Review and screening of key genes in esophageal squamous cell carcinoma

Xizi Sun (sxzintongji@126.com)
Huazhong University of Science and Technology

Research Article

Keywords: esophageal squamous cell carcinoma (ESCC), cancer, genetics, mutated genes

Posted Date: May 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-498257/v1

Abstract

Esophageal squamous cell carcinoma (ESCC) is the most common type of human esophageal cancer with high mortality due to late stage diagnosis. Efforts have been made to figure out the genetic events underlying its carcinogenesis and progression, but the molecular mechanisms of these processes remain elusive. To identify the candidate genes involved in ESCC, literature about significantly mutated genes (SMGs) was extensively reviewed and gene expression profiles of GSE161533, GSE20347 and GSE77861 were downloaded from the Gene Expression Omnibus (GEO) database. Following the identification of 230 differentially expressed genes (DEGs), hub gene identification was performed by the plug-in MCODE in Cytoscape software. 14 hub genes were identified which were enriched in cell cycle, DNA replication and p53 signaling pathway. In summary, genes mentioned in this study may provide potential targets for treatment and diagnosis of ESCC and help us better understand the pathogenesis and progression of ESCC from genetic perspective.

Introduction

Esophageal cancer is the eighth most common cancer and the sixth leading cause of cancer-related mortality in the world [1,2]. ESCC is the major histological type accounting for about 90% of the 456,000 incident esophageal cancers each year [3]. The 5-year survival rate for ESCC is about 18%, a number that reflects limited approaches of early diagnosis and treatment of ESCC [4]. Thus, there is a great need to further figure out the molecular mechanisms and to develop better diagnostic and therapeutic methods for ESCC. The pathogenesis of ESCC is believed to be a multi-step process and the genetic determinants remain elusive. Increasing evidence shows that gene mutation plays a key role in ESCC tumorigenesis and tumor progression. These genes include upregulated genes *ADAM29*, *AJUBA*, *CBX4/8, CCND1(BCL1/PRAD1), EGFR(ERBB1), ERBB2(HER-2), FAM135B, FGFR1, KMT2D(MLL2/MLL4/ALR), MMP14, MYC, NOTCH, NRF2(NFE2L2), PIK3CA, RB1, SOX2, TP53, XPO1, YAP1* and downregulated genes *CDKN2A, CREBBP/EP300, CUL3, FAT1, FBXW7, KMT2C(MLL3), PTEN, TET2, TGFBR2, ZFP36L2, ZNF750* [5-11](Table 1). Genes involved in cell cycle, the Notch signaling pathway, epigenetic processes and RTK/PI3K/AKT circuit are frequently altered [12]. Cell cycle progression is changed mostly by *TP53* mutation, *CDKN2A* deletion/mutation and *CCND1* amplification [5]. TP53 is the most significantly mutated genes (SMGs) in ESCC with mutation frequency reaching 93% [12]. NOTCH plays a dual role as both a tumor suppressor pathway and an oncogenic pathway, for which further studies are warranted [13]. Abudureheman et al. have shown that overexpression of *KMT2D* facilitates ESCC tumor progression, and that it may exert oncogenic role via activation of epithelial-to-mesenchymal transition (EMT) [14]. In a large-sample study with ESCC in China [15], *PIK3CA* was significantly overexpressed in cancer tissue and its overexpression was independently associated with higher risk of local recurrence [15]. EGFR and FGFR1 were the most often amplified RTK/RAS-related genes in ESCC [9], of which the inhibitors have been under therapeutic evaluation [16,17]. It's worth noting that conflicting results were found in studies about prognostic value of TP53 overexpression in ESCC [18-21], of which the biological functions were undoubted.

In general, there are still significant gaps to fill to figure out the exact mechanism of carcinogenesis and to develop precision treatment means of ESCC.

Gene chip or gene profile is an advanced gene detection technique that can quickly detect all the genes within the same sample at one time [22]. Last two decades have seen more and more studies of genetic alterations in cancers via microarray technology and bioinformatics analysis, which have helped us identify the differentially expressed genes (DEGs) and related pathways in ESCC. However, the results were always limited or inconsistent because of tissue or sample heterogeneity in independent studies, or the results were produced from a single cohort study. Thus, integrated bioinformatics analysis combined with gene profiling technique might be innovative and solve this disadvantage. In this work, we downloaded three microarray datasets GSE161533, GSE20347, GSE77861 from NCBI-Gene Expression Omnibus database (NCBI-GEO) (Available online: https://www.ncbi.nlm.nih.gov/geo) followed by DEGs identification via GEO2R analysis. Subsequently, Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis and protein-protein interaction (PPI) network analysis were conducted to help us understand the molecular mechanisms of carcinogenesis and progression of ESCC. In summary, 230 DEGs and 14 hub genes were found in this study, which may serve as potential biomarkers for individualized prevention, early diagnosis and precise treatment.

