Research Paper

Microarray analyses reveal genes related to progression and prognosis of esophageal squamous cell carcinoma

Mao Qixing[1,2,3,*], Dong Gaochao[1,3,*], Xia Wenjie[1,2,3,*], Wang Anpeng[1,2,3], Chen Bing[1,2,3], Ma Weidong[1,2,3], Xu Lin[1,3] and Jiang Feng[1,3]

[1]Department of Thoracic Surgery, Jiangsu Cancer Hospital, Jiangsu Institute of Cancer Research, Nanjing Medical University Affiliated Cancer Hospital, Nanjing, China

[2]The Fourth Clinical College of Nanjing Medical University, Nanjing, China

[3]Jiangsu Key Laboratory of Molecular and Translational Cancer Research, Nanjing Medical University Affiliated Cancer Hospital, Cancer Institute of Jiangsu Province, Nanjing, China

[*]These authors have contributed equally to this work

Correspondence to: Xu Lin, *email:* xulin_83@hotmail.com

Jiang Feng, *email:* zengnljf@hotmail.com

ABSTRACT

Esophageal squamous cell carcinoma is a high morbidity and mortality cancer in China. Here are few biomarkers and therapeutic targets. Our study was aimed to identify candidate genes correlated to ESCC. Oncomine, The Cancer Genome Atlas, Gene Expression Omnibus were retrieved for eligible ESCC data. Deregulated genes were identified by meta-analysis and validated by an independent dataset. Survival analyses and bioinformatics analyses were used to explore potential mechanisms. Copy number variant analyses identified upstream mechanisms of candidate genes. In our study, top 200 up/down-regulated genes were identified across two microarrays. A total of 139 different expression genes were validated in GSE53625. Survival analysis found that nine genes were closely related to prognosis. Furthermore, Gene Ontology analyses and Kyoto Encyclopedia of Genes and Genomes analyses showed that different expression genes were mainly enriched in cell division, cell cycle and cell-cell adhesion pathways. Copy number variant analyses indicated that overexpression of *ECT2* and other five genes were correlated with copy number amplification. The current study demonstrated that *ECT2* and other eight candidate genes were correlated to progression and prognosis of esophageal squamous cell carcinoma, which might provide novel insights to the mechanisms.

INTRODUCTION

Esophageal squamous cell carcinoma (ESCC), one main subtype of esophageal cancer, accounts for more than 90% of esophageal cancer cases in China [1, 2]. Based on the National Central Cancer Registry of China (NCCR), esophageal cancer is 4th lethal cancer in China [3]. Over the past decades major progress has been made in diagnosis and treatment, the morbidity and mortality of ESCC present declining tendency [2, 4]. However, 5-year survival of ESCC is still less than 25%, which is dismal [5]. Thus, further understanding the potential biological mechanisms and exploring novel biomarkers are urgently needed.

Microarray technology provides a high-throughput and powerful tool to create large amounts of data including mRNA expression, DNA methylation, and microRNA expression [6, 7]. Published studies have identified many novel tumor biomarkers and potential therapeutic targets by microarray data [8–10]. The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) are two common public platforms archiving these data [11]. The

wide range of microarrays in public archives provided an opportunity to reuse several microarrays to identify several novel targets [7, 12]. Oncomine currently includes gene expression and sample data from 500 cancer types [13]. There are more than 490 datasets and nearly 40,000 measured samples [14]. The data are normalized and analyzed using standard protocols. Based on the large amount data, Oncomine provides several analysis tools including differential expression analysis, co-expression analysis and comparing analysis [15]. For example, Clermont reported oncogenic role for CBX2 in pan-cancer [16]. Barfeld's investigation revealed a novel signature for prostate cancer [17]. However, to our knowledge, few such studies have been conducted in ESCC. To fill in this gap, we conducted a study using Oncomine analysis tool to identify novel differential expression genes and potential pathways, which might provide new therapeutic targets for ESCC.

RESULTS

Meta analysis of significantly deregulated genes in ESCC

There were two eligible independent ESCC microarray datasets in oncomine, which were Hu's ESCC cohort with 34 cases and Su's ESCC cohort covering 106 cases [18, 19]. A meta-analysis was performed across two microarray datasets. Two microarray datasets showed highly consistent trend of mRNA expression in top ranked genes (Figure 1). We selected 200 of the top ranked genes with the smallest P-values from the in over/under expression patterns (p<7.23E-07 and p<8.04E-12, fold change>2, respectively) (Supplementary Tables 1&2). In addition, different expression genes with fold change>2 were identified from GSE53625 dataset for validation. Overlap between results of Oncomine with outcome of GSE53625 showed that 139 genes were consistently and stably expression in ESCC (Figure 2). The flow chart of this study was showed in Figure 3.

Candidate genes correlated with survival

Combining with survival information of GSE53625, we next accessed the prognostic effect of candidate genes in ESCC. Log rank test indicated that nine genes were significant correlated with survival. Among them, we observed that up regulated of six genes (*ECT2, TFRC, TOPBP1, NETO2, PTDSS1, ITGA6*) were correlated with poor survival, while down regulated of three other genes (*MGLL, TP53I3, TRIP10*) were correlated with shorter survival (Figure 4). A comprehension literature search showed that *ECT2, TFRC* and *ITGA6* were reported to be prognostic biomarkers by previous ESCC investigations.

Additionally, knockdown of *ECT2* and *ITGA6* expression suppressed esophageal cancer cell growth and proliferation (Table 1). However, few investigations reported the other six genes in carcinogenesis and development of ESCC.

The interrelationship between these candidate genes was further explored by GeneMANIA tool. GeneMANIA tool provided with network relationship of candidate genes by co-expression, shared protein domains and co-localization relationships based on published papers. The network, constructed of 27 nodes and 133 edges, showed a tightly correlation between candidate genes with other 18 key functional molecules. The network indicated potential mechanisms of ESCC (Supplementary Figure 1 & Supplementary Table 3).

