Screening and identification of significant genes in esophageal squamous cell carcinoma by bioinformatical analysis

Weirui Ren
Hebei Medical University Third Affiliated Hospital
Chuang Zhang
Second Hospital of Hebei Medical University
Lei Pan
Hebei Medical University
Weijing Wang
Wuan General Hospital
Wenjuan Zhao
Hebei Medical University Third Affiliated Hospital
Yuxiang Zhang
Hebei Medical University Third Affiliated Hospital
Junmin Wang (WJMRAY@126.com)
Hebei Medical University Third Affiliated Hospital https://orcid.org/0000-0002-9007-9263

Research

Keywords: Esophageal squamous cell carcinoma, Bioinformatical analysis, Differentially expressed genes

Posted Date: March 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-16607/v1

Abstract

Background: Esophageal squamous cell carcinoma (ESCC) is one of the most common cancers with notably high incidence and mortality rates. However, the molecular mechanism underlying ESCC pathogenesis and prognosis is very complicated. The main objective of our investigation has been to obtain some knowledge of significant genes with poor outcome and their underlying mechanisms.

Methods: Gene expression profiles of GSE26886, GSE23400, GSE20347 and GSE17351 were available from GEO database. The differentially expressed genes (DEGs) were identified, and function enrichment analyses were performed. The protein-protein interaction network (PPI) was constructed and the module analysis was performed using STRING and Cytoscape software.

Results: A total of 105 DEGs were identified between normal esophagus and ESCC bioinformatical analysis samples. Functional annotations of the common DEGs indicate that extracellular matrix (ECM) remodeling plays a key role in tumor formation and progression.18 hub genes were identified and disease free survival analysis showed that CDKN3, RAD51AP1, KIF4A may be involved in poor prognosis in ESCC patients.

Conclusions: DEGs and hub genes identified in the present study help us understand the molecular mechanisms underlying the carcinogenesis and progression of ESCC, and provide candidate targets for diagnosis and treatment of ESCC.

Background

Esophageal cancer (EC) is one of the common malignant tumors, with the sixth highest mortality and the eighth highest incidence rate worldwide[1]. There are regional differences in the incidence of esophageal cancer. In the high-risk area, stretching from Northern Iran through the central Asian republics to North-Central China, 90 percent of cases are squalors cell carcinomas (SCC)[2, 3]. The typical clinical symptom of esophageal squamous cell carcinoma(ESCC) is progressive dysphagia. Early esophageal cancers are not specifically symptomatic. Dysphagia usually occurs once the esophageal lumen diameter is less than 13mm, which indicates advanced disease and leads to a high mortality rate as well as a heavy burden on patients and their families[4]. Therefore, it is urgent to find out the core genes for clinical diagnosis and poor prognosis of esophageal squamous cell carcinoma. Gene chip used for more than ten years can quickly detect differentially expressed genes and had proven to be a reliable technology[5]. It can generate many slice data and can be stored in a shared database. Therefore, a lot of valuable clues can be explored from these data for further research. Integrated bioinformatical analysis can help us further research and better explore the underlying mechanisms[6]. In this study, to identify potential biomarkers for ESCC, we chose GSE26886, GSE23400, GSE20347, and GSE17351 from Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) database. The data were downloaded and analyzed to obtain the differentially expressed genes (DEGs) between esophageal cancer tissue and non-cancerous tissue. The Database for Annotation, Visualization and Integrated Discovery (DAVID) was utilized to analyze

these DEGs including molecular function (MF), cellular component (CC), biological process (BP) and Kyoto Encyclopedia of Gene and Genome (KEGG) pathways. Moreover, the protein-protein interaction (PPI) network of DEGs was used to identify some core genes.

Material And Methods

Data Collection

Gene expression profiles of GSE26886, GSE23400, GSE20347, and GSE17351 in esophageal squamous cell carcinoma and normal tissues were downloaded from the NCBI-GEO database. Microarray data included 9 ESCC tissue samples and 19 normal esophageal squamous epithelium (ESE) tissue samples, 53 ESCC tissue samples and 53 normal ESE tissue samples, 17 ESCC tissue samples and 17 normal ESE tissue samples along with 7 ESCC tissue samples and 7 normal ESE tissue samples, respectively.

Data Identification of DEGs

The DEGs between ESCC and normal ESE tissue samples were screened using GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r) with |log fold change (FC)|≥1 and adjust P value <0.05. GEO2R is an interactive web tool that allows users to compare two or more datasets in the GEO series in to identify DEGs across experimental conditions[7] . The DEGs with log FC<0 was taken to be down-regulated genes, while the DEGs with log FC>0 was considered as an up-regulated gene.

KEGG and GO enrichment analyses of DEGs

KEGG is a bioinformatics resource that deals with genomes, diseases, enzymatic pathways, drugs and chemicals[8]. GO is a major bioinformatics initiative for interrogating gene sets in multiple annotation categories, and analyzing their biological processes[9]. DAVID is an online analysis tool designed to identify enriched gene ontological functions[10]. We used DAVID to construct the enrichment plots in order to visualize the common DEGs in the probable biological processes (BP), cellular components (CC), and molecular functions (MF) and KEGG pathways. Adjusted P value <0.05 was regarded as statistically significant.

Protein Interaction Network (PPI), Modular Analysis and Hub Genes

This research used an online database search tool (STRING, http://string-db.org) for the retrieval of interacting genes to predict the PPI network. Protein performs functions by interacting with each other. The PPI network gives a chance to understand protein functions at a system level. In this study, a PPI network of the differentially expressed genes was constructed with the STRING database. We kept the network with a confidence score > 0.4 in STRING, and then entered it into Cytoscapesoftware for visualization[11]. Cytoscape's inserted Molecular Complex Detection (MCODE) is an application for clustering a given network based on topology to find densely connected regions[12]. We used Cytoscape to draw a PPI network and use MCODE to identify the most important modules in the PPI network. The selection criteria are as follows: MCODE score> 5, degree cutoff = 2, node score cutoff = 0.2, maximum

depth = 100, and k score = 2. KEGG pathway enrichment analysis is then performed on important modules of functional interpretation. The hub genes were selected with degrees ≥10. The biological processes of the hub genes were analyzed and visualized using the Biological Networks Oncology Tool (BiNGO) plugin of Cytoscape[13] .