Table1. Expression of significantly mutated genes (SMGs) and their functions in ESCC

Gene Symbol	Expression in ESCC	Function of the Product	Reference
ADAM29	upregulated	a member of ADAM family which plays an important role in regulating cell-to-cell or cell-matrix interactions	6,23
AJUBA	upregulated	functions as a regulator of the Hippo pathway	7,24
BAP1	lack study	a deubiquitinating enzyme involved in the regulation of cell growth	25
CBX4/8	upregulated	promotes cell proliferation, colony formation,and cell invasion	8
CCND1 (BCL1/PRAD1)	upregulated	regulation of cell cycle	7,26
CDKN2A	downregulated	regulation of cell cycle	9
CREBBP	lack study	Serves as global transcriptional coactivators and integrators of numerous signaling pathways	7
CUL3	downregulated	core component of multiple cullin-RING-based BCR (BTB-CUL3-RBX1) E3 ubiquitin-protein ligase complexes	10,11
DCDC1	lack study	microtubule-binding protein which plays an important role in mediating dynein-dependent transport	11
EGFR(ERBB1)	upregulated	stimulates proliferation of different cell types	27,28
EP300	upregulated	functions as histone acetyltransferase and regulates transcription via chromatin remodeling	29
ERBB2 (HER-2/NEU/NGL/MLN19)	upregulated	regulates cell growth and differentiation	30
FAM135B	upregulated	promotes cell proliferation likely through its direct interaction with growth factor GRN	6,23,31
FAT1	downregulated	inhibits cell proliferation and migration	32
FBXW7	downregulated	one of the F-box proteins inducing the degradation of positive cell-cycle regulators	33-35
FGFR1	upregulated	regulation of cell proliferation, differentiation, migration, and angiogenesis	5,36
KDM6A(UTX)	lack study	a histone demethylase essential for cellular reprogramming	37
KMT2C(MLL3)	downregulated	regulation of expression of genes involved in cell growth and migration	38
KMT2D (MLL2/MLL4/ALR)	upregulated	histone methyltransferase	7,14
MMP14	upregulated	endopeptidase that degrades various components of the extracellular matrix	24
MYC	upregulated	transcription factor that binds DNA in a non-specific manner	39

NOTCH	upregulated	regulation of squamous differentiation	13
NRF2(NFE2L2)	upregulated	transcription factor that plays a key role in the response to oxidative stress	40,41
PIK3CA	upregulated	the p110α catalytic subunit of PI3K	15,42,43
PTCH	lack study	a key modulator of signaling in the Hh pathway	44
PTEN	downregulated	a phosphatase which inhibits cell migration, spreading and focal adhesions	45
RB1	upregulated	inhibits cell migration and invasion	46
SOX2	upregulated	transcriptional regulator having crucial roles in maintenance of progenitor and neural stem cells and neuroendocrine differentiation	8,47
TET2	downregulated	dioxygenase that catalyzes the conversion of the modified genomic base 5-methylcytosine (5mC) into 5-hydroxymethylcytosine (5hmC)	48,49
TGFBR2	downregulated	a key mediator of TGF-β signaling	50,51
TP53	upregulated	regulation of cell cycle, apoptosis and DNA damage repairing	18-21
XPO1	upregulated	mediates the nuclear export of cellular proteins (cargos) bearing a leucine-rich nuclear export signal (NES) and of RNAs	5
YAP1	upregulated	the critical downstream regulatory target in the Hippo signaling pathway that plays a pivotal role in organ size control and tumor suppression	24,52
ZFP36L2	downregulated	zinc-finger RNA-binding protein that destabilizes several cytoplasmic AU-rich element (ARE)-containing mRNA transcripts	10
ZNF750	downregulated	transcription factor involved in epidermis differentiation	53

Results

Identification of DEGs in ESCC

After analyzing with GEO2R with adj. P value < 0.01, |log FC| > 1, DEGs (1504 in GSE161533, 1680 in GSE20347 and 972 in GSE77861) were identified. The intersection of three gene sets of DEGs contains 230 genes as shown in venn diagram (Fig.1), which consists of 144 upregulated genes and 86 downregulated genes between normal and ESCC tissues.

