugs targeted it[23]. TOP2 contains two types of isozymes: TOP2A and topoisomerase II beta (TOP2B)[24], TOP2A is the only enzyme able to cleave and re-ligate the double-strand backbone of DNA, which is indispensable for DNA replication, transcription, and repair[25, 26]. Ejlersen et al.[25] showed that TOP2A was a direct molecular target of anthracyclines that can improve the sensitivity of anthracycline-containing chemotherapy in

high-risk breast cancer patients. In malignant peripheral-nerve sheath tumor, TOP2A was the most overexpressed gene compared with benign neurofibromas[27]. High expression of TOP2A was found to be correlated to worse overall survival (OS) in all non-small-cell lung cancer and lung adenocarcinoma patients, but not in lung squamous cell carcinoma patients[28, 29]. It has also been reported that miRNA's being associated with TOP2A plays an important role in lung cancer, for example, down-regulation of miRNA-144-3p whose potential target was TOP2A, was highly enriched in various key pathways like the protein digestion and absorption and the thyroid hormone signaling pathways in non-small cell lung cancer from the comprehensive meta-analysis[30].

Cell division cycle associated 7, CDCA7, was identified as a c-MYC responsive gene, and behaved as a direct c-MYC target gene. Overexpression of CDCA7 was found to enhance the transformation of lymphoblastoid cells, and it complements a transformation-defective MYC Box II mutant, suggesting its involvement in c-MYC-mediated cell transformation[31]. In quite a number of tumors, such as hepatocellular carcinoma[32], colorectal cancer[33], lymphoma[34], breast cancer[35], CDCA7 was all reported up-regulated and might be a potential prognostic factor and therapeutic target. Wang et al. have found that CDCA7 could promote lung adenocarcinoma proliferation via regulating the cell cycle and silencing CDCA7 inhibited cell proliferation in LUAD through G1 phase arrest and induction of apoptosis, which implied that CDCA7 might be identified as a potential therapeutic target for new biomarkers and LUAD[36].

TPX2, which is also known as DIL2 or p100, uses two flexibly linked elements ('ridge' and 'wedge') in a novel interaction mode to simultaneously bind across longitudinal and lateral tubulin interfaces[37, 38]. In ovarian cancer, it can promote the proliferation and migration of human ovarian cancer cells by regulating PLK1 expression[39]. Except for these function, in various cancers can TPX2 also control bladder cancer cell's proliferation and invasion via TPX2-p53-GLIPR1 regulatory circuitry[40], regulate the PI3K/AKT signaling pathway to facilitate hepatocellular carcinoma[41], interactive with miRNA such as miR-485-3p[42], miR-361-5p[43], miR-335-5p[44], miR-216b[45] and so on. Zhou et al. have verified that TPX2 can activate the epithelial-mesenchymal transition process and promote both the expression and activities of matrix metalloproteinase (MMP)2 and MMP9 in non-small cell lung cancer (NSCLC), which means TPX2 promotes the metastasis and malignant progression of NSCLC and could thus serve as a marker of poor prognosis in NSCLC[46].

Some studies have confirmed that COL1A1-related miRNAs were involved in the regulation of different kinds of tumors, including LUAD[17, 47]. Zhang et al. found that COL1A1 was positively correlated with NOTCH3 expression, and the miR-150 /NOTCH3/COL1A1 axis might be involved in EGFR-TKI resistance in LUAD, which provided a potential target for LUAD therapy and could be used as a prognostic target[48]. Hsa-let-7d-5p has been proved that it can regulate cell cycle including both G1/S and G2/M cell cycle phase transitions and telomere maintenance in human lung fibroblasts and is related to cellular senescence [17]. Hsa-miR-483-3p can also play an important role in lung cancer caused by radon exposure by regulating certain signaling pathways such as proliferative protein kinases (MAPK) and reactive oxygen species (ROS) [49].

Followed by, ROC curve analysis of these 10 hub genes was conducted based on dataset GSE43458, and we found that they all had certain diagnostic efficacy, which TOP2A ranked first. The results of Kaplan-Meier survival analysis also showed that these hub genes were significantly correlated with the prognosis of LUAD patients, suggesting that they may be potential prognostic biomarkers of LUAD.

## 5. Conclusion

In summary, our study indicated that TOP2A, CDCA7, TPX2 and COL1A1 had strong relationships with each other and also owned better diagnosis and prognosis efficacy in LUAD. The mechanism and their mutual regulation network are worthy of further research and experiments. Anyway, all of our analyses may provide some useful direction into the potential biomarkers and molecular mechanism of the occurrence and development of LUAD.

## Abbreviations

LUAD, lung adenocarcinoma; GEO, Gene Expression Omnibus; DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, The Kyoto Encyclopedia of Genes and Genomics; PPI, The protein-protein interaction; STRING, Search Tool for the Retrieval of Interacting Genes; NSLC, non-small cell lung cancer; SCLC, small cell lung cancer; LUSC, lung squamous cell carcinoma; LCC, large cell carcinoma; BP, biological process; MF, molecular function; CC, cellular component; ROC, receiver operating characteristic; AUC, area under curve; HR, hazard ratio.

## Declarations

### Acknowledgements

Not applicable.

### Funding

Not applicable.

### Availability of data and materials

All data generated or analyzed during this study are included in this published article.

### Authors' contributions

All authors contributed to the study concept and design, as well as the interpretation of the data. MZ designed and managed the whole research; ZMW analyzed the data and wrote the main manuscript text; LNZ revised the manuscript. All authors read and approved the final version of the manuscript.

### Ethics approval and consent to participate

Not applicable.

### **Patient consent for publication**

Not applicable.

### **Competing interests**

The authors declare that they have no competing interests.