Potential roles of candidate genes

In order to explore the unique roles of each candidate genes, we performed Gene Ontology (GO) and KyotoEncyclopedia of Genes and Genomes (KEGG) pathway analysis by DAVID tool. We found that cell division, mitotic nuclear division, DNA replication and G1/S transition of mitotic cell cycle were enriched by a high p-value in *ECT2* analysis, implying a potential role of *ECT2* in cell division and cell cycles (Figure 5A). Additionally, we observed that *NETO2, ITGA6* and *TOPBP1* shared the same critical pathways like cell division, DNA repaired, cell cycles and cell proliferation (Figure 5B, 5C&5E). It indicated that *ECT2, NETO2, ITGA6* and *TOPBP1* might function as oncogenes to promote the progression of ESCC by regulating these pathways. Notably, *NETO2* and *ITGA6* also took part in the p53 class mediator signal transduction that was a critical carcinogenic pathway. GO analysis unveiled that *MGLL* and *TRIP10* participated in cell-cell adhesion, regulation of Rho protein signal transduction and positive regulation of GTPase activity, indicating that *MGLL* and *TRIP10* might regulate the invasion and adhesion ability of ESCC (Figure 5D&5F). Moreover, *TRIP10* was significantly correlated with fatty acid beta-oxidation using acyl-CoA oxidase pathway, while *MGLL* play a role in oxidation-reduction process, which were significantly correlated with carcinogenesis of ESCC. To investigate the potential altered pathways in the samples, Gene Set Enrichment Analysis (GSEA) was implemented between live/death groups. According to the result of *ECT2,* genes were mainly enriched in cell cycle process and mitotic nuclear division (Figure 6A). Moreover, we noticed that regulation of growth and DNA replication pathways were significantly enriched in *TOPBP1* (Figure 6B). As to *TFRC*, co-expression genes were enriched in cytoskeletal protein binding and spliceosomal complex pathways (Figure 6C).
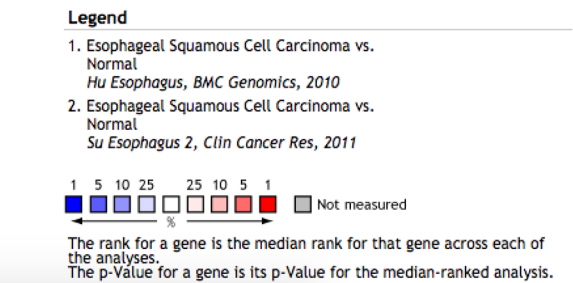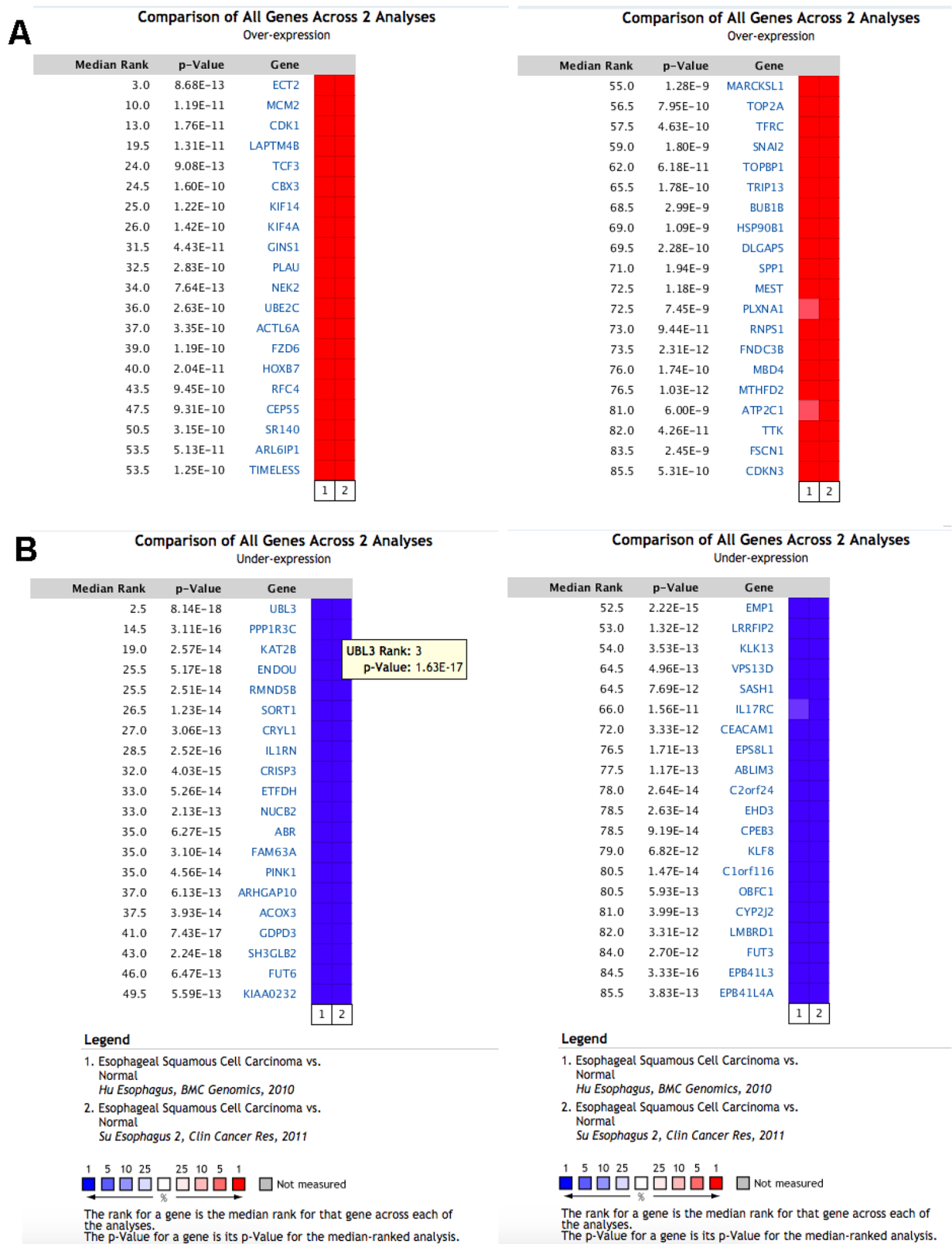
A

Comparison of All Genes Across 2 Analyses
Over-expression

Median Rank	p–Value	Gene
3.0	8.68E–13	ECT2
10.0	1.19E–11	MCM2
13.0	1.76E–11	CDK1
19.5	1.31E–11	LAPTM4B
24.0	9.08E–13	TCF3
24.5	1.60E–10	CBX3
25.0	1.22E–10	KIF14
26.0	1.42E–10	KIF4A
31.5	4.43E–11	GINS1
32.5	2.83E–10	PLAU
34.0	7.64E–13	NEK2
36.0	2.63E–10	UBE2C
37.0	3.35E–10	ACTL6A
39.0	1.19E–10	FZD6
40.0	2.04E–11	HOXB7
43.5	9.45E–10	RFC4
47.5	9.31E–10	CEP55
50.5	3.15E–10	SR140
53.5	5.13E–11	ARL6IP1
53.5	1.25E–10	TIMELESS

Comparison of All Genes Across 2 Analyses
Over-expression

Median Rank	p–Value	Gene
55.0	1.28E–9	MARCKSL1
56.5	7.95E–10	TOP2A
57.5	4.63E–10	TFRC
59.0	1.80E–9	SNAI2
62.0	6.18E–11	TOPBP1
65.5	1.78E–10	TRIP13
68.5	2.99E–9	BUB1B
69.0	1.09E–9	HSP90B1
69.5	2.28E–10	DLGAP5
71.0	1.94E–9	SPP1
72.5	1.18E–9	MEST
72.5	7.45E–9	PLXNA1
73.0	9.44E–11	RNPS1
73.5	2.31E–12	FNDC3B
76.0	1.74E–10	MBD4
76.5	1.03E–12	MTHFD2
81.0	6.00E–9	ATP2C1
82.0	4.26E–11	TTK
83.5	2.45E–9	FSCN1
85.5	5.31E–10	CDKN3

B

Comparison of All Genes Across 2 Analyses
Under-expression

Median Rank	p–Value	Gene
2.5	8.14E–18	UBL3
14.5	3.11E–16	PPP1R3C
19.0	2.57E–14	KAT2B
25.5	5.17E–18	ENDOU
25.5	2.51E–14	RMND5B
26.5	1.23E–14	SORT1
27.0	3.06E–13	CRYL1
28.5	2.52E–16	IL1RN
32.0	4.03E–15	CRISP3
33.0	5.26E–14	ETFDH
33.0	2.13E–13	NUCB2
35.0	6.27E–15	ABR
35.0	3.10E–14	FAM63A
35.0	4.56E–14	PINK1
37.0	6.13E–13	ARHGAP10
37.5	3.93E–14	ACOX3
41.0	7.43E–17	GDPD3
43.0	2.24E–18	SH3GLB2
46.0	6.47E–13	FUT6
49.5	5.59E–13	KIAA0232

UBL3 Rank: 3
p-Value: 1.63E-17

Comparison of All Genes Across 2 Analyses
Under-expression

Median Rank	p–Value	Gene
52.5	2.22E–15	EMP1
53.0	1.32E–12	LRRFIP2
54.0	3.53E–13	KLK13
64.5	4.96E–13	VPS13D
64.5	7.69E–12	SASH1
66.0	1.56E–11	IL17RC
72.0	3.33E–12	CEACAM1
76.5	1.71E–13	EPS8L1
77.5	1.17E–13	ABLIM3
78.0	2.64E–14	C2orf24
78.5	2.63E–14	EHD3
78.5	9.19E–14	CPEB3
79.0	6.82E–12	KLF8
80.5	1.47E–14	C1orf116
80.5	5.93E–13	OBFC1
81.0	3.99E–13	CYP2J2
82.0	3.31E–12	LMBRD1
84.0	2.70E–12	FUT3
84.5	3.33E–16	EPB41L3
85.5	3.83E–13	EPB41L4A

Legend

1. Esophageal Squamous Cell Carcinoma vs. Normal
   Hu Esophagus, BMC Genomics, 2010
2. Esophageal Squamous Cell Carcinoma vs. Normal
   Su Esophagus 2, Clin Cancer Res, 2011

1 5 10 25   25 10 5 1
% →
Not measured

The rank for a gene is the median rank for that gene across each of the analyses.
The p-Value for a gene is its p-Value for the median-ranked analysis.

Legend

1. Esophageal Squamous Cell Carcinoma vs. Normal
   Hu Esophagus, BMC Genomics, 2010
2. Esophageal Squamous Cell Carcinoma vs. Normal
   Su Esophagus 2, Clin Cancer Res, 2011

1 5 10 25   25 10 5 1
% →
Not measured

The rank for a gene is the median rank for that gene across each of the analyses.
The p-Value for a gene is its p-Value for the median-ranked analysis.