Expression and Survival Analysis of Hub Genes

Hierarchical clustering of hub genes was constructed using UCSC Cancer Genomics Browser (http://genome-cancer.ucsc.edu)[14]. GEPIA (http://gepia.cancer-pku.cn/)[15] is a newly web server for gene expression profiling of 9,736 tumors and 8,587 normal samples from TCGA and GTEx projects. The cBioPortal (http://cBioPortal.org) [16] is a public online tool based on TCGA. To validate these hub genes, we applied the GEPIA and cBioPortal website to analyze the data of RNA sequencing expression on the basis of thousands of samples. GEPIA was used to analyze the expression levels of central genes in cancer tissues and healthy control tissues. Kaplan-Meier curve in cBioPortal was used to perform the overall survival and disease-free survival analyses of hub genes. The log rank P value and/or hazard ratio (HR) with 95% confidence intervals were/was computed and showed on the plot. In addition, we used the Biological Network Gene Oncology Tool (BiNGO) plugin in Cytoscape to analyze and visualize biological l processes of the hub genes.

Results

Identifcation of DEGs in esophageal squamous cell carcinoma

Via GEO2R online tools, we extracted 8267, 673,1733 and 744 DEGs from GSE26886, GSE23400, GSE20347, and GSE17351, respectively. The detailed information about GEO datasets is summarized in Table 1. Then, we used Venn diagram software to identify the commonly DEGs. The overlap areas shown in Venn diagram contained 105 genes among the 4 datasets, consisting of 42 downregulated genes and 63 upregulated genes between esophageal squamous cell carcinoma tissues and healthy control tissues. The venn diagram was shown in Fig. 1.

KEGG and GO enrichment analyses of DEGs

All 105 DEGs were analyzed by DAVID software and the results of GO analysis indicated that 1) for biological processes (BP), up-regulated DEGs were particularly enriched in extracellular matrix organization, collagen catabolic process and extracellular matrix disassembly, and down-regulated DEGs in liver development and xenobiotic metabolic process; 2) for GO cell component (CC), up-regulated DEGs were enriched in collagen trimer, proteinaceous extracellular matrix and extracellular region, and down-regulated DEGs in extrinsic component of membrane and organelle membrane; 3). for molecular function (MF), up-regulated DEGs were enriched in protein binding, ATP binding and neuregulin binding, and down-regulated DEGs in monooxygenase activity and beta-catenin binding. KEGG pathway enrichment analysis showed that the DEGs mainly participated in ECM-receptor interaction, Small cell lung cancer and Protein

digestion and absorption pathways, while down regulated DEGs in no significant signaling pathways ($P < 0.05$) (Fig. 2).

Protein Interaction Network (PPI) and Modular Analysis

A total of 105 DEGs were imported into the DEGs PPI network complex which included 102nodes and 234 edges, including 42 down-regulated and 63 up-regulated genes (Fig.3a). Three significant models were obtained using the MCODE application in Cytoscape. The scores were 17.294, 5.000,3.000, respectively (Fig.3c/d/e). The most important module was selected for further analysis. Path enrichment analysis indicates that the module is mainly related to DNA replication and cell cycle（Table 2 ）. A total of 18 genes were identified as hub genes with degrees ≥10. The names, abbreviations and functions for these hub genes are shown in Table3. The biological process analysis of the hub genes is shown in Fig.3f.

Expression and Survival Analysis of Hub Genes

Hierarchical clustering showed that the hub genes could basically differentiate the esophageal squamous cell carcinoma samples from the healthy control tissues(Fig.4). According to data from the GEPIA database, the expression levels of all hub genes in cancer patients were significantly higher than those in healthy controls, except for GMNN(Fig.5). In overall survival Analysis, those hub Genes were not statistically significant. Disease Free survival (RFS) analysis of the hub genes indicates that high expression of CDKN3, RAD51AP1, KIF4A was associated with poor prognosis in ESCC patients(Fig.6).

Discussion

Esophageal cancer is one of the most aggressive cancers in the world and the sixth leading cause of cancer-related death[17]. Histologically, the subtypes of esophageal squamous cell carcinoma in Asia has the highest incidence. Escc's pathogenesis is related to multiple factors including race, genetics and dietary intake[18]. The main treatments for ESCC include surgery, chemotherapy and radio-chemotherapy. However, the 5-year survival rate of patients with esophageal cancer remain low (<15%)[19]. It is of utmost importance to understand the pathogenic mechanisms of ESCC and develop effective strategies to treat ESCC. We need to improve our understanding of the pathogenesis of esophageal squamous cell carcinoma through molecular biology research. Microarray technology has allowed us to explore the genetic alterations of ESCC, and has been proved to be a useful method for identifying new biomarkers in other diseases. In this study, four microarray data sets were analyzed to obtain DEGs between esophageal squamous cell carcinoma tissue samples and healthy control tissue samples. A total of 105 DEGs were identified in these four data sets, including 63 up-regulated genes and 42 down-regulated genes. GO and KEGG enrichment analyses were performed to explore the interaction between the DEGs. The up-regulated genes were mainly enriched in extracellular matrix organization、the breakdown of collagen in the extracellular matrix、extracellular matrix disassembly 、hemidesmosome assembly、the extracellular matrix (ECM) -receptor interaction, protein digestion and absorption, and cell-matrix adhesions, while the down-regulated genes were mainly enriched in cellular response to xenobiotic stimulus hypodermal cell differentiation, negative regulation of cell migration cell differentiation, and