## **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.* 2018;68(6):394–424.
2. McLellan R, Marshall H, Dent A, Bowman RV, Yang IA, Fong KM. Diagnosis and treatment of early lung cancer. *Aust J Gen Pract.* 2020;49(8):508–12.
3. Tan WL, Jain A, Takano A, Newell EW, Iyer NG, Lim WT, Tan EH, Zhai W, Hillmer AM, Tam WL, et al. Novel therapeutic targets on the horizon for lung cancer. *Lancet Oncol.* 2016;17(8):e347–62.
4. Remon J, Ahn MJ, Girard N, Johnson M, Kim DW, Lopes G, Pillai RN, Solomon B, Villacampa G, Zhou Q. Advanced-Stage Non-Small Cell Lung Cancer: Advances in Thoracic Oncology 2018. *J Thorac Oncol.* 2019;14(7):1134–55.
5. Hoy H, Lynch T, Beck M. Surgical Treatment of Lung Cancer. *Crit Care Nurs Clin North Am.* 2019;31(3):303–13.
6. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545–50.
7. Huang S, Li X, Zheng H, Si X, Li B, Wei G, Li C, Chen Y, Chen Y, Liao W, et al. Loss of Super-Enhancer-Regulated circRNA Nfix Induces Cardiac Regeneration After Myocardial Infarction in Adult Mice. *Circulation.* 2019;139(25):2857–76.
8. Zhang L, Qiang P, Yu J, Miao Y, Chen Z, Qu J, Zhao Q, Chen Z, Liu Y, Yao X, et al. Identification of compound CA-5f as a novel late-stage autophagy inhibitor with potent anti-tumor effect against non-small cell lung cancer. *Autophagy.* 2019;15(3):391–406.
9. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 2013;41(Database issue):D991–5.
10. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016;44(W1):W90–7.

11. Kelly-Smith M, Strain GM. STRING data mining of GWAS data in canine hereditary pigment-associated deafness. *Vet Anim Sci.* 2020;9:100118.
12. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504.
13. Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: An online resource for prediction of microRNA binding sites. *PLoS One.* 2018;13(10):e0206239.
14. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol.* 2018;63(7):07TR01.
15. Szasz AM, Lanczky A, Nagy A, Forster S, Hark K, Green JE, Boussioutas A, Busuttill R, Szabo A, Gyorffy B. Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. *Oncotarget.* 2016;7(31):49322–33.
16. Dimitrova N, Gocheva V, Bhutkar A, Resnick R, Jong RM, Miller KM, Bendor J, Jacks T. Stromal Expression of miR-143/145 Promotes Neoangiogenesis in Lung Cancer Development. *Cancer Discov.* 2016;6(2):188–201.
17. Markopoulos GS, Roupakia E, Tokamani M, Vartholomatos G, Tzavaras T, Hatzia Apostolou M, Fackelmayer FO, Sandaltzopoulos R, Polytaichou C, Kolettas E. Senescence-associated microRNAs target cell cycle regulatory genes in normal human lung fibroblasts. *Exp Gerontol.* 2017;96:110–22.
18. Zhang K, Wang L, Wei A, Jia X, Liu X. CM082, a novel angiogenesis inhibitor, enhances the antitumor activity of gefitinib on epidermal growth factor receptor mutant non-small cell lung cancer in vitro and in vivo. *Thorac Cancer.* 2020;11(6):1566–77.
19. Gomez-Contreras P, Ramiro-Diaz JM, Sierra A, Stipp C, Domann FE, Weigel RJ, Lal G. Extracellular matrix 1 (ECM1) regulates the actin cytoskeletal architecture of aggressive breast cancer cells in part via S100A4 and Rho-family GTPases. *Clin Exp Metastasis.* 2017;34(1):37–49.
20. Gotte M, Kovalszky I. Extracellular matrix functions in lung cancer. *Matrix Biol.* 2018;73:105–21.
21. Sun L, Zhang L, Chen J, Li C, Sun H, Wang J, Xiao H. Activation of Tyrosine Metabolism in CD13 + Cancer Stem Cells Drives Relapse in Hepatocellular Carcinoma. *Cancer Res Treat.* 2020;52(2):604–21.
22. Masri FA, Comhair SA, Koeck T, Xu W, Janocha A, Ghosh S, Dweik RA, Golish J, Kinter M, Stuehr DJ, et al. Abnormalities in nitric oxide and its derivatives in lung cancer. *Am J Respir Crit Care Med.* 2005;172(5):597–605.
23. Nitiss JL. Targeting DNA topoisomerase II in cancer chemotherapy. *Nat Rev Cancer.* 2009;9(5):338–50.
24. Li TK, Liu LF. Tumor cell death induced by topoisomerase-targeting drugs. *Annu Rev Pharmacol Toxicol.* 2001;41:53–77.
25. Ejlersen B, Jensen MB, Nielsen KV, Balslev E, Rasmussen BB, Willemoe GL, Hertel PB, Knoop AS, Mouridsen HT, Brunner N. HER2, TOP2A, and TIMP-1 and responsiveness to adjuvant anthracycline-containing chemotherapy in high-risk breast cancer patients. *J Clin Oncol.* 2010;28(6):984–90.