Figure 1: The top 80 genes that were significantly deregulated in ESCC across two independent microarrays retrieved from the Oncomine database. (A) The top 40 genes were significantly up regulated. **(B)** The top 40 genes were significantly down regulated. The two microarrays were Su's ESCC Statistics (52 ESCC tissues and 53 normal tissues) and Hu's ESCC Statistics (17 ESCC tissues and 17 normal tissues). The genes labeled in red and in blue represent up/down-regulated in each microarray.

Protein–protein interaction (PPI) networks of candidate genes were based on Biogrid and String database (Supplementary Figure 2). A total of 160 proteins might combine with ECT2 protein, which were detected by high-throughput analysis (Figure 7A). Among them, we found several molecules like cyclin-dependent kinase 5 (CDK5), tumor protein P53 (TP53) and E-cadherin (*CDH1*), were important proteins in cancer signal transduction pathway. Additionally, TRIP10 might interact with 57 proteins including Rho GTPase activating protein 17 (ARHGAP17), cell division cycle 42 (CDC42) and signal transducer and activator of transcription 3 (STAT3) in ESCC (Figure 7B). As to TOPBP1, E2F transcription factor1 (E2F1), mediator of DNA-damage checkpoint1 (MDC1) and BRCA1 associated RING domain 1 (BRAD1) was found in the interaction network (Figure 7C).

Coremine Medical mining was performed to identify the potential roles of the 9 candidate genes in ESCC. As shown in Figure 8, there were significant connections between gene expression, phosphorylation, cell proliferation and signal transduction of ESCC with candidate genes. For instance, *TRIP10, TP53I3, ECT2, ITGA6, TFRC,*
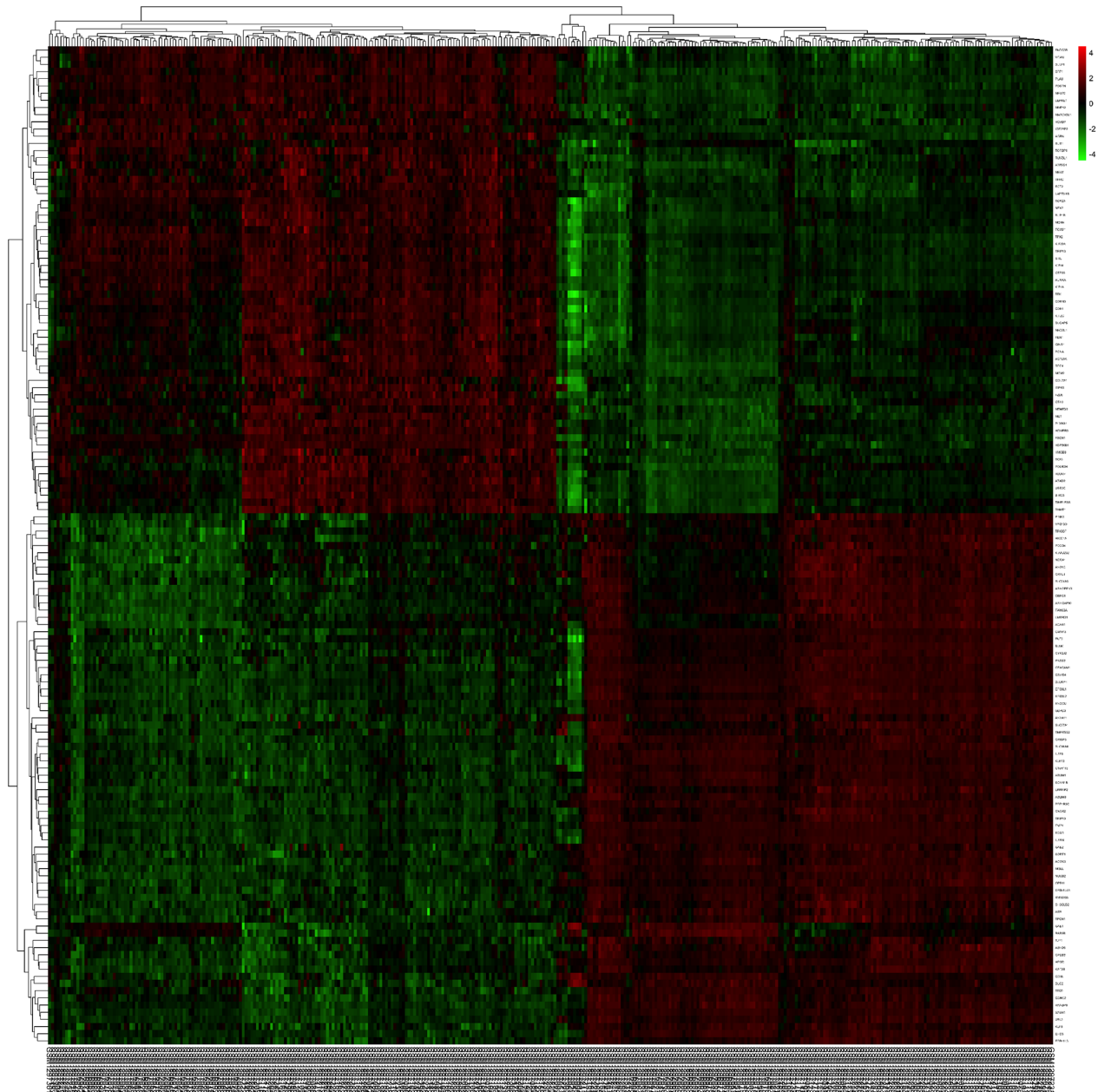
Figure 2: The heatmap revealed the overlapped differentially expressed genes between tumor and normal samples.
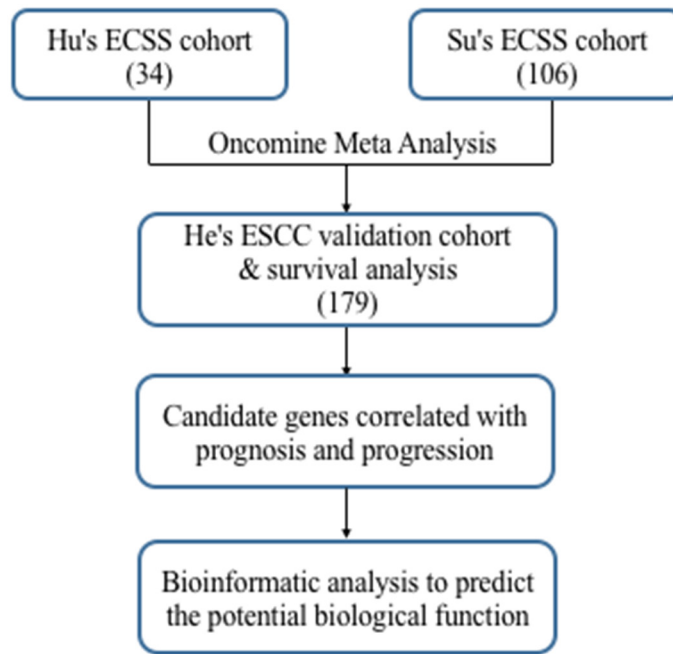
Figure 3: Flowchart for comprehensive analysis of the expression profiles and identification of the candidate genes correlated with progression and prognosis.
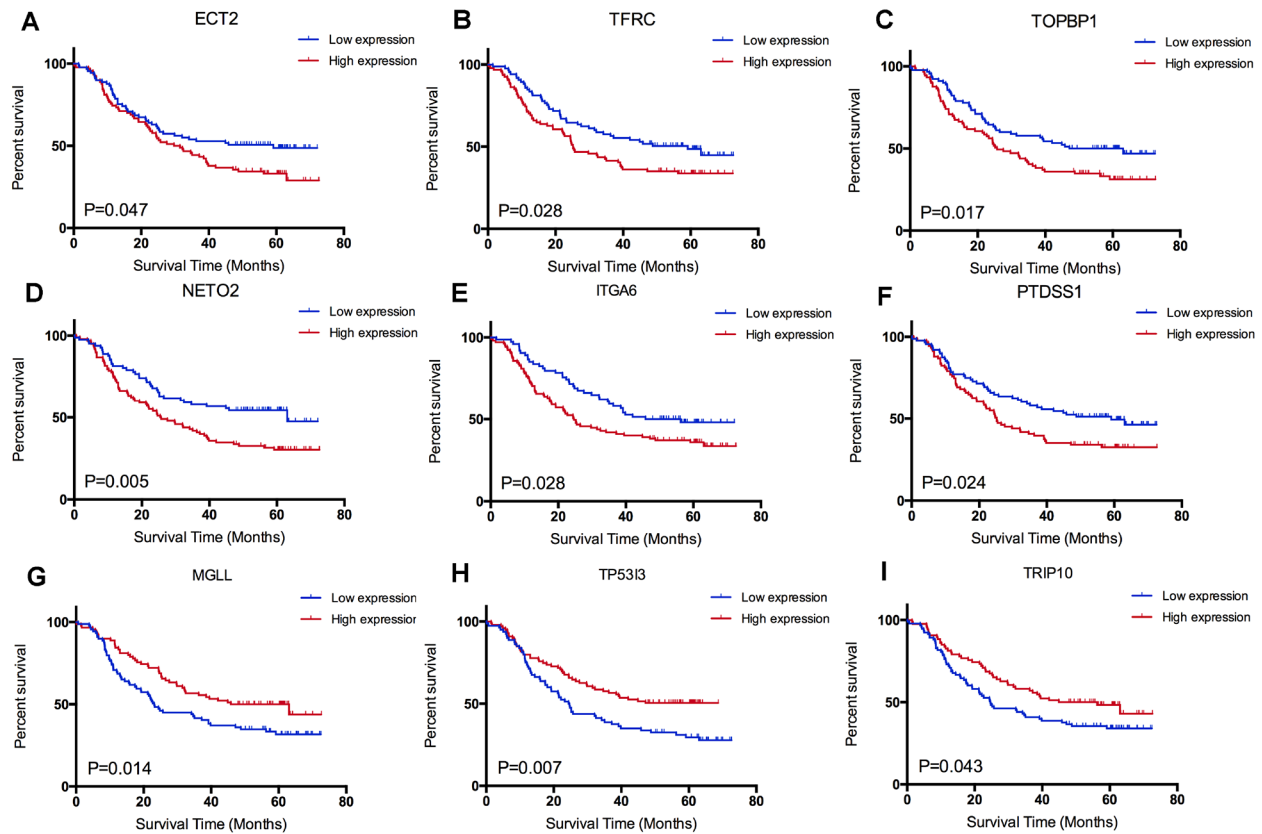
Figure 4: Survival analysis of candidate genes: **(A)** *ECT2*, **(B)** *TFRC*, **(C)** *TOPBP1*, **(D)** *NEOT2*, **(F)** *PTDSS1*, **(E)** *ITGA6*, **(G)** *MGLL*, **(H)** *TP53I3* and **(I)** *TRIP10*.

