Original Article

Identification of the key transcription factors in esophageal squamous cell carcinoma

Yuefeng Zhang, Yanzhao Xu, Zhenhua Li, Yonggang Zhu, Shiwang Wen, Mingbo Wang, Huilai Lv, Fan Zhang, Ziqiang Tian

Department of Thoracic Surgery, The Fourth Hospital of Hebei Medical University, Shijiazhuang 050035, China

Contributions: (I) Conception and design: Y Zhang, Z Tian; (II) Administrative support: Y Zhang, Z Tian; (III) Provision of study materials or patients: Y Zhang, S Wen, M Wang, H Lv, F Zhang; (IV) Collection and assembly of data: Y Zhang, Y Xu, Z Li, Y Zhu; (V) Data analysis and interpretation: Y Zhang, Y Xu, Z Li, Y Zhu, S Wen, M Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Ziqiang Tian. Department of Thoracic Surgery, The Fourth Hospital of Hebei Medical University, No. 169 Tianshan Street, Economic and Technological Development District, Shijiazhuang 050035, China. Email: tianziq@sina.com or tizq12@vip.163.com.

Background: Esophageal cancer (EC) is a common human malignancy worldwide. Esophageal squamous cell carcinoma (ESCC) is the predominant subtype in China. The tumorigenesis mechanism in ESCC is unclear. The aim of this study was to identify key transcription factors (TFs) in ESCC and elucidate the mechanism of it.

Methods: A total of ten published microarray datasets of ESCC was downloaded from the Gene Expression Omnibus (GEO). Then, bioinformatics analyses including differentially expressed genes (DEGs) analysis, gene ontology (GO) annotation, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment, TFs-genes regulatory network construction was performed. Quantitative real-time polymerase chain reactions (qRT-PCR) were used to detect the expression levels of TFs and DEGs in ESCC. The association between stage and TFs and the association between survival and TFs were evaluated based on The Cancer Genome Atlas (TCGA), respectively.

Results: A total of 1,248 dysregulated genes were selected as DEGs in ESCC. A total of 26 TFs and corresponding target-genes were identified. The ESCC-specific transcriptional regulatory network was constructed. The network was consisted of 882 edges and 631 nodes. *BRCA1*, *SOX10*, *ARID3A*, *ZNF354C* and *NFIC* had the highest connectivity with DEGs, and regulated 92, 89, 82, 79 and 78 DEGs in the network, respectively. All these 1,248 DEGs were significantly enriched in cell cycle, DNA replication and oocyte meiosis pathways. The qRT-PCR results were consistent with our microarray analysis. High expression of *SREBF1* and *TFAP2A* were significantly correlated with the longer overall survival time of patients with ESCC.

Conclusions: *BRCA1*, *SOX10*, *ARID3A*, *ZNF354C* and *NFIC* might be the key TFs in carcinogenesis and development of ESCC by regulating their corresponding target-genes involved in cell cycle, DNA replication and oocyte meiosis pathways. *SREBF1* and *TFAP2A* may be two potential prognostic biomarkers of ESCC.

Keywords: Transcription factors (TFs); esophageal squamous cell carcinoma (ESCC); regulatory network; expression profiling

Submitted Mar 27, 2017. Accepted for publication Dec 06, 2017.
doi: 10.21037/jtd.2017.12.27
View this article at: http://dx.doi.org/10.21037/jtd.2017.12.27

Introduction

Esophageal cancer (EC) is a common human malignancy with poor survival. The incidence rate of EC was seventh and sixth in male and female cancer worldwide, respectively. Moreover, EC was usually diagnosed at an advanced stage. It is known that an estimated 455,800 new EC cases and 400,200 deaths occurred in 2012 worldwide (1).

EC is classified as two major subtypes including esophageal adenocarcinoma (EAC) and esophageal squamous cell carcinoma (ESCC) according to histopathology appearance (2). In recent years, the incidence rates of EAC had a markedly decrease in western countries such as the United States, Australia, France and England. On the contrary, the incidence rates of ESCC have been increasing in Asian countries (1). ESCC subtype accounts for more than 90% EC patients in Asian countries such as China and Japan.