26. Bartlett JM, McConkey CC, Munro AF, Desmedt C, Dunn JA, Larsimont DP, O'Malley FP, Cameron DA, Earl HM, Poole CJ, et al. Predicting Anthracycline Benefit: TOP2A and CEP17-Not Only but Also. *J Clin Oncol*. 2015;33(15):1680–7.
27. Skotheim RI, Kallioniemi A, Bjerkhagen B, Mertens F, Brekke HR, Monni O, Mousses S, Mandahl N, Soeter G, Nesland JM, et al. Topoisomerase-II alpha is upregulated in malignant peripheral nerve sheath tumors and associated with clinical outcome. *J Clin Oncol*. 2003;21(24):4586–91.
28. Hou GX, Liu P, Yang J, Wen S. Mining expression and prognosis of topoisomerase isoforms in non-small-cell lung cancer by using Oncomine and Kaplan-Meier plotter. *PLoS One*. 2017;12(3):e0174515.
29. Ni M, Liu X, Wu J, Zhang D, Tian J, Wang T, Liu S, Meng Z, Wang K, Duan X, et al. Identification of Candidate Biomarkers Correlated With the Pathogenesis and Prognosis of Non-small Cell Lung Cancer via Integrated Bioinformatics Analysis. *Front Genet*. 2018;9:469.
30. Chen YJ, Guo YN, Shi K, Huang HM, Huang SP, Xu WQ, Li ZY, Wei KL, Gan TQ, Chen G. Down-regulation of microRNA-144-3p and its clinical value in non-small cell lung cancer: a comprehensive analysis based on microarray, miRNA-sequencing, and quantitative real-time PCR data. *Respir Res*. 2019;20(1):48.
31. Gill RM, Gabor TV, Couzens AL, Scheid MP. The MYC-associated protein CDCA7 is phosphorylated by AKT to regulate MYC-dependent apoptosis and transformation. *Mol Cell Biol*. 2013;33(3):498–513.
32. Gao J, Dai C, Yu X, Yin XB, Zhou F. LncRNA LEF1-AS1 silencing diminishes EZH2 expression to delay hepatocellular carcinoma development by impairing CEBPB-interaction with CDCA7. *Cell Cycle*. 2020;19(8):870–83.
33. Li S, Huang J, Qin M, Zhang J, Liao C. High expression of CDCA7 predicts tumor progression and poor prognosis in human colorectal cancer. *Mol Med Rep*. 2020;22(1):57–66.
34. Martin-Cortazar C, Chiodo Y, Jimenez RP, Bernabe M, Cayuela ML, Iglesias T, Campanero MR. CDCA7 finely tunes cytoskeleton dynamics to promote lymphoma migration and invasion. *Haematologica*. 2020;105(3):730–40.
35. Ye L, Li F, Song Y, Yu D, Xiong Z, Li Y, Shi T, Yuan Z, Lin C, Wu X, et al. Overexpression of CDCA7 predicts poor prognosis and induces EZH2-mediated progression of triple-negative breast cancer. *Int J Cancer*. 2018;143(10):2602–13.
36. Wang H, Ye L, Xing Z, Li H, Lv T, Liu H, Zhang F, Song Y. CDCA7 promotes lung adenocarcinoma proliferation via regulating the cell cycle. *Pathol Res Pract*. 2019;215(11):152559.
37. Zhang R, Roostalu J, Surrey T, Nogales E. **Structural insight into TPX2-stimulated microtubule assembly**. *Elife* 2017, 6.
38. Alfaro-Aco R, Petry S. How TPX2 helps microtubules branch out. *Cell Cycle*. 2017;16(17):1560–1.
39. Ma S, Rong X, Gao F, Yang Y, Wei L. TPX2 promotes cell proliferation and migration via PLK1 in OC. *Cancer Biomark*. 2018;22(3):443–51.
40. Yan L, Li Q, Yang J, Qiao B. TPX2-p53-GLIPR1 regulatory circuitry in cell proliferation, invasion, and tumor growth of bladder cancer. *J Cell Biochem*. 2018;119(2):1791–803.

41. Huang DH, Jian J, Li S, Zhang Y, Liu LZ. TPX2 silencing exerts antitumor effects on hepatocellular carcinoma by regulating the PI3K/AKT signaling pathway. *Int J Mol Med*. 2019;44(6):2113–22.
42. Taherdangkoo K, Kazemi Nezhad SR, Hajjari MR, Tahmasebi Birgani M. miR-485-3p suppresses colorectal cancer via targeting TPX2. *Bratisl Lek Listy*. 2020;121(4):302–7.
43. Long N, Chu L, Jia J, Peng S, Gao Y, Yang H, Yang Y, Zhao Y, Liu J. CircPOSTN/miR-361-5p/TPX2 axis regulates cell growth, apoptosis and aerobic glycolysis in glioma cells. *Cancer Cell Int*. 2020;20:374.
44. Gu X, Yao X, Liu D. Up-regulation of microRNA-335-5p reduces inflammation via negative regulation of the TPX2-mediated AKT/GSK3beta signaling pathway in a chronic rhinosinusitis mouse model. *Cell Signal*. 2020;70:109596.
45. Feng C, Zhang HL, Zeng A, Bai M, Wang XJ. Tumor-Suppressive MicroRNA-216b Binds to TPX2, Activating the p53 Signaling in Human Cutaneous Squamous Cell Carcinoma. *Mol Ther Nucleic Acids*. 2020;20:186–95.
46. Zhou F, Wang M, Aibaidula M, Zhang Z, Aihemaiti A, Aili R, Chen H, Dong S, Wei W, Maimaitiaili A. TPX2 Promotes Metastasis and Serves as a Marker of Poor Prognosis in Non-Small Cell Lung Cancer. *Med Sci Monit*. 2020;26:e925147.
47. Fang SS, Guo JC, Zhang JH, Liu JN, Hong S, Yu B, Gao Y, Hu SP, Liu HZ, Sun L, et al. A P53-related microRNA model for predicting the prognosis of hepatocellular carcinoma patients. *J Cell Physiol*. 2020;235(4):3569–78.
48. Zhang Y, Chen B, Wang Y, Zhao Q, Wu W, Zhang P, Miao L, Sun S. NOTCH3 Overexpression and Posttranscriptional Regulation by miR-150 Were Associated With EGFR-TKI Resistance in Lung Adenocarcinoma. *Oncol Res*. 2019;27(7):751–61.
49. Cui FM, Li JX, Chen Q, Du HB, Zhang SY, Nie JH, Cao JP, Zhou PK, Hei TK, Tong J. Radon-induced alterations in micro-RNA expression profiles in transformed BEAS2B cells. *J Toxicol Environ Health A*. 2013;76(2):107–19.