Table 1: Fold changes and correlations between ESCC and nine candidate genes

Gene	Microarray data			Correlations with ESCC
	Up/down-regulated	Su's ESCC	Hu's ESCC	
ECT2	Up	4.346	5.87	Cancer progression and poor prognosis
TFRC	Up	2.544	3.652	Prognostic biomarker
TOPBP1	Up	2.34	2.11	-
NETO2	Up	2.689	2.566	-
PTDSS1	Up	2.083	2.214	-
ITGA6	Up	2.738	3.032	Proliferation of ESCC
MGLL	Down	-9.531	-3.109	-
TP53I3	Down	-6.323	-2.676	-
TRIP10	Down	-4.285	-2.408	-

NETO2, TOPBP1 and *PTDSS1* were correlated with cell proliferation in this network.

All genes had at least 4 connections with factors related to ESCC.

Correlation of CNV and expression of the candidate genes

To investigate the correlations between copy number variation and the expression of candidate genes, TCGA CNV and expression profiles were downloaded. We explored the up/down-regulated genes and amplified/deletion in copy number. By bivariate analysis, 6 up-regulated genes, *ECT2, ITGA6, TFRC, NETO2, TOPBP1* and *PTDSS1* (Figure 9) revealed positive correlation with amplified in copy number. It indicated that up regulated of 6 genes might be resulted from DNA copy number amplification.

DISCUSSION

ESCC presents high prevalence and mortality in China [3]. Distant metastasis and recurrence are two
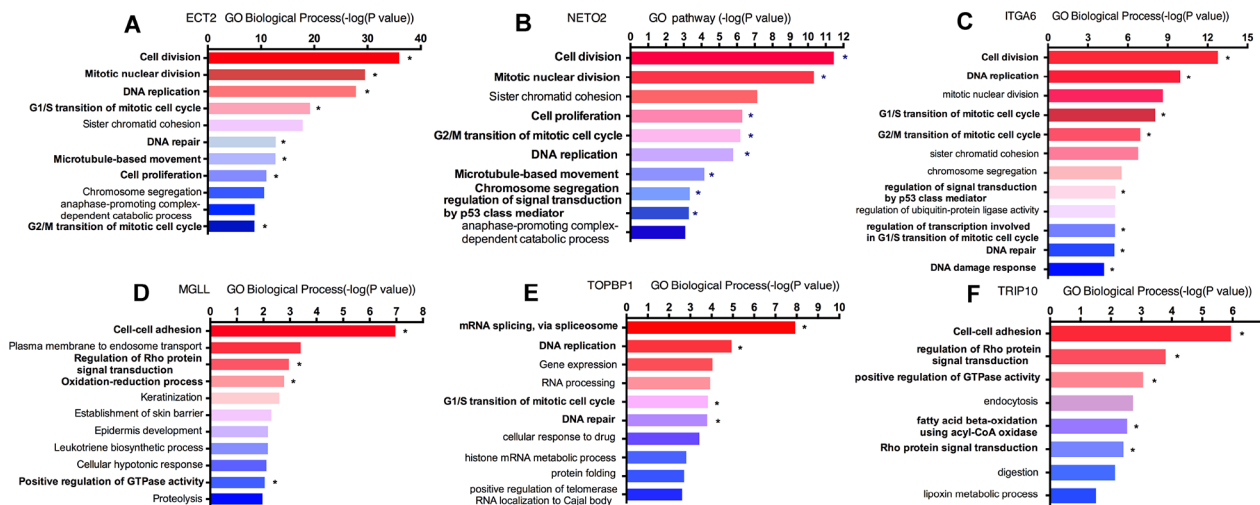
Figure 5: Enrichment analyses of candidate genes in Gene Ontology pathway. "*" represented significant carcinogenic pathway. **(A)** ECT2 GO biological pathway analysis; **(B)** NETO2 GO biological pathway analysis; **(C)** ITGA6 GO biological pathway analysis; **(D)** MGLL GO biological pathway analysis; **(E)** TOPBP1 GO biological pathway analysis; **(F)** TRIP10 GO biological pathway analysis.

major lethal issues [2]. The main objective for this study was to discover more powerful and useful biomarkers or therapeutic targets that involved in the pathogenesis of ESCC by utilizing public datasets.