Currently, the major etiological factors of ESCC are known as environmental factors, human papillomavirus (HPV) infection and genetic susceptibility. The environmental factors include alcohol consumption and tobacco exposure. The case-control studies show that exposure to secondhand smoking and heavy alcohol intake are risk factors for ESCC (3,4). HPV is one of risk factors of ESCC and HPV status is an independent prognostic factor for survival among patients with ESCC (5). Beyond that, *CYP1A1* Val/Val and *CYP2E1* c1/c1 genotypes are significantly associated with ESCC risk (6). MiR-34b/c rs4938723 T>C, miR-423 rs6505162 C>A, miR-196a2 rs11614913 and miR-499 rs3746444 polymorphisms are associated with susceptibility to ESCC in Chinese population (7,8).

Transcription factors (TFs) commonly regulating gene expression through binding to specific DNA sequences and involves in tumor cell processes including cell proliferation, apoptosis and migration. It is reported that over-expression of zinc finger E-box binding to homeobox factor 1 promotes aggressive ESCC progression and contributes to the unfavorable prognosis in ESCC (9). *SOX2* promotes the EMT of ESCC cells by modulating slug expression through the activation of *STAT3/HIF-α* signaling, and it suggests that *SOX2* promotes the metastasis process of ESCC (10).

In this study, bioinformatics analyses were applied to integrate mRNA expression profiling of ESCC, which were available in the Gene Expression Omnibus (GEO) database. Identification of differentially expressed genes (DEGs) in ESCC and construction of TFs-target genes regulatory network were subjected to screen the key TFs in ESCC. This study aimed to provide valuable information for further pathogenesis mechanism elucidation, and identify cancer biomarkers of diagnosis, prognosis and therapeutic targets in ESCC.

Methods

Microarray data collection

The microarray expression profiles of ESCC were obtained from GEO (http://www.ncbi.nlm.nih.gov/geo/) (11). The inclusion criteria of datasets were as follows: (I) expression profiling of ESCC tumor tissues/cell lines and normal controls were available in the datasets; (II) patients without preoperative treatment before esophagectomy and cell lines without drug stimulation. A total of ten mRNA expression datasets of ESCC were incorporated into our study.

Data preprocessing and identification of DEGs

The raw expression datasets were downloaded and pre-processed to perform background subtraction, quantile normalization and $\log_2$ transformation before using moderated *t*-tests within Bioconductor package Limma (Linear Models for Microarray Data). Subsequently, the Linear Models Limma (12) package for microarray data was used to screen DEGs between ESCC and normal controls. The P values was adjusted to false discovery rate (FDR) through Benjamini and Hochberg method (13). The threshold for the DEGs screening was set as FDR $<10^{-5}$. Hierarchical cluster analysis of group samples based on expression levels of mRNA were visualized through "pheatmap" package in R language.

Gene ontology (GO) and pathway enrichment analysis of DEGs

The GO annotation analysis has commonly used for functional studies of large-scale transcriptomics data. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database contains biochemistry pathways. Online software of Gorilla and GeneCoDis3 were utilized for GO annotation and KEGG pathway enrichment of DEGs, respectively (14,15). The threshold of GO and KEGG enrichment of DEGs were P<0.001 and FDR <0.05, respectively.

Table 1 Basic information for mRNA expression profiling of ESCC

GEO ID	Samples (case control)	Platform	Year	Author
GSE70409	17:17	GPL13287 halanx Human OneArray	2015	Wu IC
GSE61587	3:1	GPL16699 Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray 039381 (Feature Number version)	2014	Osamu Ishibashi
GSE63941	22:4	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	2014	Niki T
GSE26886	9:19	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	2013	Wang Q
GSE45168	5:5	GPL13497 Agilent-026652 Whole Human Genome Microarray 4x44K v2 (probe Name version)	2013	Wang L
GSE47404	71:0	GPL6480 Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (probe Name version)	2013	Sawadagenta
GSE29001	21:24	GPL571 [HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array	2011	Yan W
GSE20347	17:17	GPL571 [HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array	2011	Clifford RJ
GSE23400	53:53	GPL96 [HG-U133A] Affymetrix Human Genome U133A Array; GPL97 [HG-U133B] Affymetrix Human Genome U133B Array	2010	Yang HH
GSE9982	20:2	GPL1928 CodeLink Human 20K ver4.1	2006	Shimokuni T

ESCC, esophageal squamous cell carcinoma; GEO, Gene Expression Omnibus.