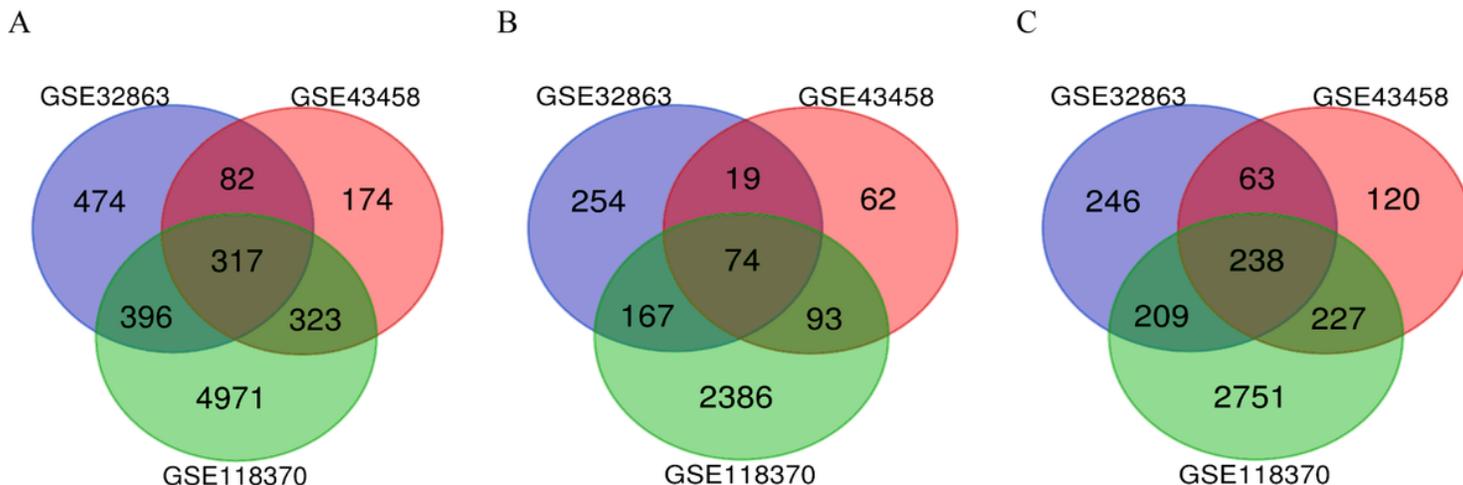
## Tables

**Table 2 Differentially expressed genes**

Up- or Down-regulated	Number	Differentially expressed genes
Up-regulated	74	GDF15 TPX2 C15orf48 IGF2BP3 COL1A1 ADGRF1 METTL7B ANKRD22 SLC50A1 TMPRSS4 SPP1 CLDN10 PLPP2 TIMP1 MMP12 HABP2 UBE2T SLC2A1 EPCAM ECT2 MMP11 FAM83A CDH3 CRABP2 CEACAM5 FUT3 LAD1 TNFRSF21 TFAP2A SULF1 GCNT3 COL11A1 FUT2 KIAA0101 AGR2 SPINK1 S100P HMGB3 DSP GOLM1 PROM2 MUC16 CST1 SERINC2 ETV4 LGSN PYCR1 CDCA7 SFN CDC20 COL3A1 ASPM OCIAD2 CD24 TCN1 SLC7A5 IL37 NQO1 PSAT1 SPDEF TOP2A GJB2 THBS2 SMPDL3B PCP4 AIM2 COL10A1 LCN2 CTHRC1 FBXO32 ATP10B CENPF MMP9
Down-regulated	238	RTKN2 FRAS1 MSRB3 AQP9 NOSTRIN PAPSS2 A2M HIGD1B ANXA3 TEK GRK5 VGLL3 C10orf67 GHR MS4A2 ZNF385B SLC02A1 LAMP3 BCHE STARD13 ID1 PTRF SPTBN1 TSPAN7 LRRK2 AFAP1L1 ACADL AHNAK FHL1 KLF4 PLSR4 NDRG2 TCF21 GLDN NEDD4L LDLR SFTPD OLR1 GLIPR2 AGER HOXA5 CD93 MSR1 WIF1 RAMP3 CLIC5 NPNT MME LAMA3 ICAM2 PDK4 GBP4 EFEMP1 EMCN ADARB1 LMO7 STARD8 LRRC32 SASH1 ESAM SCGB1A1 CA4 TACC1 FMO3 LYVE1 GIMAP4 P3H2 PHACTR2 FEZ1 PDE5A CPA3 TBX3 PDZD2 LHFP ADIRF HBB HEY1 NTNG1 GNG11 RHOJ SRPX TSPAN18 ZNF106 SLC39A8 SFTPC MYL9 TGFB3 LPL CAT PTPRM NDRG4 SOSTDC1 SBSPON GPC3 ADCY4 FPR2 PGM5 LMCD1 FOSB MARCO GSTM5 DCN EPAS1 SLIT2 CX3CR1 FZD4 PCOLCE2 CYBRD1 C2orf40 AOX1 HSD17B6 GKN2 CYP4B1 SMAD6 AFF3 HSPB8 SLC6A4 TMEM204 GIMAP8 CDH13 ACVRL1 TIMP3 DACH1 EML1 RERG HBEGF CRYAB CD36 ETS2 PTPRB ADGRL2 PPARG AQP4 MT1M TNNC1 IL1RL1 MCEMP1 CPB2 KLF6 EMP1 ABCA8 SLC19A3 AOC3 TMEM47 ZBTB16 EDNRB SLIT3 FAM167A MYH10 STX11 MYH11 DUSP1 RGCC VWF ABI3BP ARRB1 VSIG4 CYR1 MS4A15 HPGD C7 S100A8 HEG1 IL6 CPED1 PRICKLE2 ANGPT1 SEMA6A TMEM100 DUOX1 FERMT2 CD34 KIAA1324L DPT MFAP4 PECAM1 PRG4 MAOB B3GALNT1 CDO1 SEPP1 SDPR EPB41L2 ADAMTS1 CLEC14A ANOS1 C8B STXBP6 S1PR1 GIMAP7 LIMCH1 SPOCK2 RNF144B LDB2 CALCRL RECK CAV1 CLEC1A JAM2 SPARCL1 CA2 ANKRD29 ADH1A METTL7A ITM2A FGD5 FAM189A2 LRRN3 LGI3 IL7R FMO2 SOCS2 ACTG2 TMOD1 FGFBP2 ARHGEF6 IL33 PLLP HHIP PGC ADH1B NEDD9 FABP4 FCN3 PDLIM3 FAM107A GPM6B SEMA5A RAMP2 VIPR1 PLAC8 ANKRD1 MAMDC2 CLDN18 CDH5 OLFML1 COL6A6 FBLN5