In the present study, we identified 200 deregulated genes between cancer and normal tissues of ESCC by meta-analysis across two independent microarrays with 140 cases. By validating in another ESCC microarray with 179 paired samples, we screened 139 consistently and stably deregulated genes. In addition, survival analysis identified nine candidate genes related to prognosis of ESCC. Furthermore, by conducting a series of bioinformatics analysis, we revealed that nine candidate genes mainly regulated cell cycle and cell-cell adhesion pathways that were closely related with progression of ESCC. Finally, we found that up-regulated of *ECT2* and other five genes were correlated with amplification of copy number. Together, our data clarified nine candidate genes related to progression and prognosis of ESCC and corresponding potential mechanisms.
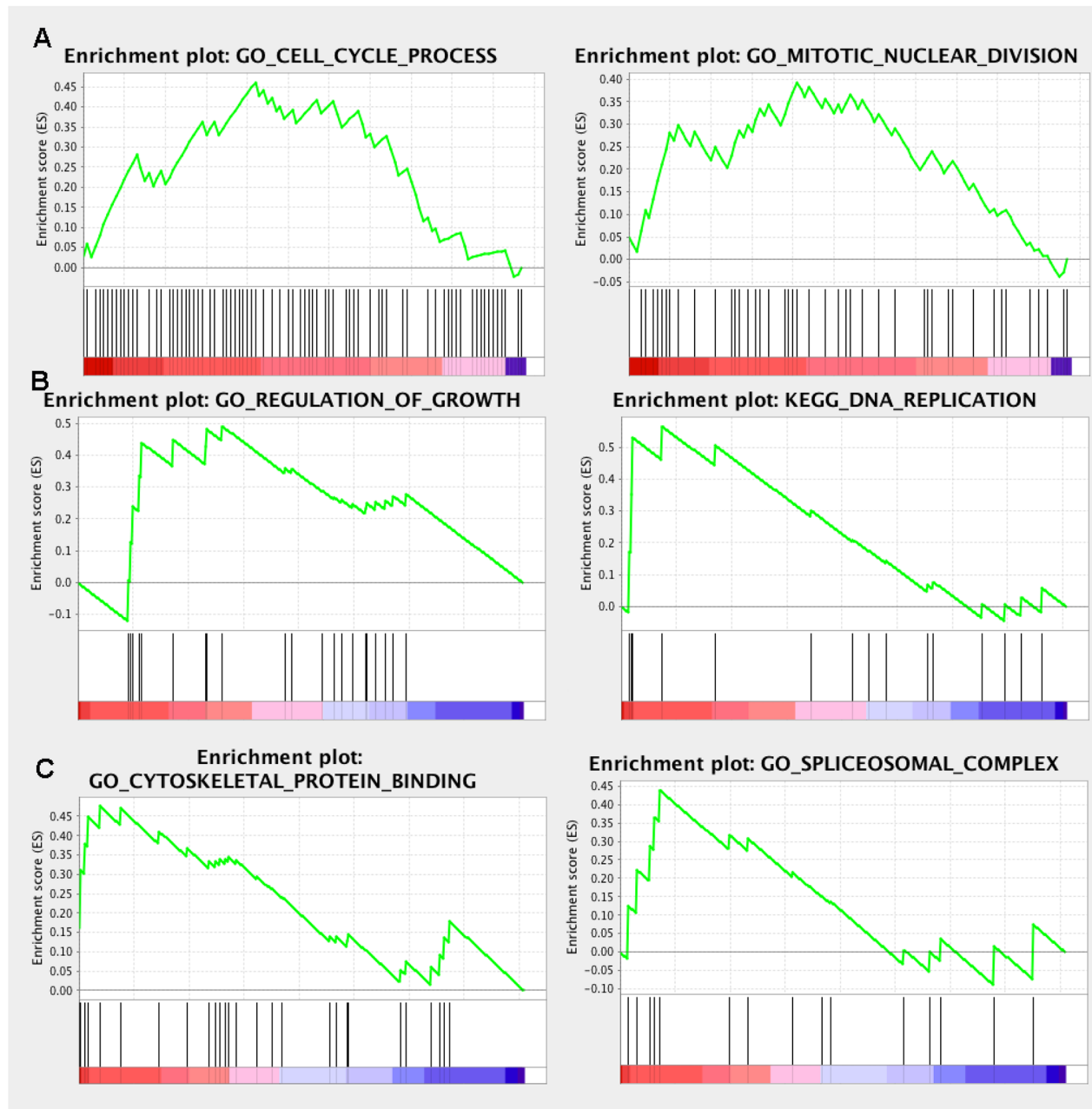
Figure 6: GSEA enrichment analysis of the co-expressed genes. (A) ECT2 GSEA enrichment analysis; **(B)** TOPBP1 GSEA enrichment analysis; **(C)** TFRC GSEA enrichment analysis.

Among nine candidate genes, *ECT2, TFRC* and *ITGA6* have been explored in ESCC [20]. *ECT2* encodes a guanine nucleotide exchange factor and transforming protein related to Rho-specific exchange factors. Hirata reported to identify *ECT2* as a candidate prognostic biomarker and regulate cancer cell growth [21]. *TFRC* is a cell surface receptor that is necessary for cellular iron uptake by the process of receptor-mediated endocytosis. *TFRC* was identified as a prognostic factor in patients with ESCC and correlated with amplification of the chromosome 3q [22]. *ITGA6* belongs to integrin alpha chain family, which activates with other extracellular matrix proteins. Kwon's study found that *ITGA6* promoted the proliferation and invasion of ESCC (24042193). These results were consistent with our findings. To further explore the possible molecular mechanisms, we performed GO term analysis and KEGG pathway analysis. We found that the three genes mainly enriched in cell division, cell cycle and cell-cell adhesion pathways. In addition, protein-protein network analyses indicated that ECT2 might interact with other 160 proteins including CDK5, TP53 and E-cadherin. CDK5 is a serine/threonine kinase that controls the cell cycle and proliferation. TP53 encodes

a tumor suppressor transcription factor that regulates cell cycle arrest, apoptosis. E-cadherin is a important bio-marker in EMT progression. These results indicated that ECT2 might activate cell cycle, apoptosis and cell invasion by these proteins. Previous studies demonstrated that ECT2 could promote the growth and invasion of esophageal cancer, which were consisted with our results.

Besides above-mentioned genes, we also identified novel candidate genes related to progression of ESCC. TOPBP1 has been reported to accelerate tumor development in multi-cancers [23]. However, little is known to its functions in ESCC. Previous studies reported that TOPBP1 played conserved roles in the initiation of DNA replication and activation of DNA damage checkpoint signaling by serving as scaffolding protein [24]. In our study, TOPBP1 was found to be up regulated in cancer tissues and related to survival. Pathway analysis showed up-regulated TOPBP1 could regulate the mRNA splicing, DNA replication and DNA repair pathways. Protein-protein network showed that TOPBP1 could activate or inhibit with other 50 proteins including E2F1, MDC1 and BARD1. These proteins mainly served in DNA replication and
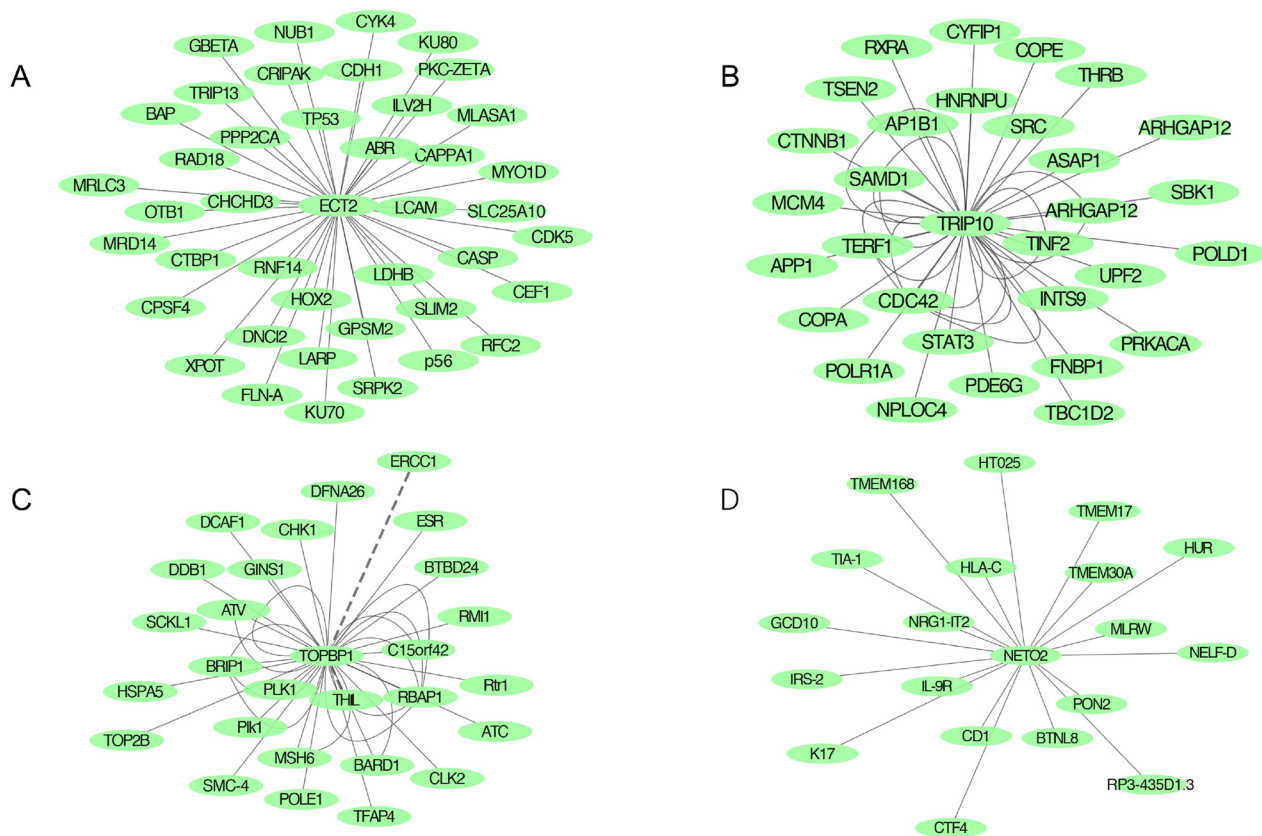

Figure 7: Protein-protein network predicting highly potential interactions with candidate genes based on BioGrid and SRTING databases. (A) PPI network of ECT2; **(B)** PPI network of TRIP10; **(C)** PPI network of TOPBP1; **(D)** PPI network of NETO2.