extrinsic component of membrane cell cortex. Increasing evidence shows that a comprehensive understanding of the molecular composition of esophageal cancer requires attention not only to tumor cells but also to the tumor microenvironment[20]. The cells are surrounded by the extracellular matrix (ECM), the complex network consisting of molecules, proteins, and polysaccharides. Related research shows that cancer-related fibroblasts secrete growth factors and change the ECM to create a tumor niche and enhancing tumor cell migration and metastasis[20]. Extracellular matrix (ECM) remodeling plays a key role in tumor formation and progression, especially invasion[21]. Stromal cells such as fibroblasts secrete ECM remodeling enzymes, such as lysyl oxidase (LOX) and matrix metalloproteinase (MMP), which can help with the formation of primary tumors or metastatic micro-ecologies and the down-regulation of cell adhesion to achieve invasion, migration and intravascular infiltration[22, 23]. ECM remodeling is associated with esophageal cancer, especially ESCC. For example, LOX-L2 is overexpressed in more than 90% of ESCC[24]. In addition, several matrix metalloproteinases (MIMPs) in ESCC, including MMP-2, MMP-7, and MMP-9, are all up-regulated and related to tumor stage[25]. These theories are consistent with our results. 18 hub genes were identified from the PPI network, and 3 of these genes, namely CDKN3, RAD51AP1, and KIF4A, were related to DFS in ESCC patients. The high expression levels of these genes were related to a shorter DFS and the poor prognosis of patients with esophageal cancer. Cyclin-dependent kinase inhibitor 3 (CDKN3) performs crucial roles in the modulation of tumor development. Many previous studies reported that CDNK3 plays an inhibitory role in tumor development [26]. New research finds that CDKN3 plays different roles in different types of cancer. CDKN3 is reported to be involved in the occurrence and development of some types of tumors [27, 28]. CDKN3 facilitated the promotion of ESCC cell proliferation, invasion and migration by activating the AKT signaling pathway[29]. CDKN3 knockdown significantly inhibited ESCC cell proliferation, migration, and invasion, and suppressed the G1 / S transition of tumor cells, providing a potential target for ESCC treatment[29]. RAD51-associated protein 1 (RAD51AP1) is an emerging protein that is essential for RAD51-mediated Homologous recombination (HR). HR serves to repair DNA double-strand breaks and damaged replication forks, which is essential for maintaining genome stability and suppressing tumors. However, there are no reports about the clinical value of RAD51AP1 in ESCC. Related studies have found that the expression of RAD51AP1 is up-regulated in primary hepatocellular carcinoma and intrahepatic bile duct cancer[30, 31]. High expression of RAD51AP1 is associated with shortened survival in patients with breast and ovarian cancer[32, 33]. Although the exact molecular role of RAD51AP1 in cancer development and progression has not been fully understood, research has uncovered a possible link between RAD51AP1 and tumor metastasis. Studies have reported that the high expression of RAD51AP1 is associated with poor prognosis in patients with breast and ovarian cancer[34]. In some tumor cells and tissues, the increased RAD51AP1 may be related to the advantages of selective growth, enabling the replication of early tumor cells and metastatic cells[35, 36]. However, we still don't know the precise molecular mechanisms by which RAD51AP1 carries out the HR reaction in cells and stabilizes DNA replication forks, and how posttranslational modifications affect the activities and biology of RAD51AP1. Therefore, we need to further study and explore the impact of RAD51AP1 on the pathogenesis and prognosis of ESCC. Filling the knowledge gap in the above fields will also provide prospects for targeted treatment of high RAD51AP1 cancer. Human kinesin family member 4A (KIF4A) is a 140 kD protein that plays a key role in

a variety of cellular processes, including chromosome condensation and segregation, cytok inesis during mitotic division, and middle-spindle formation [37, 38] Recently, it has become apparent that KIF4A plays an essential role in cancer development and progression. Recent studies have discovered that KIF4A might serve as a biomarker in breast cancer based on bioinformatics analysis[39]. KIF4A overexpression is observed in colorectal cancer and pancreatic ductal adenocarcinoma as well as in lung cancer, for which it is an independent prognostic risk factor[40-42]. However, the clinical value of KIF4A in esophageal squamous cell carcinoma is rarely mentioned. In the future, in-depth studies on the roles of KIF4A in ESCC might provide new clues for inhibiting ESCC.

Conclusions

In conclusion, with integrated bioinformatics analysis, we identified 105 DEGs and 18 hub genes in ESCC from four gene profile datasets. Functional annotations of the common DEGs in the four datasets indicate that extracellular matrix (ECM) remodeling plays a key role in tumor formation and progression. Among them, CDKN3, RAD51AP, and KIF4A may emerge as potential prognostic biomarkers or therapeutic targets for ESCC. However, these predictions should be verified by a series of experiments in the future. Anyway, these data may provide some useful information and direction into the potential biomarkers and biological mechanisms of ESCC.

Abbreviations

EC: Esophageal cancer; SCC: squalors cell carcinomas; ESCC: Esophageal squamous cell carcinoma; DEGs: Differentially expressed genes; PPI: Protein-protein interaction network; ECM: extracellular matrix; MF: molecular function ; CC: cellular component; BP: biological process; GEO: Gene Expression Omnibus; DAVID: Database for Annotation, Visualization and Integrated Discovery; GO: Gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; BiNGO: Biological Network Gene Oncology Tool; RFS: Disease Free survival; HR: Homologous recombination.

Declarations

Acknowledgements

Not applicable.

Funding

This work was supported by the Key Research and Development Projects of Hebei Province (19277729D)

Availability of data and materials

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

Authors' contributions

WR and CZ contributed to the conception and design of the study. WW, LP, WZ, and YZ contributed to the acquisition, analysis, and interpretation of data. WR, CZ, WW, LP, and JW drafted the article and revised it critically for important intellectual content. All the authors approved the final version of the manuscript to be submitted.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Chen X, Sina C, Baoxia L, Xiaona Z, Wenhui L, Henglun L, Xiaolong C: Identification of key genes and pathways for esophageal squamous cell carcinoma by bioinformatics analysis. Experimental & Therapeutic Medicine.
2. Gholipour C, Shalchi R, Abbasi M: A histopathological study of esophageal cancer on the western side of the Caspian littoral from 1994 to 2003. Diseases of the Esophagus Official Journal of the International Society for Diseases of the Esophagus 2008, 21:322-327.
3. Tran GD, Sun X-D, Abnet CC, Fan J-H, Dawsey SM, Dong Z-W, Mark SD, Qiao Y-L, Taylor PR: Prospective study of risk factors for esophageal and gastric cancers in the Linxian general population trial cohort in China. International Journal of Cancer, 113:456-463.
4. Humbert IA, Robbins J: Dysphagia in the Elderly. Journal of the American Geriatrics Society, 19:853-866.
5. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW: Cancer Genome Landscapes. Science, 339:1546-1558.
6. Hao, Feng, Zhong-Yi, Gu, Qin, Li, Qiong-Hua, Liu, Xiao-Yu, Yang: Identification of significant genes with poor prognosis in ovarian cancer via bioinformatical analysis.
7. Li L, Lei Q, Zhang S, Kong L, Qin B: Screening and identification of key biomarkers in hepatocellular carcinoma: Evidence from bioinformatic analysis. Oncology Reports.