Construction of TFs-target genes regulatory network

TRANSFAC (http://www.gene-regulation.com/pub/databases.html) provides data of eukaryotic TFs, experimentally-proven binding sites of TFs and consensus binding sequences (16). Data of position weight matrix (PMW), binding sites motif of TFs were downloaded from TRANSFAC and used for identification of TFs and target-genes. The identified TFs and target-genes were used for regulatory network construction. The regulatory network was visualized by Cytoscape (http://cytoscape.org/) (17). Nodes represented TFs or target-genes and solid line represented correlation between TF and target-gene.

qRT-PCR verification of DEGs and TFs in ESCC

Five pairs of ESCC tumor and adjacent non-tumor tissues were obtained from five Chinese patients underwent surgery in Department of Thoracic Surgery, The Fourth Hospital of Hebei Medical University. The detail information of subjects was shown in *Table S1*. Our study was approved by the Ethics Committee of The Fourth Hospital of Hebei Medical University (No. 2015MEC073). All of patients signed the consent form and our study compiled with the Declaration of Helsinki.

Total RNA of fresh five paired ESCC tumor and adjacent non-tumor specimens were extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). The SuperScript III Reverse Transcription Kit (Invitrogen, Carlsbad, CA, USA) was used to synthesize the cDNA. qRT-PCR reactions were performed using Power SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA) on the Applied Biosystems 7500 (Applied Biosystems, Foster City, CA, USA). β-actin was used as internal control for mRNA detected. The relative expression of DEGs was calculated through the comparative Ct methods (18). The PCR primers were used as *Table S2* shown.

Stage and survival analysis of TFs in ESCC

To better research the clinical relevance of these TFs in ESCC. We downloaded an illumina HiSeq RNA-Seq data from The Cancer Genome Atlas (TCGA) which consisted of 94 ESCC samples. This TCGA data were used to evaluate the correlation between survival and the expression of TFs in ESCC. In addition, the correlation between stage and the expression of TFs in ESCC were evaluated based on this TCGA data as well.

Results

DEGs in ESCC

We collected ten mRNA expressions profiling including 238 ESCC cases and 142 normal controls (*Table 1*) from GEO database. With a cut-off value of FDR $<10^{-5}$, a total

Table 2 The top 10 up- and down-regulated genes in ESCC

ID	Symbol	Log_2 FC	P value
Up-regulated genes (top 10)			
1469	*CST1*	5.96	6.02^{-10}
3226	*HOXC10*	5.39	2.55^{-14}
10643	*IGF2BP3*	5.02	2.22^{-11}
4314	*MMP3*	5.02	3.71^{-9}
4321	*MMP12*	4.91	1.10^{-12}
4312	*MMP1*	4.74	5.35^{-10}
3206	*HOXA10*	4.57	1.43^{-15}
6364	*CCL20*	4.54	1.38^{-8}
3227	*HOXC11*	4.54	1.36^{-13}
3209	*HOXA13*	4.54	1.06^{-6}
Down-regulated genes (top 10)			
643847	*PGA4*	−11.2	1.98^{-15}
643834	*PGA3*	−10.6	1.43^{-15}
8513	*LIPF*	−10.2	2.92^{-11}
5222	*PGA5*	−9.69	1.27^{-19}
495	*ATP4A*	−9.11	1.21^{-19}
56287	*GKN1*	−9.03	8.79^{-15}
496	*ATP4B*	−8.28	7.78^{-19}
2694	*GIF*	−7.85	1.81^{-12}
27159	*CHIA*	−7.71	3.47^{-25}
200504	*GKN2*	−7.44	5.06^{-14}

FC, fold change; ESCC, esophageal squamous cell carcinoma.

of 1,248 DEGs including 663 up- and 585 down-regulated genes were identified in ESCC compared to normal controls. The top 10 up- and down-regulated genes were listed in *Table 2*. The most up- and down-regulated gene was *CST1* and *PGA4* in ESCC, respectively. The full list of DEGs was shown in http://jtd.amegroups.com/public/addition/jtd/supp-jtd.2017.12.27.pdf. The top 200 up- and down-regulated DEGs in ESCC compared with normal controls were subjected to heatmap analyses. As *Figure 1* shown, the difference of expression pattern of DEGs between ESCC and normal controls was notable.