DNA repair pathways, indicating that TOPBP1 might function as an oncogene through these mechanisms. NETO2 is a single-pass transmembrane protein with two extracellular CUB domains. Increasing evidence showed that NETO2 was an oncogene in colorectal carcinoma and hepatocellular carcinoma [25, 26]. One report found that NETO2 was related to rapidly growing hepatocellular carcinoma [26]. Another study pointed out that up-regulation of NETO2 expression correlated with tumor progression and poor prognosis in colorectal carcinoma [25]. NETO2 was a prognostic biomarker and participated in cell division and cell proliferation pathways by our analysis. This notion is in line with published studies. Among the down-regulated genes, TRIP10, MGLL and TP53I3 were well investigated in previous studies. TRIP10, also named CIP4, was related to lung cancer and chronic lymphocytic leukemia [27, 28]. It was defined as functional interactions during biogenesis of epithelial junctions.

By our analysis, TRIP10 was correlated with cell-cell adhesion, regulation of Rho protein signal transduction and positive regulation of GTPase activity pathways, suggesting potential roles in cell invasion and migration of ESCC. TP53I3 was correlated with melanoma and glioblastoma [29, 30]. Intriguingly, TP53I3 was up regulated in these cancers. It functioned as oncogenes in glioblastoma and papillary thyroid cancer [31]. However, TP53I3 was down regulated in our analysis, indicating a unique role in ESCC. MGLL is a serine hydrolase of the AB hydrolase superfamily. Published studies demonstrated that MGLL was correlated with colorectal cancer and gastric cancer [32, 33]. In current analyses, we identified that MGLL was correlated with survival and participated multi-pathways in ESCC.

The present study facilitated access and interpretation of candidate genes related to progression and prognosis of ESCC. However, some limitations should be mentioned. Our data were analyzed by bioinformatics.
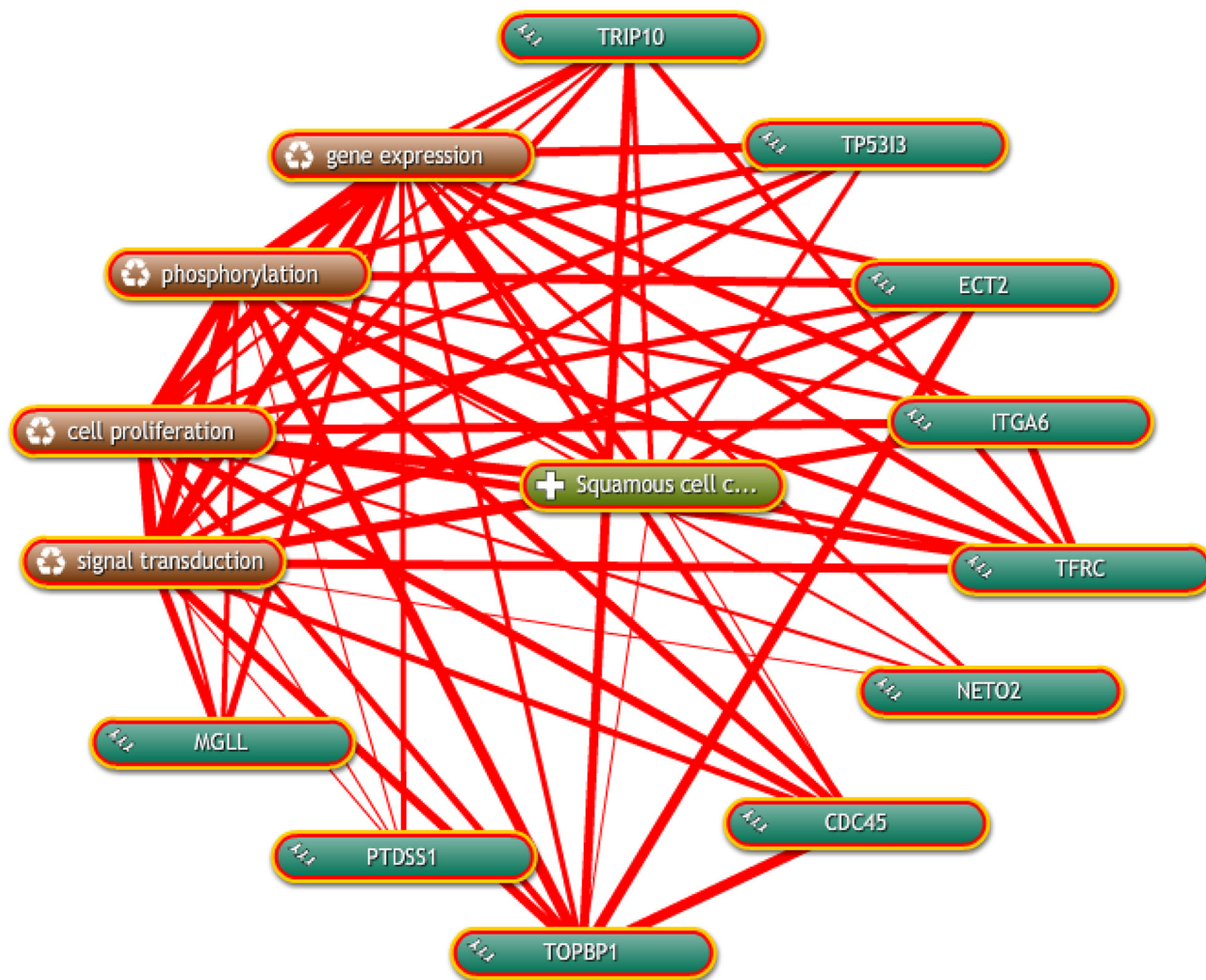

Figure 8: Inter-relationship of candidate genes with ESCC was determined by text mining using Coremine Medical.

Experiment data are needed to support our analyses. In addition, only two studies were included in the analysis. The sample size was too small. Studies with large samples were needed to further validate. Furthermore, several studies built sub-network for cancer prognosis by systematically analyzing multi-dimensional cancer genomics data [34, 35]. However, there was lack of epigenetic and DNA mutations in our analyses. Integrated analyses of genetic, epigenetic and proteomic could be included in further study.