8. Kanehisa M, Goto S: KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 2000, 28:27-30.
9. Ashburner M, Ball CA, Blake JA, Botstein D, Cherry JM: Gene ontology: Tool for the unification of biology. Nature Genetics 2000, 25:25-29.
10. Da WH, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 2008, 4:44-57.
11. 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 research 2003, 13:2498-2504.
12. Bader GD, Hogue CW: An automated method for finding molecular complexes in large protein interaction networks. 4:2-0.
13. Maere S: BiNGO : a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. Bioinformatics 2005, 21.
14. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler aD: The Human Genome Browser at UCSC. Genome Research, 12:996-1006.
15. Tang Z, Li C, Kang B, Ge G, Cheng L, Zhang Z: GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Research 2017:W1.
16. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E: Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Science Signaling, 6:pl1-pl1.
17. Chen J, Kwong DL, Cao T, Hu Q, Zhang L, Ming X, Chen J, Fu L, Guan X: Esophageal squamous cell carcinoma (ESCC): advance in genomics and molecular genetics. Diseases of the Esophagus, 28:84-89.
18. Abnet CC, Arnold M, Wei WQ: Epidemiology of Esophageal Squamous Cell Carcinoma. Gastroenterology 2017, 154.
19. Y Y, X S, C C, Y L, C L, D X, X S, X L: Upregulation of long noncoding RNA LINC00152 promotes proliferation and metastasis of esophageal squamous cell carcinoma. Cancer Manag Res:4643-4654.
20. Lin EW, Karakasheva TA, Hicks PD, Bass AJ, Rustgi AK: The tumor microenvironment in esophageal cancer. Oncogene.
21. Bonnans C, Chou J, Werb Z: Remodelling the extracellular matrix in development and disease. Nature Reviews Molecular Cell Biology, 15:786-801.
22. Cox TR, Erler JT: Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. Radiotherapy & Oncology, 4:165-178.
23. Kai K, Plaks V, Werb Z: Matrix Metalloproteinases: Regulators of the Tumor Microenvironment. 141:0-67.

24. Fong SFT, Dietzsch E, Fong KSK, Hollosi P, Asuncion L, He Q, Parker MI, Csiszar K: Lysyl oxidase-like 2 expression is increased in colon and esophageal tumors and associated with less differentiated colon tumors. Genes Chromosomes Cancer, 46:644-655.
25. Zhang Y, Wang Q, Ma A, Li Y, Li R, Wang Y: Functional expression of TLR9 in esophageal cancer. Oncology Reports, 31:2298–2304.
26. Dai W, Miao H, Fang S, Fang T, Chen N, Li M: CDKN3 expression is negatively associated with pathological tumor stage and CDKN3 inhibition promotes cell survival in hepatocellular carcinoma. Molecular Medicine Reports.
27. Wang H, Chen H, Zhou H, Yu W, Lu Z: Cyclin-Dependent Kinase Inhibitor 3 Promotes Cancer Cell Proliferation and Tumorigenesis in Nasopharyngeal Carcinoma by Targeting p27. Oncology Research, 25:1431-1440.
28. Zang X, Chen M, Zhou Y, Xiao G, Xie Y, Wang X: Identifying CDKN3 Gene Expression as a Prognostic Biomarker in Lung Adenocarcinoma via Meta-analysis. Cancer Informatics, 14s2:CIN.S17287.
29. Yu H, Yao J, Du M, Ye J, He X, Yin L: CDKN3 promotes cell proliferation, invasion and migration by activating the AKT signaling pathway in esophageal squamous cell carcinoma. Oncology Letters 2019, 19:542-548.
30. Song H, Xia SL, Liao C, Li YL, Wang Y-F, Li TP, Zhao MJ: Genes encoding Pir51, Beclin 1, RbAp48 and aldolase b are up or down-regulated in human primary hepatocellular carcinoma. World Journal of Gastroenterology, 10:p.509-513.
31. Obama K, Satoh S, Hamamoto R, Sakai Y, Nakamura Y, Furukawa Y: Enhanced Expression of RAD51 Associating Protein-1 Is Involved in the Growth of Intrahepatic Cholangiocarcinoma Cells. Clinical Cancer Research, 14:1333-1339.
32. Preethi S, E. ST, A. AK, Orly A, D. HJr: Tensor GSVD of Patient- and Platform-Matched Tumor and Normal DNA Copy-Number Profiles Uncovers Chromosome Arm-Wide Patterns of Tumor-Exclusive Platform-Consistent Alterations Encoding for Cell Transformation and Predicting Ovarian Cancer Survival. Plos One, 10:e0121396-.
33. Pathania R, Ramachandran S, Mariappan G, Thakur P, Shi H, Choi J-H, Manicassamy S, Kolhe R, Prasad PD, Sharma S: Combined inhibition of DNMT and HDAC blocks the tumorigenicity of cancer stem-like cells and attenuates mammary tumor growth. Cancer Research:0008-5472.CAN-0015-2249.
34. Pires E, Sung P, Wiese C: Role of RAD51AP1 in homologous recombination DNA repair and carcinogenesis. 2017, 59:76-81.
35. Halazonetis TD, Gorgoulis VG, Bartek J: An Oncogene-Induced DNA Damage Model for Cancer Development. Science, 319:1352-1355.
36. Gorgoulis VG, Vassiliou LF, Karakaidos P, Zacharatos P, Kotsinas A, Liloglou T, Venere M, Ditullio RA, Kastrinakis NG, Levy B: Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. Nature 2005, 434:907-913.
37. Mazumdar, M.: Human chromokinesin KIF4A functions in chromosome condensation and segregation. Journal of Cell Biology, 166:613-620.

38. Hu CK, Coughlin M, Field CM, Mitchison TJ: KIF4 Regulates Midzone Length during Cytokinesis. 21:815-824.
39. Xue D, Pu C, Mengjiao H, Xiyong L, Lijun X, Chenyi Y, Ke W, Jian H: An integrated bioinformatical analysis to evaluate the role of KIF4A as a prognostic biomarker for breast cancer. Oncotargets & Therapy, Volume 11:4755-4768.
40. Hou PF, Jiang T, Chen F, Shi P-C, Li H-Q, Bai J, Song J: KIF4A facilitates cell proliferation via induction of p21-mediated cell cycle progression and promotes metastasis in colorectal cancer. Cell Death & Disease, 9:477.
41. Haider S, Wang J, Nagano A, Desai A, Arumugam P, Dumartin L, Fitzgibbon J, Hagemann T, Marshall JF, Kocher HM: A multi-gene signature predicts outcome in patients with pancreatic ductal adenocarcinoma. Genome Medicine, 6:105.
42. Taniwaki M, Takano A, Ishikawa N, Yasui W, Inai K, Nishimura H, Tsuchiya E, Kohno N, Nakamura Y, Daigo Y: Activation of KIF4A as a Prognostic Biomarker and Therapeutic Target for Lung Cancer. Clinical Cancer Research An Official Journal of the American Association for Cancer Research, 13:6624-6631.