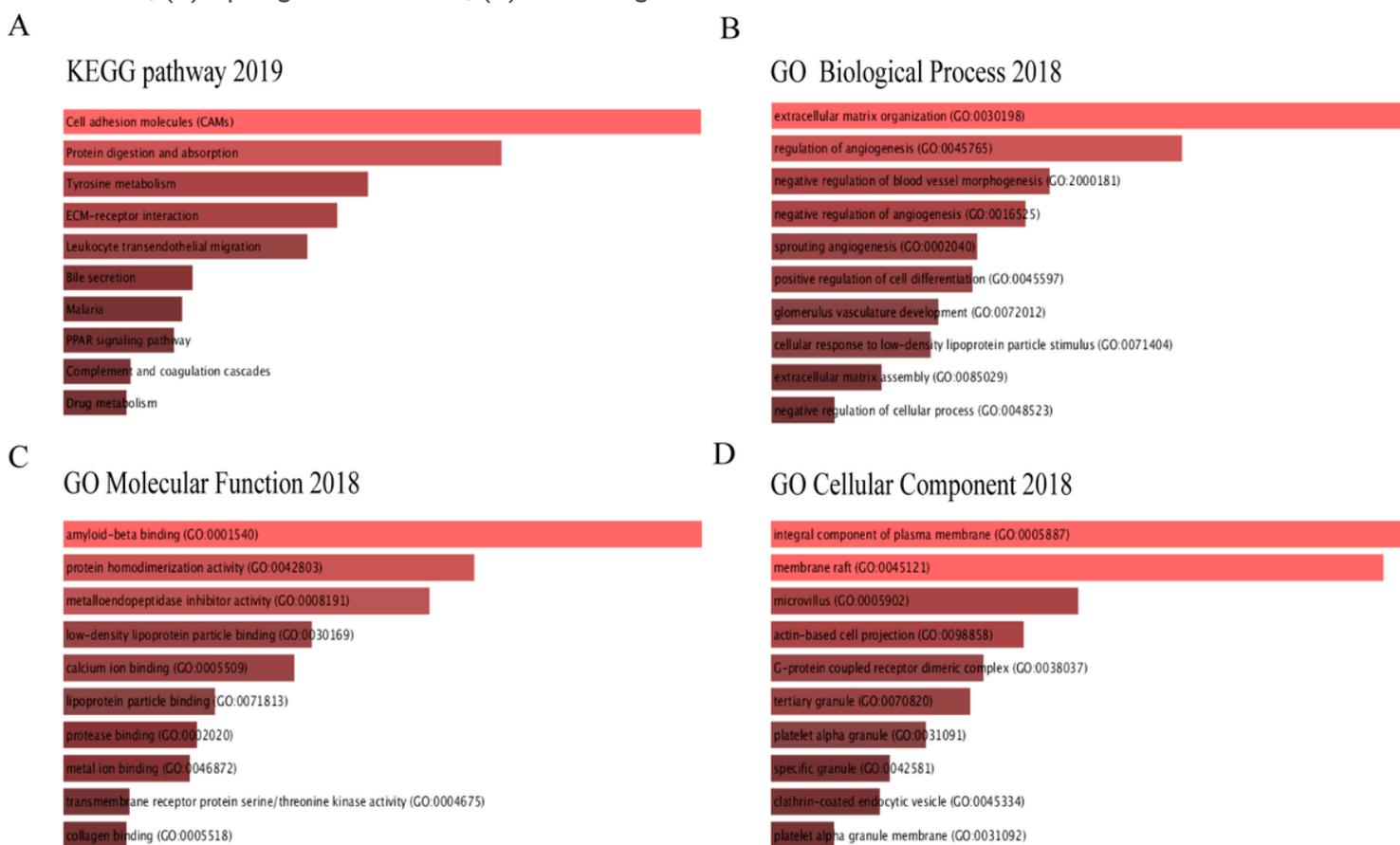
<b>Table 1 The Full Name and Functions of Hub Genes</b>			
No.	Name	Full name	Function
1	CDC20	Cell division cycle 20	CDC20 is a regulatory protein interacting with several other proteins at multiple points in the cell cycle.
2	CENPF	Centromere protein F	CENPF is a component of the nuclear matrix during the G2 phase of interphase
3	TPX2	TPX2 microtubule nucleation factor	TPX2 can bind to and increases the activity of a protein kinase, an enzyme which phosphorylates a protein.
4	TOP2A	DNA topoisomerase II alpha	TOP2A is involved in processes such as chromosome condensation, chromatid separation, and the relief of torsional stress.
5	KIAA0101	PCLAF(PCNA clamp associated factor)	KIAA0101 can prevent doxorubicin-induced apoptosis by inhibiting p53 activation
6	CDCA7	Cell division cycle associated 7	Overexpression of this gene is found to enhance the transformation of lymphoblastoid cells
7	ASPM	Assembly factor for spindle microtubules	ASPM is essential for normal mitotic spindle function in embryonic neuroblasts.
8	ECT2	Epithelial cell transforming 2	The expression of this gene is elevated with the onset of DNA synthesis and remains elevated during G2 and M phases.
9	UBE2T	Ubiquitin-conjugating enzyme E2T	The protein encoded by UBE2T catalyzes the covalent attachment of ubiquitin to protein substrates.
10	COL1A1	Collagen type I alpha 1 chain	COL1A1 encodes the pro-alpha1 chains of type I collagen whose triple helix comprises two alpha1 chains and one alpha2 chain.