Due to the restriction of GEO, only little clinical information of samples of these ten datasets can be obtained by GEO. The information of gender and age of samples can be obtained from only four datasets (GSE70409, GSE61587, GSE45168 and GSE47404), which accounts for 31.5% (n=119) of all samples. The information of TNM staging or histological differentiation of samples can be obtained from only two datasets (GSE45168 and GSE47404), which accounts for 21.4% (n=81) of all samples. Hence, it is difficult to analyze the confounding effects of age, gender, TNM staging, and histological differentiation on expression of TFs and DEGs in ESCC. Considering the race of ESCC samples in these ten datasets, 57.2% (n=135) ESCC samples were obtained from Asian (China and Japan), 42.4% ESCC samples were obtained from Non-Asian (USA and Germany). Considering the race of control samples in these ten datasets, 19.7% (n=28) samples were obtained from Asian (China and Japan), 79.6% (n=113) samples were obtained from Non-Asian (USA and Germany).

GO annotation of DEGs in ESCC

To understand the biological roles of DEGs in ESSC, GO analysis were performed. The threshold of GO terms was P value <0.001. The system process (GO:0003008) was the most significant enrichment of biological process; substrate-specific transmembrane transporter activity (GO:0022891) was the highest enrichment term of molecular function; and membrane part (GO:0044425) was the highest enrichment term of cellular component, as *Table 3* shown.

KEGG pathway enrichment of DEGs in ESCC

KEGG enrichment analysis was performed to obtain the signaling pathways of DEGs in ESCC. The threshold was FDR <0.05. The significantly enriched signaling pathways consisted of cell cycle, DNA replication and oocyte meiosis, as *Table 4* shown.

Construction of TFs-genes regulatory network

TF-gene pairs were selected based on the TRANSFAC database and a transcriptional regulatory network was constructed. Total of 26 TFs including 15 up- and 11 down-regulated genes (*Table S4*), and corresponding target-genes were identified in the DEGs of ESCC. TFs and targets-genes were carried out to construct regulatory network by Cytoscape software, as *Figure 2* shown. The network was consisted of 631 nodes and 882 edges between 21 TFs and 610 target-genes. The TFs had high connectivity with target-genes including *BRCA1* (degree =92), *SOX10*

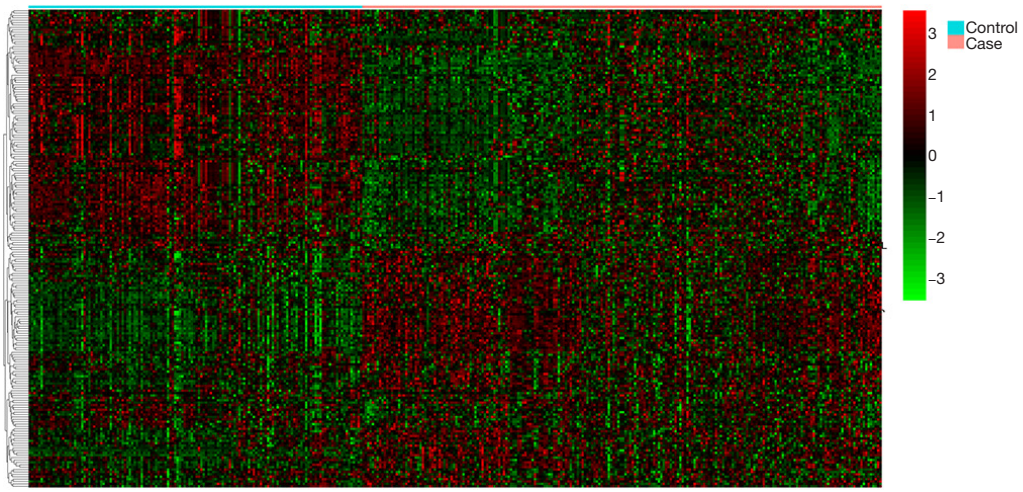

Figure 1 Hierarchical clustering analysis based on the expression profile of the top 200 up-regulated DEGs and top 200 down-regulated DEGs in ESCC and normal controls. The color scale shown at the right illustrated the relative expression levels of DEGs across all samples; red color represented an expression level above mean, blue color represented expression lower than the mean. Case indicated ESCC tissues and control indicated normal controls. DEGs, differentially expressed genes; ESCC, esophageal squamous cell carcinoma.