Tables

Table1 Information of GEO datasets

Datasets	Platform Samples	(tumor/normal)
GSE26886	Affymetrix Human Genome U133 Plus 2.0 Array	9/19
GSE23400	Affymetrix Human Genome U133A Array	53/53
GSE20347	Affymetrix Human Genome U133A 2.0 Array	17/17
GSE17351	Affymetrix Human Genome U133 Plus 2.0 Array	7/7

Table2 KEGG pathway analyses of DEGs in Module 1

Term	Count	PValue	Genes
hsa03030:DNA replication	2	0.020774017	RFC4, MCM6
hsa04110:Cell cycle	2	0.070192126	TTK, MCM6

Table3 Functional roles of 18 hub genes with degree ≥10

Gene symbol	Full name	Function
TOP2A	DNA topoisomerase II alpha	an enzyme that controls and alters the topologic states of DNA during transcription
RFC4	replication factor C subunit 4	The core complex possesses DNA-dependent ATPase activity
CENPF	centromere protein F	This gene encodes a protein that associates with the centromere-kinetochore complex.
CDKN3	cyclin dependent kinase inhibitor 3	a cyclin-dependent kinase inhibitor
UBE2C	ubiquitin conjugating enzyme E2 C	The encoded protein is required for the destruction of mitotic cyclins and for cell cycle progression, and may be involved in cancer progression.
FOXM1	forkhead box M1	The protein encoded by this gene is a transcriptional activator involved in cell proliferation
TPX2	TPX2 microtubule nucleation factor	Mediates AURKA localization to spindle microtubules
MCM6	minichromosome maintenance complex component 6	a role in the regulation of DNA replication
CKS1B	CDC28 protein kinase regulatory subunit 1B	CKS1B protein binds to the catalytic subunit of the cyclin dependent kinases
TTK	TTK protein kinase	Associated with cell proliferation
DTL	denticleless E3 ubiquitin protein ligase homolog	cell cycle control, DNA damage response and translesion DNA synthesis.
TRIP13	thyroid hormone receptor interactor 13	a protein that interacts with thyroid hormone receptors,
ATAD2	ATPase family AAA domain containing 2	assist in the assembly, operation, or disassembly of protein complexes
ECT2	epithelial cell transforming 2	an important role in the regulation of cytokinesis
RAD51AP1	RAD51 associated protein 1	May participate in a common DNA damage response pathway associated with the activation of homologous recombination and double-strand break repair.
KIF4A	kinesin family member 4A	involved in the intracellular transport of membranous organelles
KIF14	kinesin family member 14	This gene encodes a member of the kinesin-3 superfamily of microtubule motor proteins.
GMNN	geminin DNA replication inhibitor	This gene encodes a protein that plays a critical role in cell cycle regulation.