## Figures



**Figure 1**

The expression of DEGs in GSE32863, GSE43458 and GSE118370 via Venn diagrams (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). Different colors represented different datasets. (A) total DEGs; (B) up-regulation DEGs; (C) down-regulation DRGs.



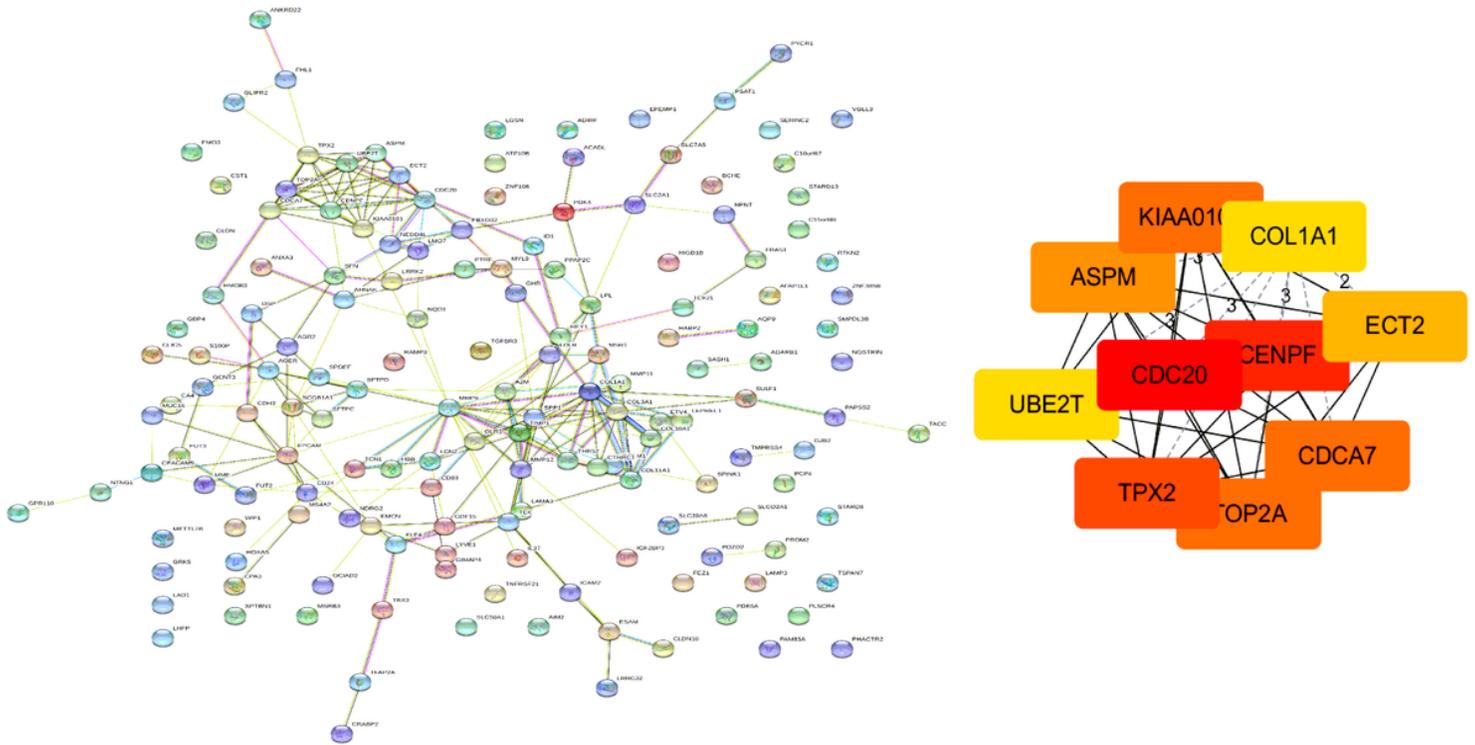
**Figure 2**

The KEGG and GO enrichment analyses about 311 DEGs using Enricher online tool (<https://amp.pharm.mssm.edu/Enrichr>). (A) Kyoto Encyclopedia of Genes and Genomes (KEGG)

pathways 2019. (B) Enrichment of GO Biological Process (2018) terms in proteomic analysis. (C) Enrichment of GO Molecular Function (2018) terms in proteomic analysis. (D) Enrichment of GO Cellular Component (2018) terms in proteomic analysis. The ratio of the percent composition of terms in proteomic data versus percent composition in the genome annotation was performed in bar graph. The length of the bar means the significance of that specific gene-set or term, and the significance rises as the bar goes deeper.