(degree =89), *NFIC* (degree =82), *ARID3A* (degree =82) and *ZNF354C* (degree =79), as *Figure 2* and *Table S5* shown.

qRT-PCR verification of DEGs and TFs in ESCC

To validate the microarray analysis data, a number of TFs and top 20 up- and down-regulated DEGs were used to perform qRT-PCR verification. The expression levels of DEGs and TFs including *BRCA1*, *SOX10*, *FEV*, *C16orf89*, *HOXA13*, *SERPINB5*, *GRIN2D* and *CCL20* were detected by qRT-PCR in five paired ESCC tumor and adjacent non-tumor tissues derived from Chinese patients with ESCC. As shown in *Figure 3A*, *BRCA1* was significant up-regulated in ESCC. *SOX10*, had the up-regulation tendency in ESCC (*Figure 3B*). *FEV* had the down-regulation tendency in ESCC (*Figure 3C*). *C16orf89* was significantly down-regulated in ESCC (*Figure 3D*). *HOXA13*, *SERPINB5*, *CRIN2D* and *CCL20* had the up-regulation tendency in ESCC (*Figure 3E,F,G,H*), respectively. Although most of the qRT-PCR results were not statistically significant, the regulation trends of candidate DEGs and TFs in our qRT-PCR were generally consistent with our integrated analysis.

Stage and survival analysis of selected TFs in ESCC

Due to lack of clinical information in GEO, we downloaded an illumina HiSeq RNA-Seq data that consisted of 94 ESCC samples from TCGA. Based on this illumina HiSeq RNA-Seq data, the correlation between survival and the expression levels of the 26 TFs were evaluated. Among which, high expression of *SREBF1* (P=0.0427) and *TFAP2A* (P=0.0422) were significantly correlated with the longer overall survival time of patients with ESCC (*Figure 4*). The correlation between stage and the expression levels of eight TFs including *BRCA1*, *SOX10*, *FEV*, *ARID3A*, *ZNF354C*, *NFIC*, *TFAP2A* and *SREBF1* were evaluated. The expression of these eight TFs in stage I–IV of patients with ESCC in TCGA were displayed in *Figure 5* which shown the association between the expression of TFs and the severity of ESCC.

Discussion

Due to the differences of samples and platforms in various microarray studies, integrated analysis of various microarray datasets could obtain more accurate disease-related regulators with a larger sample size than an individual microarray. Our previous study has identified key DEGs in ESCC by integrated analysis of five microarray datasets from GEO (19). With GEO database updated, we identified DEGs in ESCC by integrated analysis of ten microarray datasets with larger sample size in this study. Moreover, the key TFs in ESCC were further identified based these ESCC-related DEGs.