GO and KEGG enrichment analysis of DEGs

To annotate the DEGs, GO and KEGG enrichment analysis were performed using DAVID, with P value<0.05 considered significant. GO analysis results (Fig.2) showed that changes in biological processes (BP) were mainly enriched in oxidation-reduction process, positive regulation of cell proliferation, cell-cell adhesion

and inflammatory response. Changes in cellular components (CC) of DEGs were significantly enriched in cytoplasm, extracellular exosome, cytosol and extracellular space. Changes in molecular function (MF) of DEGs were enriched in calcium ion binding, protein homodimerization activity, cadherin binding involved in cell-cell adhesion and actin binding. And KEGG analysis showed that DEGs mainly enriched in transcriptional misregulation in cancer and p53 signaling pathway.

PPI network construction and hub genes identification

Prediction of the functional interaction was conducted by STRING online and the PPI network of DEGs was constructed by Cytoscape (Fig.3). Subsequently, 14 hub genes were identified with MCODE score > 10 (Fig.4).

Hub gene analysis

Among all hub genes, only ESPL1 is down-regulated. GO and KEGG analysis network of hub genes was performed using ClueGO (Fig.5). Result showed that hub genes were mainly enriched in positive regulation of mitotic cell cycle phase transition, regulation of cytokinesis and DNA replication origin binding. Subsequently, we conducted an extensive literature search on the hub genes.

AURKA, which is significantly overexpressed in various cancers including ESCC [54] has been reported to contribute to the malignant development of ESCC [55] and Jin et al. have revealed part of the mechanism underlying this process [56]. Ke et al. showed that downregulated CDC6 functioned as downstream molecule of RYBP in the inhibition of cell proliferation in ESCC [57]. An association between overexpression of *DTL* and detrimental outcome in basal-like and luminal breast cancer and non-small cell lung adenocarcinomas but not esophagus-stomach cancer was found in another bioinformatic analysis [58]. The gene *ECT2* encoding guanine nucleotide exchange factor (GEF) functions as an oncogene in a wide spectrum of cancers [59]. Sun et al. [60] showed that ECT2 promoted proliferation and metastasis of ESCC via the RhoA-ERK signaling pathway. And the overexpression of ECT2 often suggests a poor outcome in cancers, especially in breast cancer [61,62] , gastric cancer [63], non-small-cell lung cancer [64,65], and ESCC [65]. *ESPL1*, the dysregulation of which plays an important role in the development of aneuploidy [66], is a candidate oncogene in luminal B breast cancers [67]. Studies have revealed the significant association between high expression of FOXM1 and poor outcome of ESCC [68,69]. GTSE1 was found to promote malignant behavior in hepatocellular carcinoma (HCC) [70], confer to cisplatin resistance in gastric cancer cells [71] and to be involved in breast cancer progression [72]. Overexpression of *KIF14* in ESCC was validated in a study, in which *KIF14* was a downstream gene regulated by the *miR-375/MMP13* axis [73]. The overexpression of MCM10 in ESCC has been confirmed in an experiment using semiquantitative reverse transcription-PCR [74]. MCM2, of which the diagnostic value has been confirmed in different types of cancers [75-77] was found to be a more reliable and useful marker than Ki67 in assessing tumor growth and tumor aggressiveness in patients with ESCC [78] and in screening patients at high risk of ESCC in mass surveys [79]. RFC4 has been found to be overexpressed in several cancers [80], while study about its

role in ESCC remains blank. *RRM1* has been found to be an oncogene in lung cancer [81], the overexpression of which is involved in tumor progression [82] and is transforming to the therapeutic target [83,84]. A large-scale, long-term follow-up retrospective analysis [85] showed that TOP2A expression was not only associated with perineural invasion and poorer differentiation, but it could be also an independent prognostic factor. Additionally, as TOP2A is a specific marker for the use of chemotherapeutic drugs such as anthracycline, therapy targeting TOP2A protein may be an appropriate way of individualized treatment and improving the prognosis of ESCC patients. Studies [86,87] using immunohistochemical analysis confirmed that UBE2C protein expression was upregulated in all ESCC cases, but absent in the histologically normal tumor surrounding tissues, pointing out its role as a diagnostic biomarker for ESCC. Besides, high expression of UBE2C is a marker of poor prognosis in ESCC [87].