Figures

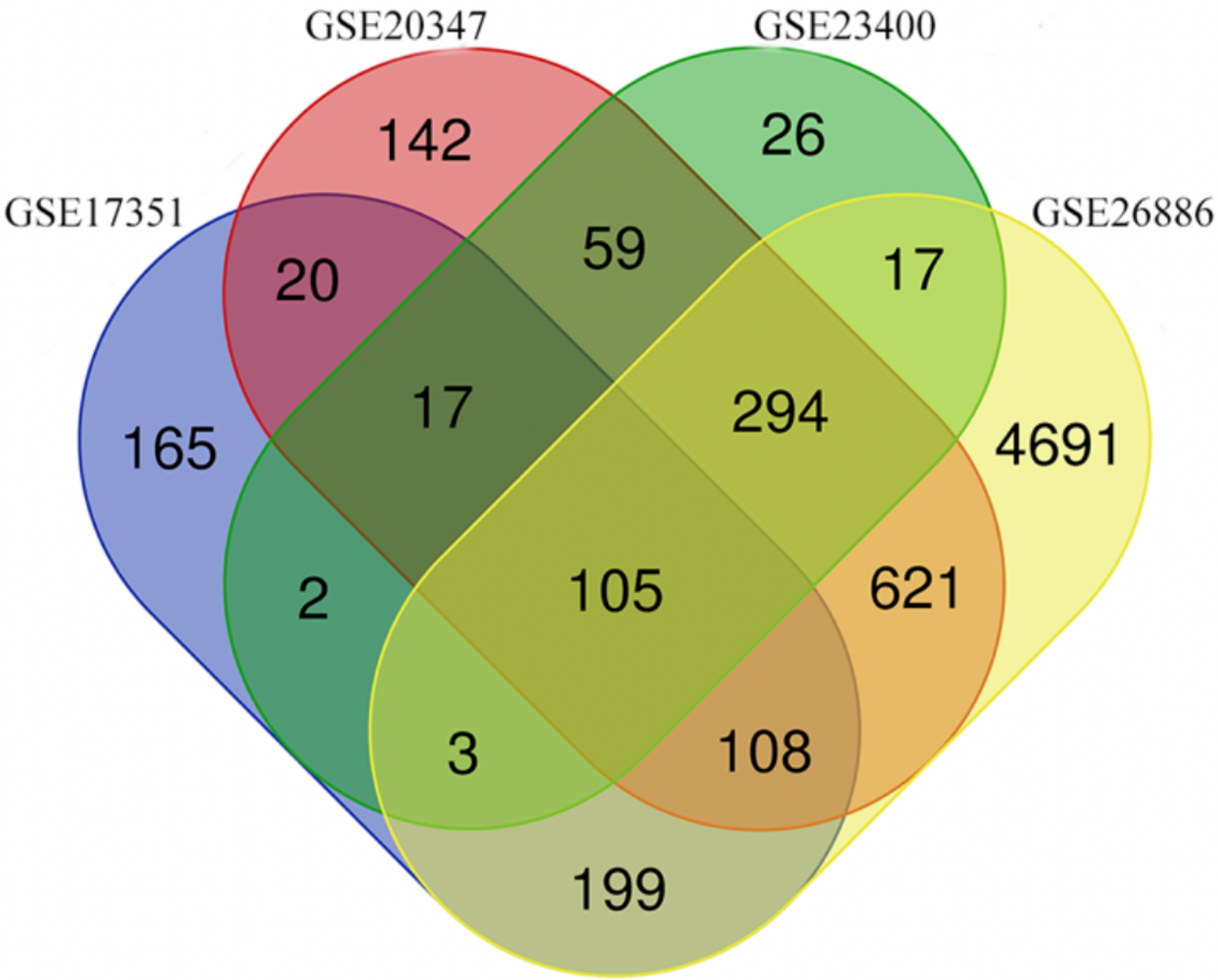

Figure 1

Venn diagram of 105 DEGs from the microarray datasets

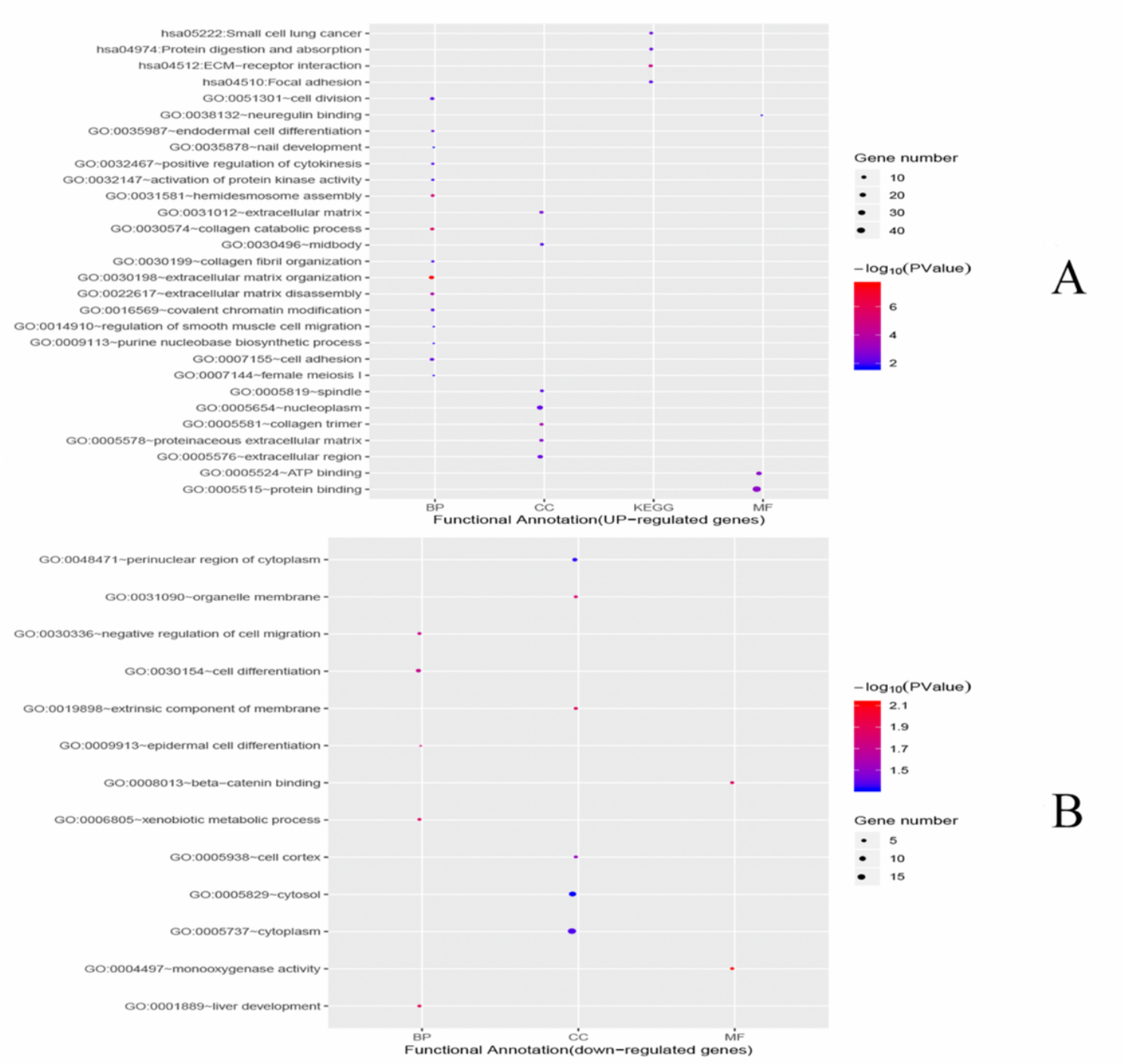

Figure 2

GO and KEGG pathway analyses of DEGs in ESCC. (a) BP, CC, MF and KEGG pathway of the upregulated genes. (b) BP, CC and ,MF of the downregulated genes.