Table 3 GO enrichment of DEGs in ESCC

GO ID	GO term	Count	P value
Biological process			
GO:0003008	System process	80	4.19^{-12}
GO:0044765	Single-organism transport	124	3.88^{-10}
GO:0043269	Regulation of ion transport	39	4.16^{-9}
GO:0007268	Synaptic transmission	40	4.17^{-9}
GO:0034220	Ion transmembrane transport	51	6.18^{-9}
GO:0006811	Ion transport	66	2.57^{-8}
GO:0055085	Transmembrane transport	64	3.55^{-8}
GO:0050804	Modulation of synaptic transmission	29	6.03^{-8}
GO:0007586	Digestion	9	1.09^{-7}
GO:1902578	Single-organism localization	147	1.45^{-7}
GO:0007186	G-protein coupled receptor signaling pathway	47	1.63^{-7}
GO:0098660	Inorganic ion transmembrane transport	34	1.69^{-7}
GO:0090087	Regulation of peptide transport	25	2.51^{-7}
GO:0006810	Transport	143	3.21^{-7}
GO:0042391	Regulation of membrane potential	29	3.69^{-7}
Molecular function			
GO:0022891	Substrate-specific transmembrane transporter activity	51	5.59^{-8}
GO:0022857	Transmembrane transporter activity	54	6.40^{-8}
GO:0015075	Ion transmembrane transporter activity	45	1.02^{-7}
GO:0046873	Metal ion transmembrane transporter activity	27	2.01^{-7}
GO:0022892	Substrate-specific transporter activity	52	5.38^{-7}
GO:0038023	Signaling receptor activity	62	5.57^{-7}
GO:0022890	Inorganic cation transmembrane transporter activity	28	5.72^{-7}
GO:0004871	Signal transducer activity	70	7.56^{-7}
GO:0008324	Cation transmembrane transporter activity	31	1.35^{-6}
GO:0004930	G-protein coupled receptor activity	30	1.74^{-6}
GO:0004872	Receptor activity	61	3.86^{-6}
GO:0022832	Voltage-gated channel activity	17	5.26^{-6}
GO:0005244	Voltage-gated ion channel activity	17	5.26^{-6}
GO:0060089	Molecular transducer activity	78	5.84^{-6}
GO:0004888	Transmembrane signaling receptor activity	53	7.57^{-6}
Cellular component			
GO:0044425	Membrane part	217	1.31^{-17}
GO:0031224	Intrinsic component of membrane	169	5.18^{-14}
GO:0016021	Integral component of membrane	160	5.25^{-13}
GO:0005886	Plasma membrane	147	5.29^{-12}
GO:0044459	Plasma membrane part	117	2.29^{-10}
GO:0097458	Neuron part	70	9.13^{-10}
GO:0044456	Synapse part	40	2.64^{-7}
GO:0031226	Intrinsic component of plasma membrane	67	8.61^{-7}
GO:0005887	Integral component of plasma membrane	64	1.04^{-6}

DEGs, differentially expressed genes; ESCC, esophageal squamous cell carcinoma; GO, gene ontology.