Table2. Fuction and clinical significance of 14 hub genes with MCODE score ≥10. "→" indicates "was significantly associated with"

Gene Symbol	Protein Name	Biological Function of the Product	Clinical Significance in ESCC
AURKA	Aurora kinase A	Mitotic serine/threonine kinase that contributes to the regulation of cell cycle progression	higher expression → poorer prognosis
CDC6	Cell division control protein 6 homolog	essential to initiate the DNA replication and G1-S transition	lack study
DTL	Denticleless protein homolog	Substrate-specific adapter of a complex required for cell cycle control,DNA damage response and translesion DNA synthesis.	lack study
ECT2	Epithelial cell-transforming sequence 2 oncogene	Guanine nucleotide exchange factor (GEF) that catalyzes the exchange of GDP for GTP	higher expression → poorer prognosis
ESPL1	Extra spindle poles-like 1	separate sister chromatids through proteolytic cleavage of cohesin protein Rad21 during the metaphase to anaphase transition	lack study
FOXM1	Forkhead box protein M1	transcriptional factor regulating the expression of cell cycle genes essential for DNA replication and mitosis	higher expression → poorer prognosis
GTSE1	G2 and S phase-expressed-1	involved in p53-induced cell cycle arrest in G2/M phase by interfering with microtubule rearrangements that are required to enter mitosis	lack study
KIF14	Kinesin-like protein KIF14	Microtubule motor protein that binds to microtubules with high affinity	lack study
MCM10	Minichromosome maintenance protein 10	Acts as a replication initiation factor that initiates DNA replication	lack study
MCM2	Minichromosome maintenance protein 2	Acts as component of the MCM2-7 complex which is the putative replicative helicase essential for 'once per cell cycle' DNA replication initiation and elongation	higher expression → more aggressiveness & poorer prognosis
RFC4	Replication factor C subunit 4	subunit of a complex required for DNA replication which functions by "load" onto the proliferating cell nuclear antigen (PCNA)	lack study
RRM2	Ribonucleoside-diphosphate reductase subunit M2	subunit of the enzyme that reduces ribonucleoside diphosphates (NDPs) to deoxyribonucleoside diphosphates (dNDPs)	lack study
TOP2A	DNA topoisomerase 2-alpha	a key enzyme involved in DNA replication	higher expression → poorer differentiation & perineural invasion & poorer prognosis
UBE2C	ubiquitin conjugating enzyme E2	a cyclin-selective ubiquitin carrier protein required in the cell-cycle transition from metaphase to anaphase	accurate diagnosis higher expression → poorer prognosis