A

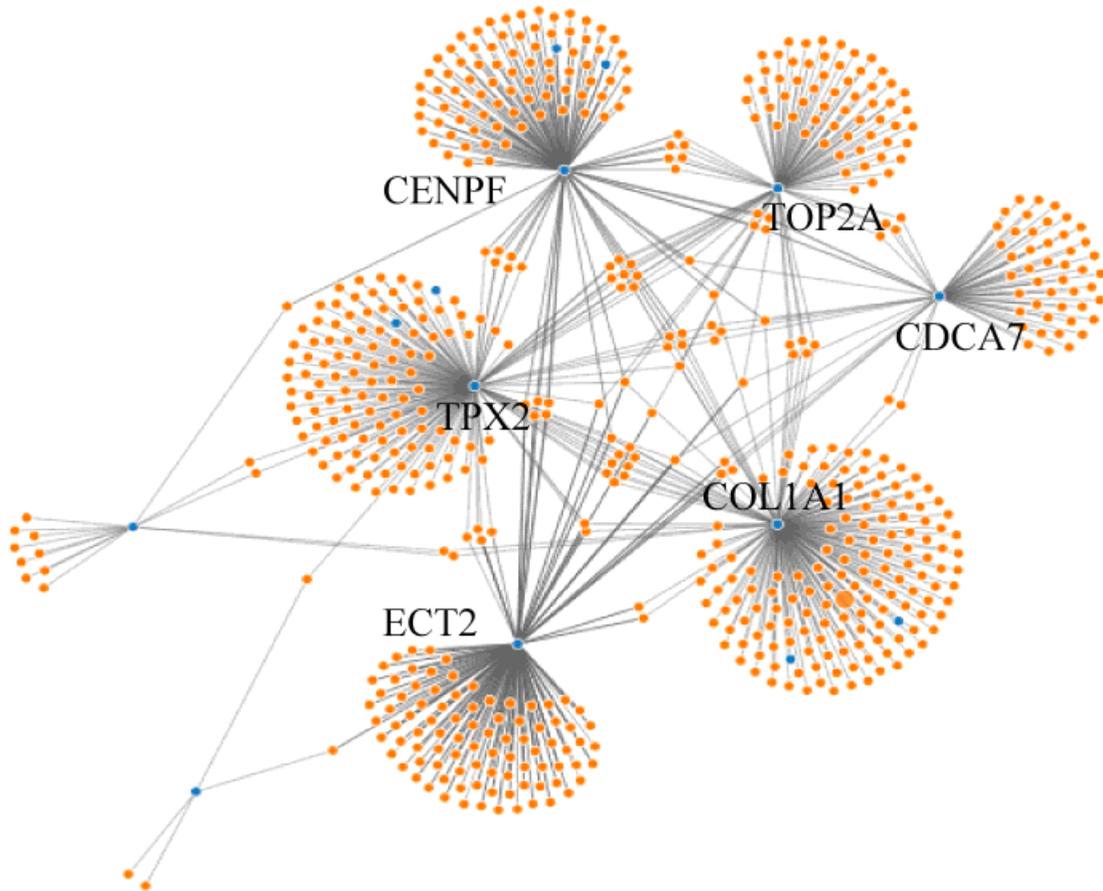
B



**Figure 3**

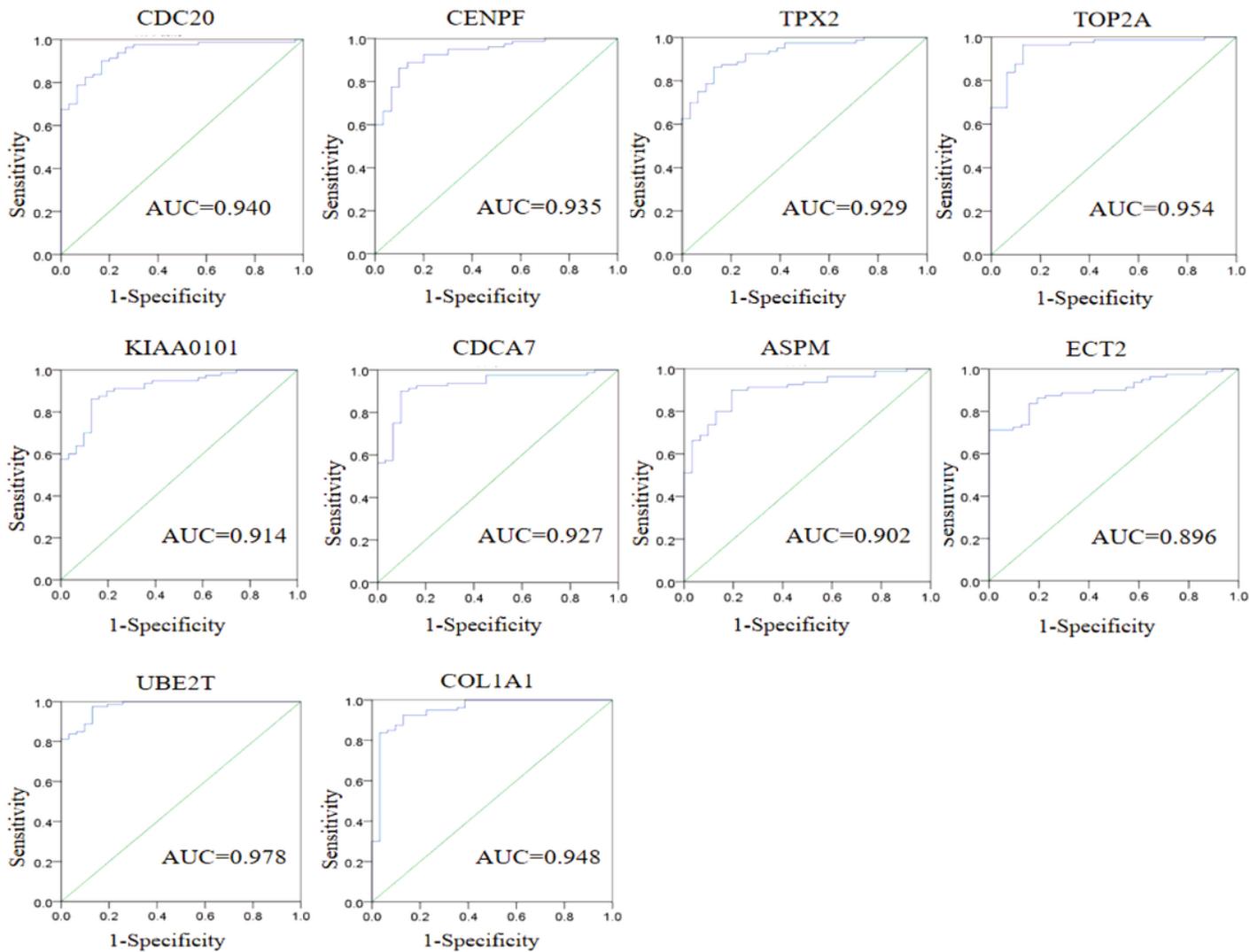
The PPI network constructed by STRING online database and the identification of 10 hub genes. (<https://stringdb.org/>). (A) The PPI network of 311 DEGs, the nodes meant different proteins and the lines meant interaction between the two proteins. (B) 10 hub genes were screened out by cytohubba plug-in. The squares meant genes and the lines meant interaction between two genes, and the deeper the color was, the more significant the gene was.