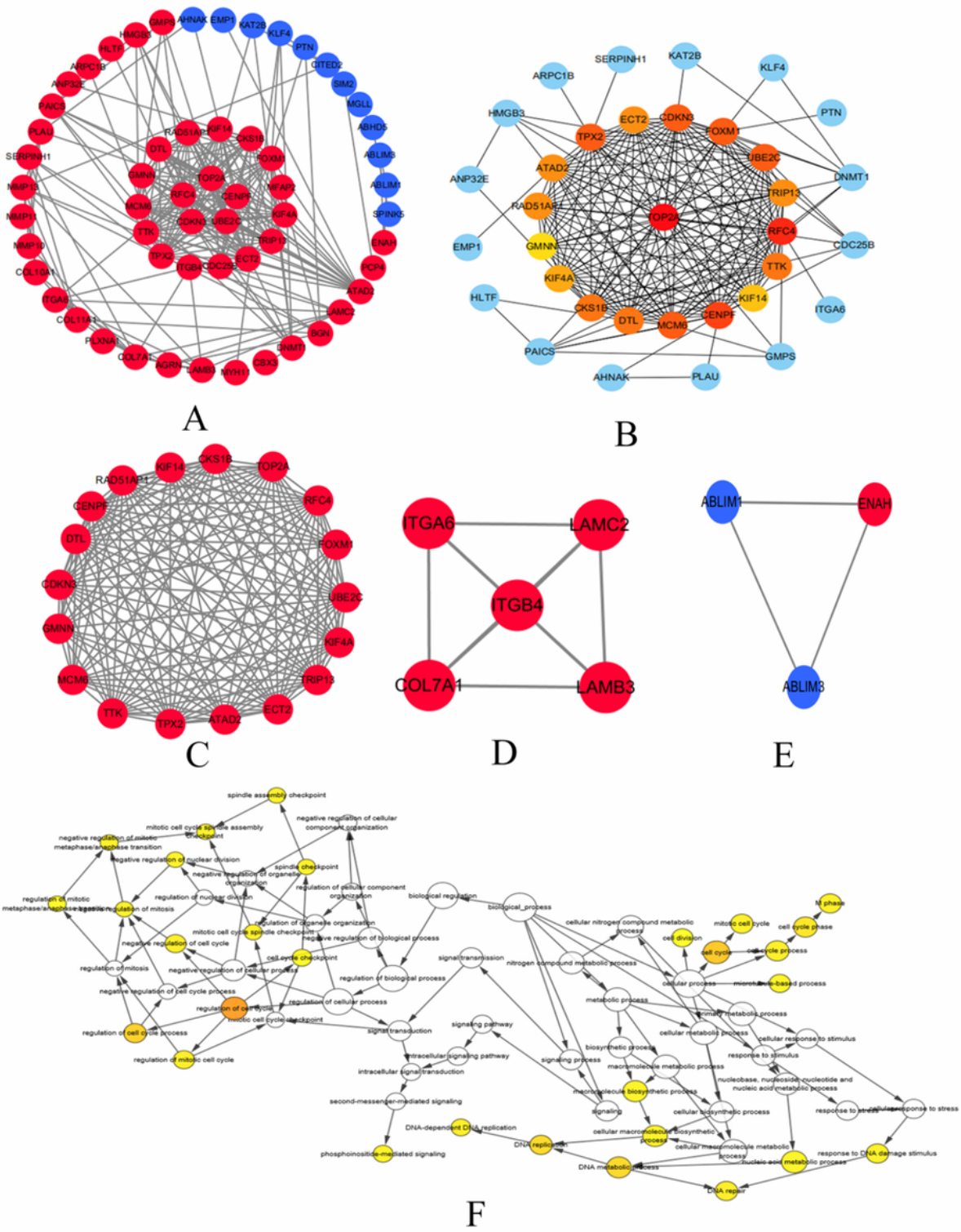

Figure 3

Protein-protein interaction (PPI) network.(a) The PPI network of overlapping DEGs; blue circles meant down-regulated DEGs and redcircles meant up-regulated DEGs. (b) The PPI network of hub genes; red, orange, yellow circles represente the central gene;□c□Module 1 (MCODE score =17.294). (d) Module 2 (MCODE score = 5.000). (e)Module 3 (MCODE score = 3.00) .□f□The biological process analysis of hub genes was constructed using BiNGO.

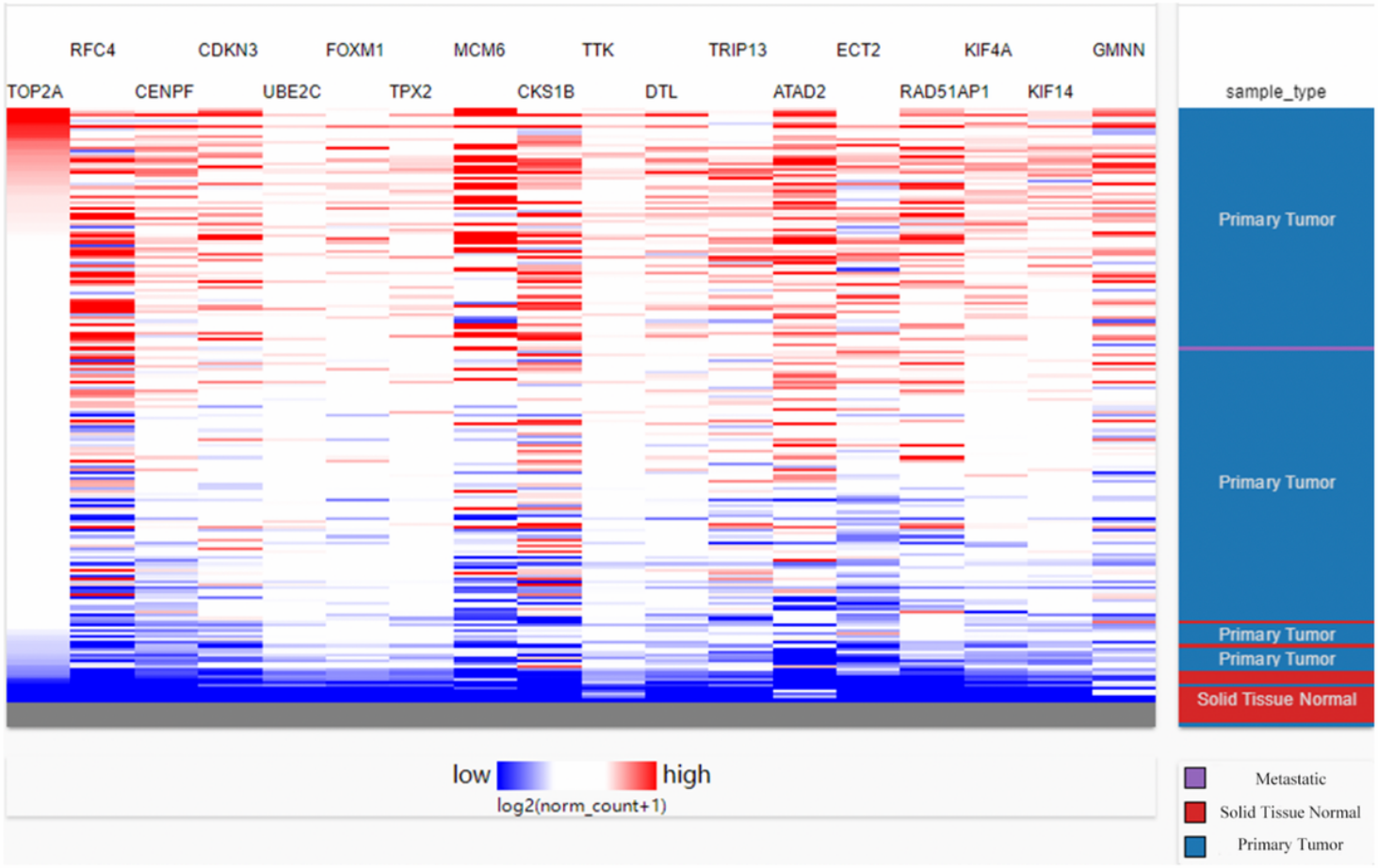

Figure 4

Hierarchical clustering of hub genes was constructed using UCSC.