Figures

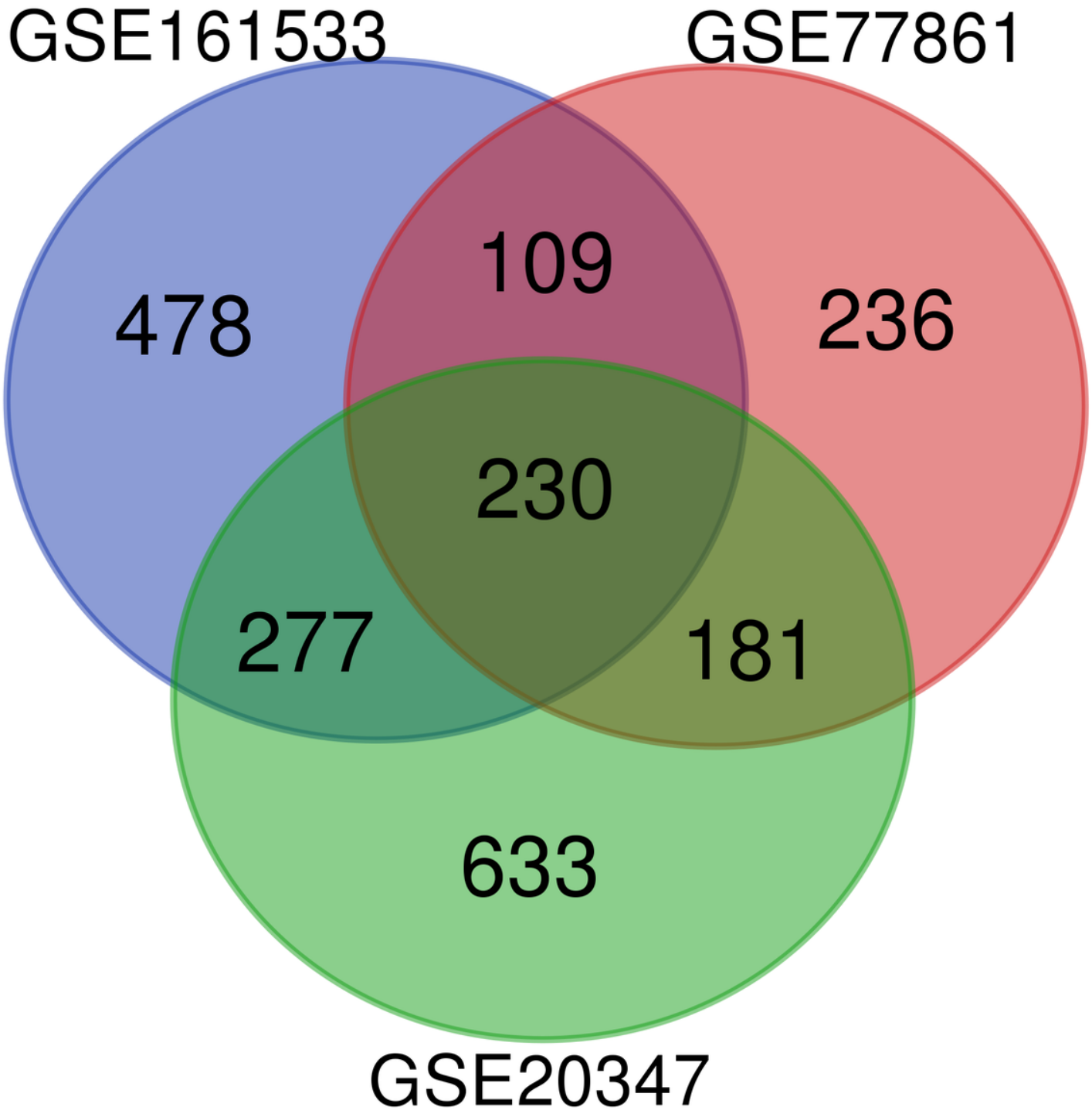

Figure 1

Venn Diagram

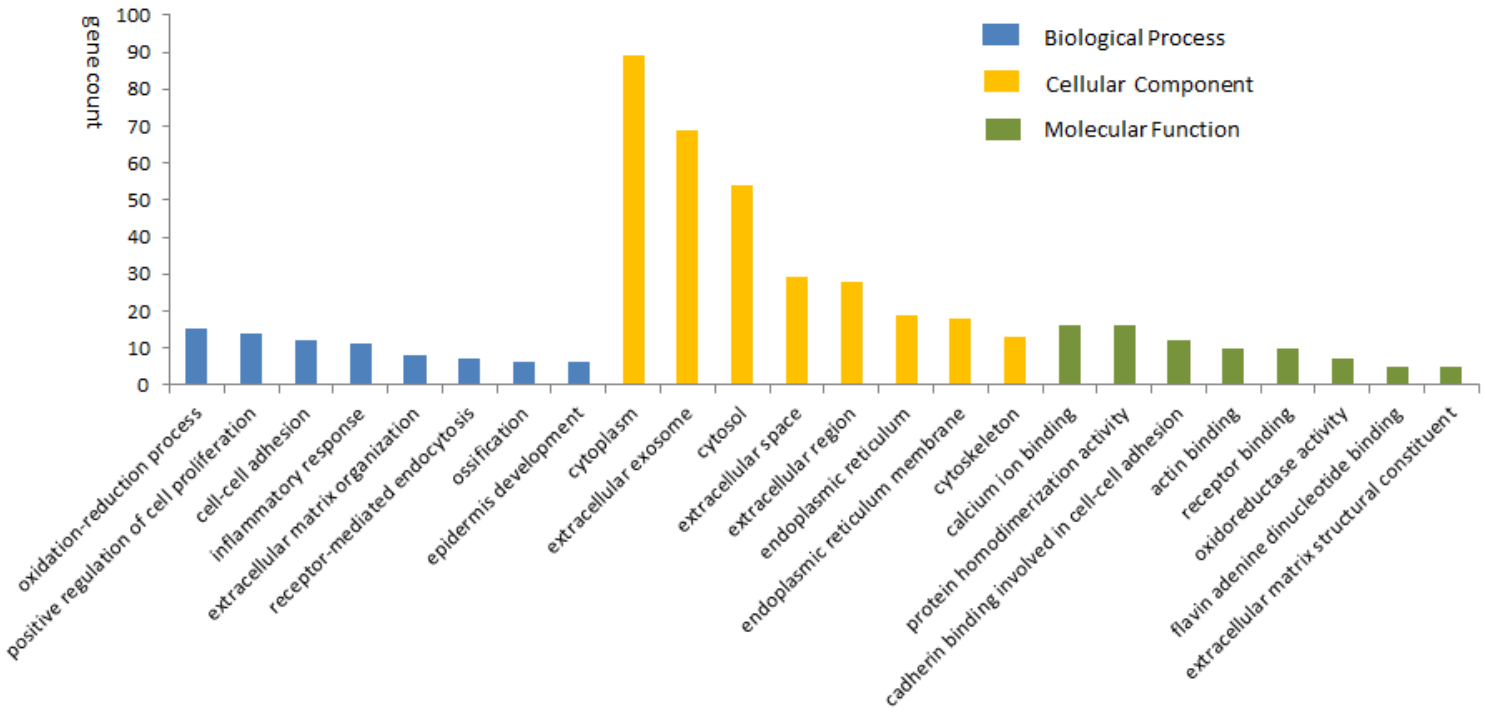

Figure 2

GO enrichment analysis of DEGs in ESCC samples.

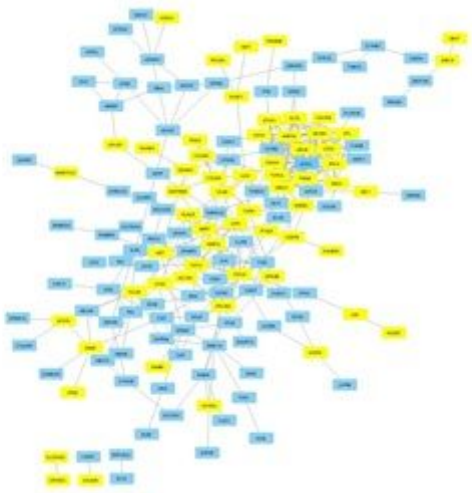

Figure 3