In summary, the current study demonstrated nine genes related to progression and prognosis of ESCC using the public portals and bioinformatics, which provided novel prognostic biomarkers and contributed to a better understanding of cancer molecular mechanisms. This will ultimately accelerate the translation of bench work into bedside.
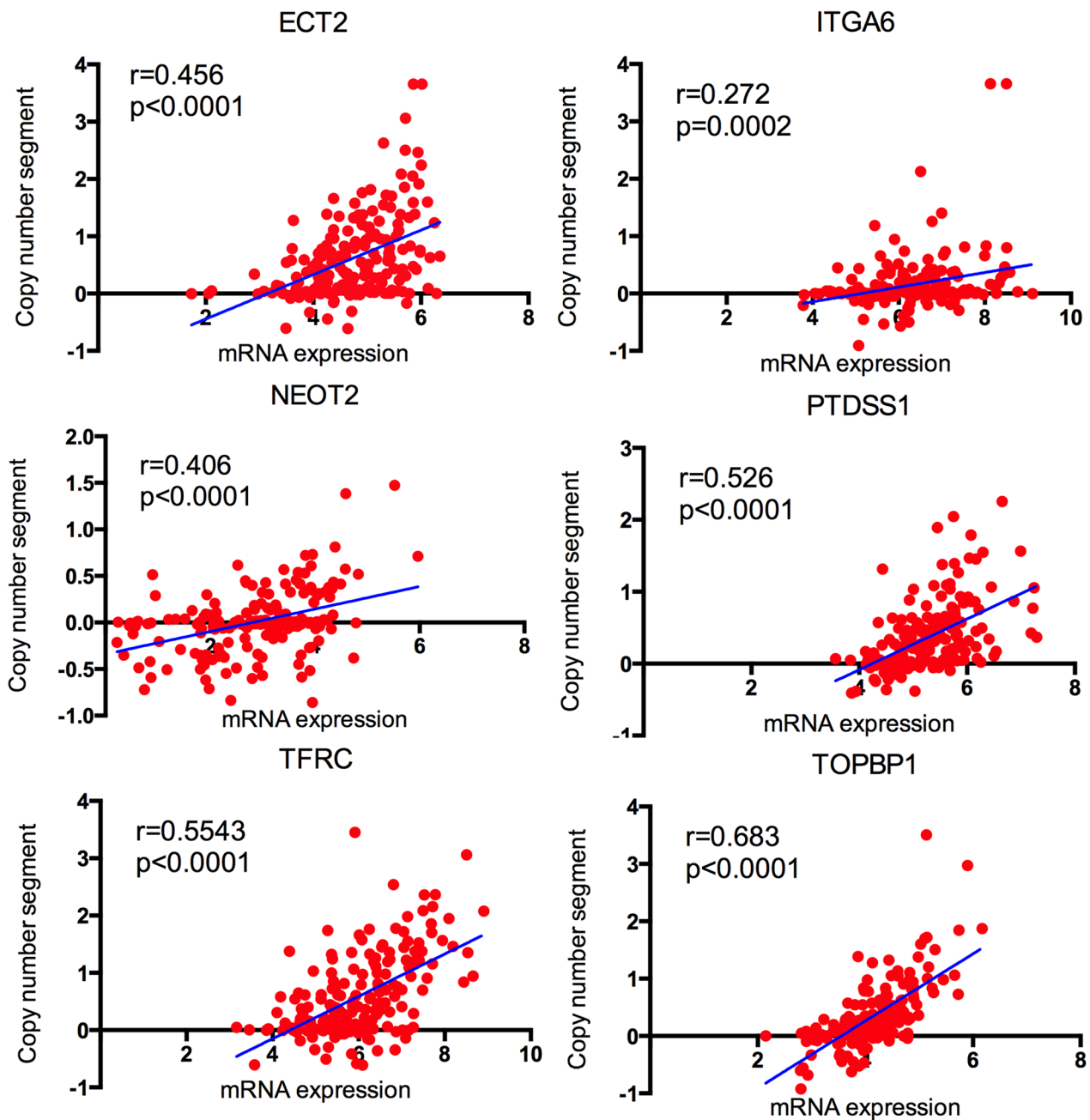
Figure 9: The correlation between copy number segment and the corresponding mRNA expression in TCGA.

MATERIALS AND METHODS

Meta-analysis & data source

We searched eligible ESCC microarray datasets from Oncomine platform by adding filters like mRNA data, clinical specimens and cancer vs normal. To identify candidate genes, meta-analysis of gene expression profiles was performed across ESCC datasets in ONCOMINE 4.5 (https://www.oncomine.org/resource/). The deregulated genes were ranked by median ranks across the selected analyses and displayed corresponding p-values. GSE53625 and corresponding survival information were downloaded from GEO profiles databases for validation (http://www.ncbi.nlm.nih.gov/geoprofiles/)[36]. The limma R package was also utilized to identify the deregulated gene expression between cancer and normal. The level 3 TCGA data TCGA_ESCA_GSNP6noCNV-2015-02-24 (delete germline CNV) and TCGA_ESCA_exp_HiSeq-2015-02-24 were downloaded from Cancer Browser (http://www.genome-cancer.ucsc.edu).

Survival analysis

Kaplan-Meier method was performed to identify genes correlated with survival. Log rank test was used to evaluate the significant differences. The cutoff of expression value was defined as median expression level in each gene expression profile. Survival curves were constructed using GraphPad Prism v6.0 (GraphPad Software, San Diego, CA, USA). Hazard ratio and 95% confidence interval were also calculated.

Bioinformatics analysis

Co-expression analysis was performed based on the functions of Oncomine. By selected a co-expression analysis in the datasets tab, the system displayed a co-expression heat map. Typed the target gene into search bar, co-expression genes were displayed by their correlations to target gene. A correlation value closer to 1.0 indicates genes that are more highly correlated with the selected gene of interest. The list of co-expression genes were submitted to a functional annotation tool named DAVID Bioinformatics Resources 6.8 for KEGG pathway and GO biological process enrichment analysis (https://david.ncifcrf.gov). Protein-protein interaction was conducted in Cytoscape public databases and GeneMANIA plug-in (http://www.genemania.org/). Text mining was used to display potential roles in ESCC by the Coremine Medical online database (http://www.coremine.com/ medical/). Gene Set Enrichment Analysis (GSEA) was performed to determine whether a priori defined set of genes shows statistically significant, consistent differences between two biological states [37].

Expression and CNV analysis

There were 184 esophageal cancer samples in DNA copy number variant (CNV) of TCGA and 198 esophageal cancer samples in expression profile of TCGA. Among them, a total of 183 samples were matched between expression and CNV. The correlation between the CNVs and expression of genes were evaluated by bivariate correlations. Pearson correlation was used. Significant differences were defined as $p<0.05$.

Abbreviations

ESCC: esophageal squamous cell carcinoma
TCGA: The Cancer Genome Atlas
GEO: Gene Expression Omnibus
CNV: copy number variant
GO: Gene Ontology
KEGG: Kyoto Encyclopedia of Genes and Genomes
NCCR: National Central Cancer Registry of China

Author contributions

MQX and XWJ performed most analysis. MQX and DGC wrote the manuscript. JF and XL revised manuscript. MQX, WAP and DGC designed the study and helped draft manuscript. All authors reviewed the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015; 65:87-108. https://doi.org/10.3322/caac.21262.

2. Rustgi AK, El-Serag HB. Esophageal carcinoma. N Engl J Med. 2014; 371:2499-509. https://doi.org/10.1056/NEJMra1314530.

3. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. CA Cancer J Clin. 2016; 66:115-32. https://doi.org/10.3322/caac.21338.

4. Qixing M, Dong G, Wenjie X, Rong Y, Feng J, Lin X, Mantang Q, Qiang C. Predictive value of Ercc1 and Xpd polymorphisms for clinical outcomes of patients receiving neoadjuvant therapy: a prisma-compliant meta-analysis. Medicine (Baltimore). 2015; 94:e1593. https://doi.org/10.1097/MD.0000000000001593.

5. Pennathur A, Gibson MK, Jobe BA, Luketich JD. Oesophageal carcinoma. Lancet. 2013; 381:400-12. https://doi.org/10.1016/S0140-6736(12)60643-6.

6. Yin F, Shu L, Liu X, Li T, Peng T, Nan Y, Li S, Zeng X, Qiu X. Microarray-based identification of genes associated with cancer progression and prognosis in hepatocellular carcinoma. J Exp Clin Cancer Res. 2016; 35:127. https://doi.org/10.1186/s13046-016-0403-2.

7. Di Mauro C, Rosa R, D'Amato V, Ciciola P, Servetto A, Marciano R, Orsini RC, Formisano L, De Falco S, Cicatiello V, Di Bonito M, Cantile M, Collina F, et al. Hedgehog signalling pathway orchestrates angiogenesis in triple-negative breast cancers. Br J Cancer. 2017; 116:1425-35. https://doi.org/10.1038/bjc.2017.116.

8. Dai F, Mei L, Meng S, Ma Z, Guo W, Zhou J, Zhang J. The global expression profiling in esophageal squamous cell carcinoma. Genomics. 2017; 109:241-250. https://doi.org/10.1016/j.ygeno.2017.04.005.

9. Luo J, Wang W, Tang Y, Zhou D, Gao Y, Zhang Q, Zhou X, Zhu H, Xing L, Yu J. mRNA and methylation profiling of radioresistant esophageal cancer cells: the involvement of Sall2 in acquired aggressive phenotypes. J Cancer. 2017; 8:646-56. https://doi.org/10.7150/jca.15652.

10. Zhang F, Ren C, Zhao H, Yang L, Su F, Zhou MM, Han J, Sobie EA, Walsh MJ. Identification of novel prognostic indicators for triple-negative breast cancer patients through integrative analysis of cancer genomics data and protein interactome data. Oncotarget. 2016; 7:71620-34. https://doi.org/10.18632/oncotarget.12287.

11. Qixing M, Juqing X, Yajing W, Gaochao D, Wenjie X, Run S, Anpeng W, Lin X, Feng J, Jun W. The expression levels of CYP3A4 and CYP3A5 serve as potential prognostic biomarkers in lung adenocarcinoma. Tumour Biol. 2017; 39. https://doi.org/10.1177/1010428317698340.

12. Dong G, Mao Q, Yu D, Zhang Y, Qiu M, Dong G, Chen Q, Xia W, Wang J, Xu L, Jiang F. Integrative analysis of copy number and transcriptional expression profiles in esophageal cancer to identify a novel driver gene for therapy. Sci Rep. 2017; 7:42060. https://doi.org/10.1038/srep42060.