A



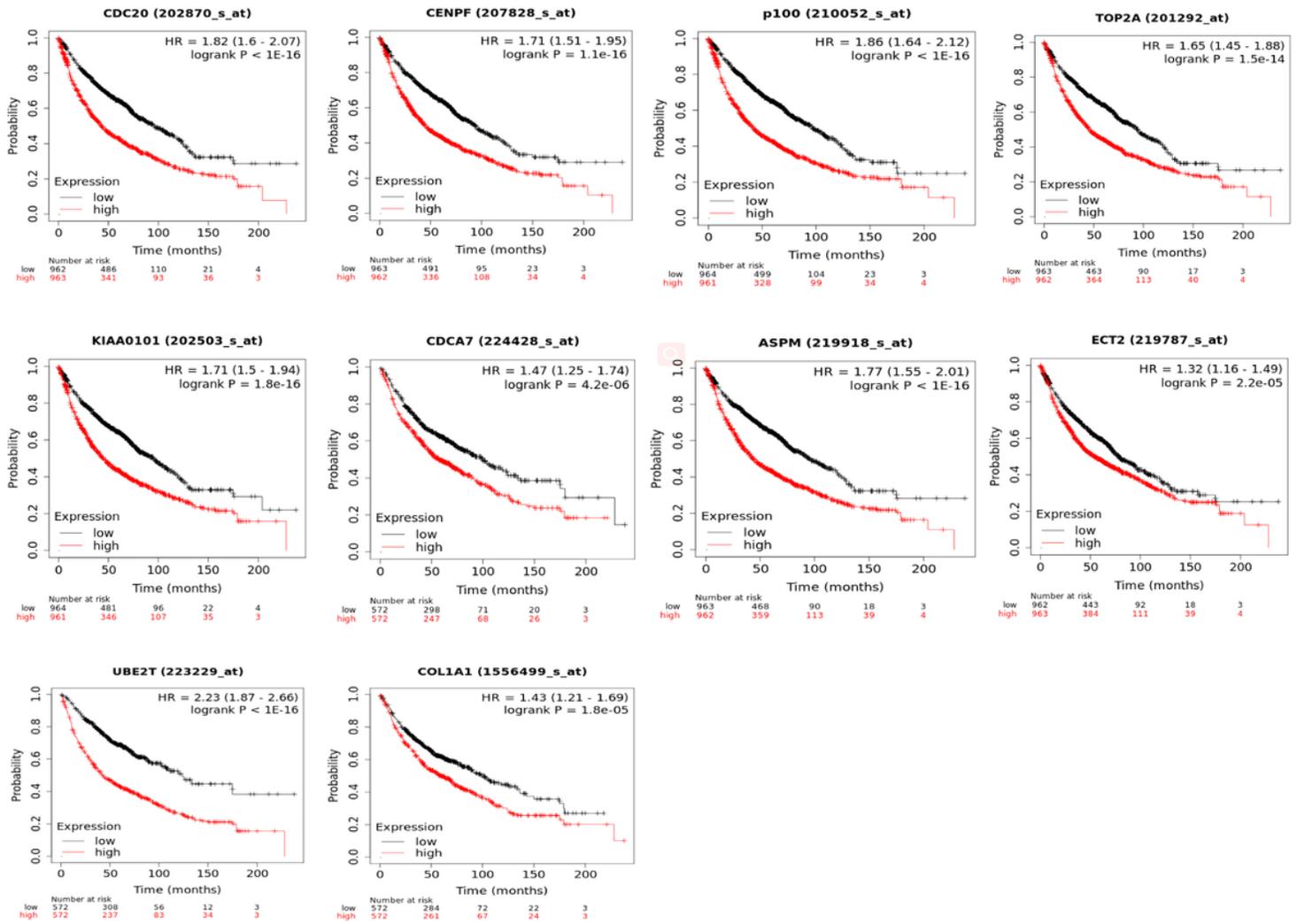
**Figure 4**

The interaction network between hub genes and its related miRNAs via the website miRWalk (<http://mirwalk.umm.uni-heidelberg.de/>). The blue dot meant gene while the orange one meant its related potential miRNAs. The line revealed the relationship between them.



**Figure 5**

The ROC curves for 10 hub genes. AUC, area under the receiver operating characteristic curve,  $P < 0.05$  was considered statistically significant.



**Figure 6**

The Kaplan-Meier plotter survival analyses of 10 hub genes. (<http://kmplot.com/>). TPX2 was also known as p100. OS, overall survival; HR, hazard ratio, Logrank P < 0.05 was considered statistically significant.