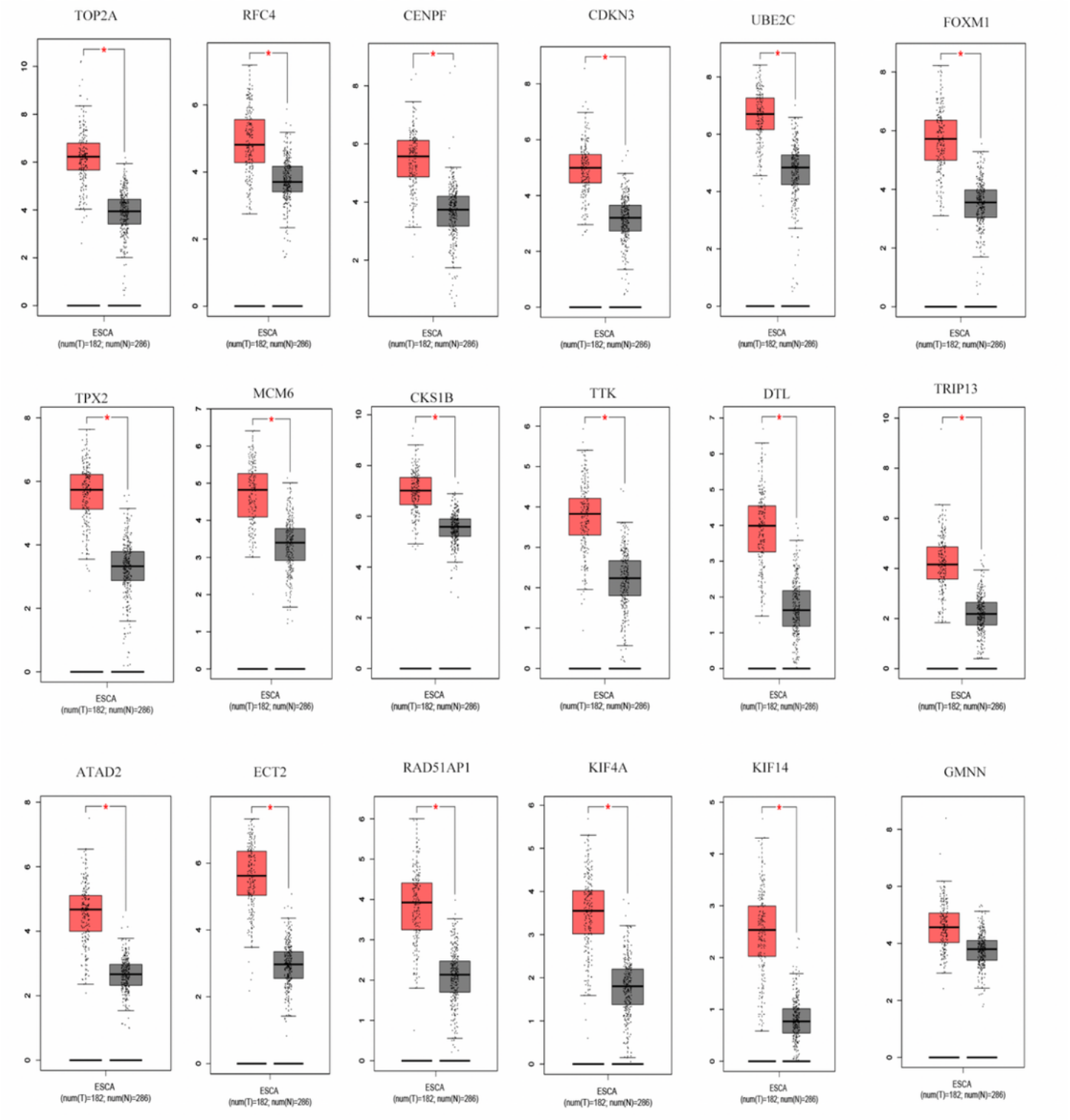

Figure 5

Expression of hub genes in ESCC and normal tissues using GEPIA

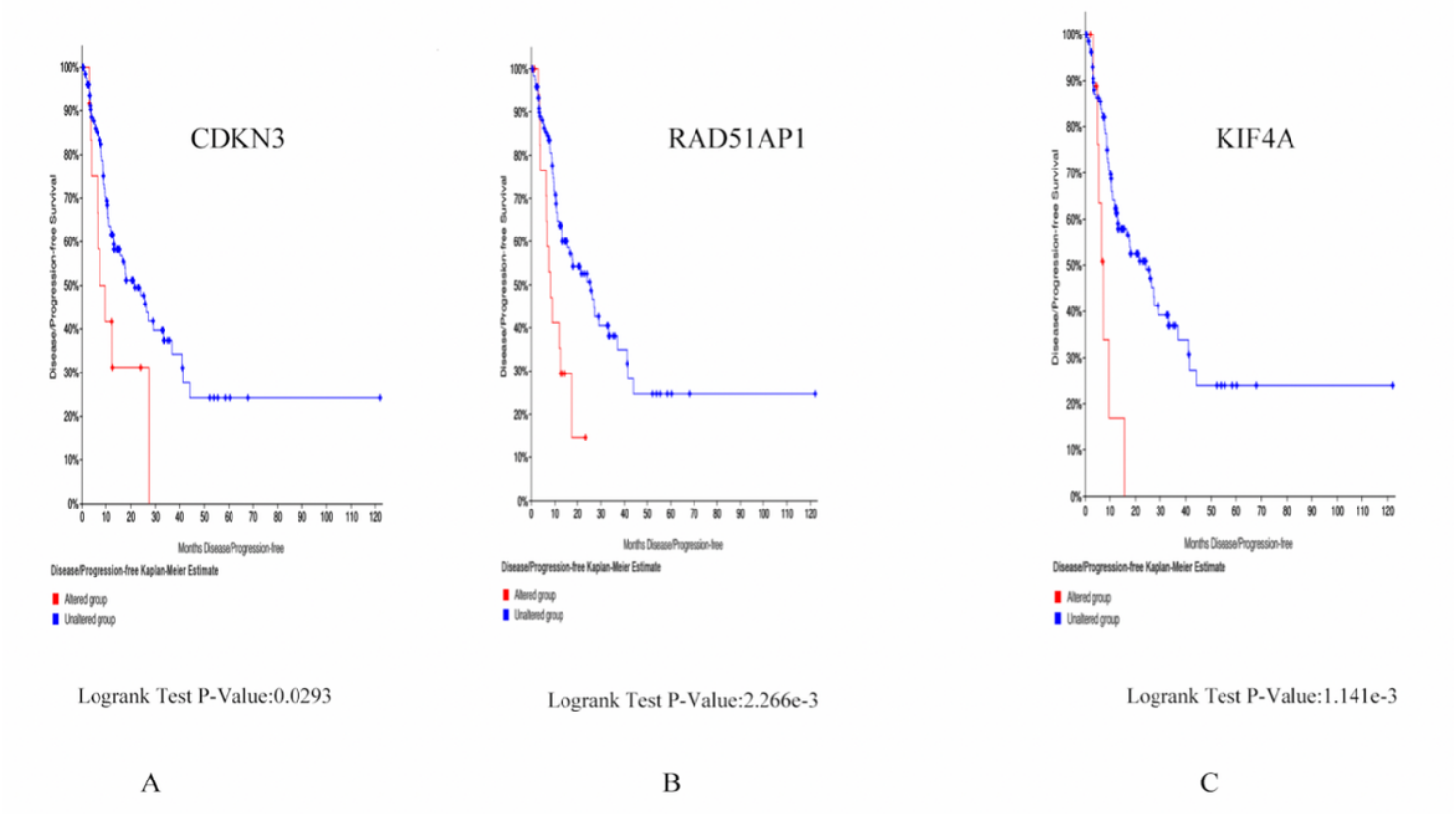

Figure 6

Prognostic value of hub genes in ESCC obtained from GEPIA. High expression levels of CDKN3 (a)„ RAD51AP1(b), KIF4A(c)were associated with poor disease-free survival.