13. Mesci A, Huang X, Taeb S, Jahangiri S, Kim Y, Fokas E, Bruce J, Leong HS, Liu SK. Targeting of CCBE1 by miR-330-3p in human breast cancer promotes metastasis. Br J Cancer. 2017; 116:1350-7. https://doi.org/10.1038/bjc.2017.105.

14. Xu HX, Wu KJ, Tian YJ, Liu Q, Han N, He XL, Yuan X, Wu GS, Wu KM. Expression profile of SIX family members correlates with clinic-pathological features and prognosis of breast cancer: a systematic review and meta-analysis. Medicine (Baltimore). 2016; 95:e4085. https://doi.org/10.1097/MD.0000000000004085.

15. Tsai CY, Wang CS, Tsai MM, Chi HC, Cheng WL, Tseng YH, Chen CY, Lin CD, Wu JI, Wang LH, Lin KH. Interleukin-32 increases human gastric cancer cell invasion associated with tumor progression and metastasis. Clin Cancer Res. 2014; 20:2276-88. https://doi.org/10.1158/1078-0432.CCR-13-1221.

16. Clermont PL, Sun L, Crea F, Thu KL, Zhang A, Parolia A, Lam WL, Helgason CD. Genotranscriptomic meta-analysis of the Polycomb gene CBX2 in human cancers: initial evidence of an oncogenic role. Br J Cancer. 2014; 111:1663-72. https://doi.org/10.1038/bjc.2014.474.

17. Barfeld SJ, East P, Zuber V, Mills IG. Meta-analysis of prostate cancer gene expression data identifies a novel discriminatory signature enriched for glycosylating enzymes. BMC Med Genomics. 2014; 7:513. https://doi.org/10.1186/s12920-014-0074-9.

18. Hu N, Clifford RJ, Yang HH, Wang C, Goldstein AM, Ding T, Taylor PR, Lee MP. Genome wide analysis of DNA copy number neutral loss of heterozygosity (CNNLOH) and its relation to gene expression in esophageal squamous cell carcinoma. BMC Genomics. 2010; 11:576. https://doi.org/10.1186/1471-2164-11-576.

19. Su H, Hu N, Yang HH, Wang C, Takikita M, Wang QH, Giffen C, Clifford R, Hewitt SM, Shou JZ, Goldstein AM, Lee MP, Taylor PR. Global gene expression profiling and validation in esophageal squamous cell carcinoma and its association with clinical phenotypes. Clin Cancer Res. 2011; 17:2955-66. https://doi.org/10.1158/1078-0432.CCR-10-2724.

20. Kwon J, Lee TS, Lee HW, Kang MC, Yoon HJ, Kim JH, Park JH. Integrin alpha 6: a novel therapeutic target in esophageal squamous cell carcinoma. Int J Oncol. 2013; 43:1523-30. https://doi.org/10.3892/ijo.2013.2097.

21. Hirata D, Yamabuki T, Miki D, Ito T, Tsuchiya E, Fujita M, Hosokawa MF, Chayama K, Nakamura Y, Daigo Y. Involvement of epithelial cell transforming sequence-2 oncoantigen in lung and esophageal cancer progression. Clin Cancer Res. 2009; 15:256-66. https://doi.org/10.1158/1078-0432.CCR-08-1672.

22. Wada S, Noguchi T, Takeno S, Kawahara K. PIK3CA and TFRC located in 3q are new prognostic factors in esophageal squamous cell carcinoma. Ann Surg Oncol. 2006; 13:961-6. https://doi.org/10.1245/ASO.2006.08.006.

23. Liu Y, Cussiol JR, Dibitetto D, Sims JR, Twayana S, Weiss RS, Freire R, Marini F, Pellicioli A, Smolka MB. TOPBP1Dpb11 plays a conserved role in homologous recombination DNA repair through the coordinated recruitment of 53BP1Rad9. J Cell Biol. 2017; 216:623-39. https://doi.org/10.1083/jcb.201607031.

24. Gallina I, Christiansen SK, Pedersen RT, Lisby M, Oestergaard VH. TopBP1-mediated DNA processing during mitosis. Cell Cycle. 2016; 15:176-83. https://doi.org/10.1080/15384101.2015.1128595.

25. Hu L, Chen HY, Cai J, Yang GZ, Feng D, Zhai YX, Gong H, Qi CY, Zhang Y, Fu H, Cai QP, Gao CF. Upregulation of NETO2 expression correlates with tumor progression and poor prognosis in colorectal carcinoma. BMC Cancer. 2015; 15:1006. https://doi.org/10.1186/s12885-015-2018-y.

26. Villa E, Critelli R, Lei B, Marzocchi G, Camma C, Giannelli G, Pontisso P, Cabibbo G, Enea M, Colopi S, Caporali C, Pollicino T, Milosa F, et al. Neoangiogenesis-related genes are hallmarks of fast-growing hepatocellular carcinomas and worst survival. Gut. 2016; 65:861-9. https://doi.org/10.1136/gutjnl-2014-308483.

27. Truesdell P, Ahn J, Chander H, Meens J, Watt K, Yang X, Craig AW. CIP4 promotes lung adenocarcinoma metastasis and is associated with poor prognosis. Oncogene. 2015; 34:3527-35. https://doi.org/10.1038/onc.2014.280.

28. Malet-Engra G, Viaud J, Ysebaert L, Farce M, Lafouresse F, Laurent G, Gaits-Iacovoni F, Scita G, Dupre L. CIP4 controls CCL19-driven cell steering and chemotaxis in chronic lymphocytic leukemia. Cancer Res. 2013; 73:3412-24. https://doi.org/10.1158/0008-5472.CAN-12-3564.

29. Herraiz C, Calvo F, Pandya P, Cantelli G, Rodriguez-Hernandez I, Orgaz JL, Kang N, Chu T, Sahai E, Sanz-Moreno V. Reactivation of p53 by a cytoskeletal sensor to control the balance between DNA damage and tumor dissemination. J Natl Cancer Inst. 2015; 108. https://doi.org/10.1093/jnci/djv289.

30. Quan J, Li Y, Jin M, Chen D, Yin X, Jin M. Suppression of p53-inducible gene 3 is significant for glioblastoma progression and predicts poor patient prognosis. Tumour Biol. 2017; 39. https://doi.org/10.1177/1010428317694572.

31. Xu J, Cai J, Jin X, Yang J, Shen Q, Ding X, Liang Y. PIG3 plays an oncogenic role in papillary thyroid cancer by activating the PI3K/AKT/PTEN pathway. Oncol Rep. 2015; 34:1424-30. https://doi.org/10.3892/or.2015.4096.

32. Shen X, Yue M, Meng F, Zhu J, Zhu X, Jiang Y. Microarray analysis of differentially-expressed genes and linker genes associated with the molecular mechanism of colorectal cancer. Oncol Lett. 2016; 12:3250-8.

33. Li CF, Chuang IC, Liu TT, Chen KC, Chen YY, Fang FM, Li SH, Chen TJ, Yu SC, Lan J, Huang HY. Transcriptomic reappraisal identifies MGLL overexpression as an unfavorable prognosticator in primary gastrointestinal stromal tumors. Oncotarget. 2016; 7:49986-97. https://doi.org/10.18632/oncotarget.10304.

34. Zhang F, Ren C, Lau KK, Zheng Z, Lu G, Yi Z, Zhao Y, Su F, Zhang S, Zhang B, Sobie EA, Zhang W, Walsh MJ. A network medicine approach to build a comprehensive atlas for the prognosis of human cancer. Brief Bioinform. 2016; 17:1044-59. https://doi.org/10.1093/bib/bbw076.

35. Buzolin AL, Moreira CM, Sacramento PR, Oku AY, Fornari A, Antonio DS, Quaio C, Baratela WR, Mitne-Neto M. Development and validation of a variant detection workflow for BRCA1 and BRCA2 genes and its clinical application based on the Ion Torrent technology. Hum Genomics. 2017; 11:14. https://doi.org/10.1186/s40246-017-0110-x.

36. Li J, Chen Z, Tian L, Zhou C, He MY, Gao Y, Wang S, Zhou F, Shi S, Feng X, Sun N, Liu Z, Skogerboe G, et al. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. Gut. 2014; 63:1700-10. https://doi.org/10.1136/gutjnl-2013-305806.

37. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545-50. https://doi.org/10.1073/pnas.0506580102.