PPI network of DEGs. Upregulated genes are marked in yellow; downregulated genes are marked in blue.

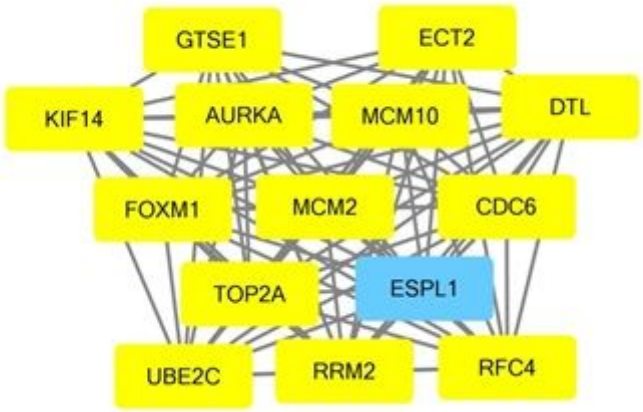

Figure 4

Hub genes.

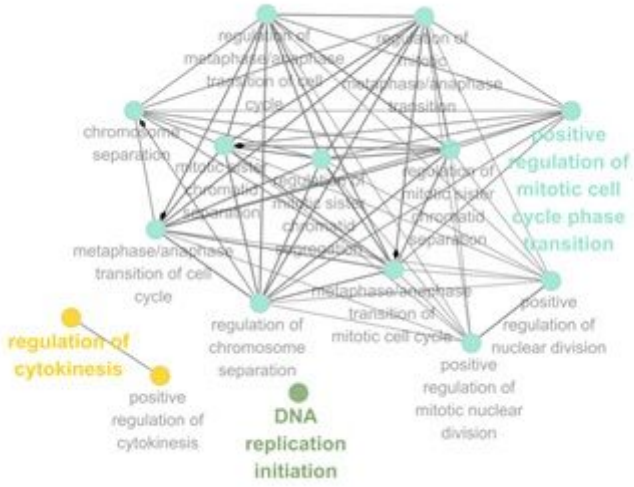

Figure 5

GO of hub genes.