Table 4 KEGG pathway enrichment of DEGs

KEGG ID	KEGG term	Count	FDR	Genes
hsa04110	Cell cycle	39	1.16^{-24}	*ESPL1*, *MCM4*, *CCNB2*, *ABL1*, *CDC20*, *CCNA2*, *PTTG1*, *E2F3*, *CDC7*, *MCM7*, *CCNE1*, *CHEK1*, *BUB3*, *E2F1*, *CHEK2*, *CDC25C*, *E2F2*, *CCNB1*, *CDC6*, *MAD2L1*, *CDC25B*, *SKP2*, *MCM3*, *DBF4*, *CDK2*, *MCM5*, *MCM6*, *PLK1*, *BUB1B*, *CDC45*, *CDC25A*, *PKMYT1*, *TTK*, *BUB1*, *RBL1*, *HDAC1*, *PCNA*, *MCM2*, *CDK1*
hsa03030	DNA replication	18	1.12^{-15}	*RNASEH2A*, *MCM4*, *MCM7*, *LIG1*, *RPA3*, *POLE2*, *POLD1*, *RFC5*, *RFC3*, *MCM3*, *MCM5*, *MCM6*, *PRIM1*, *FEN1*, *PCNA*, *MCM2*, *DNA2*, *RFC4*
hsa04114	Oocyte meiosis	22	1.61^{-9}	*ESPL1*, *CCNB2*, *CDC20*, *PTTG1*, *FBXO5*, *CPEB1*, *SGOL1*, *CCNE1*, *CDC25C*, *AR*, *CCNB1*, *RPS6KA6*, *MAD2L1*, *CDK2*, *PLK1*, *PKMYT1*, *BUB1*, *ITPR1*, *AURKA*, *FBXW11*, *PRKACB*, *CDK1*
hsa03050	Proteasome	12	3.15^{-7}	*PSMA5*, *PSMA6*, *PSMD11*, *PSME2*, *PSMC4*, *PSMB3*, *PSMB4*, *PSMA3*, *PSMD14*, *PSMA7*, *PSMB2*, *PSMA4*
hsa05200	Pathways in cancer	34	4.33^{-7}	*MSH2*, *MMP9*, *STAT1*, *FGF10*, *ABL1*, *KIT*, *BIRC5*, *E2F3*, *TRAF2*, *CCNE1*, *E2F1*, *ITGA2*, *RAD51*, *TPM3*, *AR*, *E2F2*, *WNT2*, *IL8*, *SKP2*, *FIGF*, *BID*, *FZD2*, *CDK2*, *RAF1*, *MITF*, *CKS1B*, *BRCA2*, *LAMC2*, *JUP*, *BAX*, *ZBTB16*, *RXRG*, *HDAC1*, *MMP1*
hsa04080	Neuroactive ligand-receptor interaction	30	6.06^{-7}	*GRIA3*, *CCKBR*, *LEPR*, *CHRNA5*, *GLP2R*, *GRIA2*, *CHRM2*, *SCTR*, *NMUR1*, *HRH2*, *VIPR2*, *GABRD*, *ADCYAP1R1*, *CTSG*, *GRIN2D*, *GRIA1*, *GRIK3*, *PTGER3*, *HTR1E*, *P2RX2*, *GRIA4*, *NPY2R*, *GABRA1*, *CHRNA1*, *AVPR1B*, *GRIK1*, *CCKAR*, *ADRA1A*, *GALR1*, *P2RY14*
hsa04914	Progesterone-mediated oocyte maturation	16	9.30^{-7}	*CCNB2*, *CCNA2*, *CPEB1*, *CDC25C*, *CCNB1*, *RPS6KA6*, *MAD2L1*, *CDC25B*, *CDK2*, *RAF1*, *PLK1*, *CDC25A*, *PKMYT1*, *BUB1*, *PRKACB*, *CDK1*
hsa03430	Mismatch repair	9	1.04^{-6}	*MSH2*, *LIG1*, *RPA3*, *POLD1*, *EXO1*, *RFC5*, *RFC3*, *PCNA*, *RFC4*
hsa04115	p53 signaling pathway	14	1.27^{-6}	*CCNB2*, *CCNE1*, *CHEK1*, *CHEK2*, *CCNB1*, *BID*, *CDK2*, *GTSE1*, *CCNG1*, *RPRM*, *BAX*, *SERPINB5*, *RRM2*, *CDK1*
hsa03040	Spliceosome	18	2.36^{-6}	*SNRPC*, *SNRPA*, *LSM7*, *PRPF19*, *SNRPA1*, *SNRPB*, *MAGOH*, *THOC4*, *PPIL1*, *EIF4A3*, *SNRPG*, *HNRNPC*, *EFTUD2*, *SNRPE*, *SF3B4*, *U2AF2*, *SNRNP40*, *SNRPD1*
hsa04971	Gastric acid secretion	14	2.47^{-6}	*SLC9A4*, *CCKBR*, *KCNE2*, *SLC26A7*, *HRH2*, *KCNJ16*, *ATP4B*, *KCNK2*, *MYLK3*, *SST*, *ITPR1*, *ATP4A*, *ATP1A2*, *PRKACB*
hsa03440	Homologous recombination	9	2.53^{-6}	*RAD54B*, *RPA3*, *RAD51*, *EME1*, *POLD1*, *XRCC2*, *BRCA2*, *RAD54L*, *BLM*
hsa05322	Systemic lupus erythematosus	15	5.05^{-6}	*HIST1H2BJ*, *H2AFX*, *HIST3H2BB*, *HIST2H2AC*, *CD80*, *C6*, *HIST1H2BH*, *CTSG*, *SNRPB*, *HIST1H2AE*, *C7*, *HIST1H2BO*, *HIST1H3G*, *ELANE*, *SNRPD1*
hsa05219	Bladder cancer	9	1.57^{-4}	*MMP9*, *E2F3*, *E2F1*, *E2F2*, *TYMP*, *IL8*, *FIGF*, *RAF1*, *MMP1*
hsa00240	Pyrimidine metabolism	13	2.52^{-4}	*TYMS*, *UCK2*, *AK3*, *TYMP*, *POLE2*, *POLD1*, *DTYMK*, *NT5C1A*, *PRIM1*, *NME1*, *RRM2*, *TK1*, *NME5*

DEGs, differentially expressed genes; FDR, false discovery rate; KEGG, Kyoto Encyclopedia of Genes and Genomes.

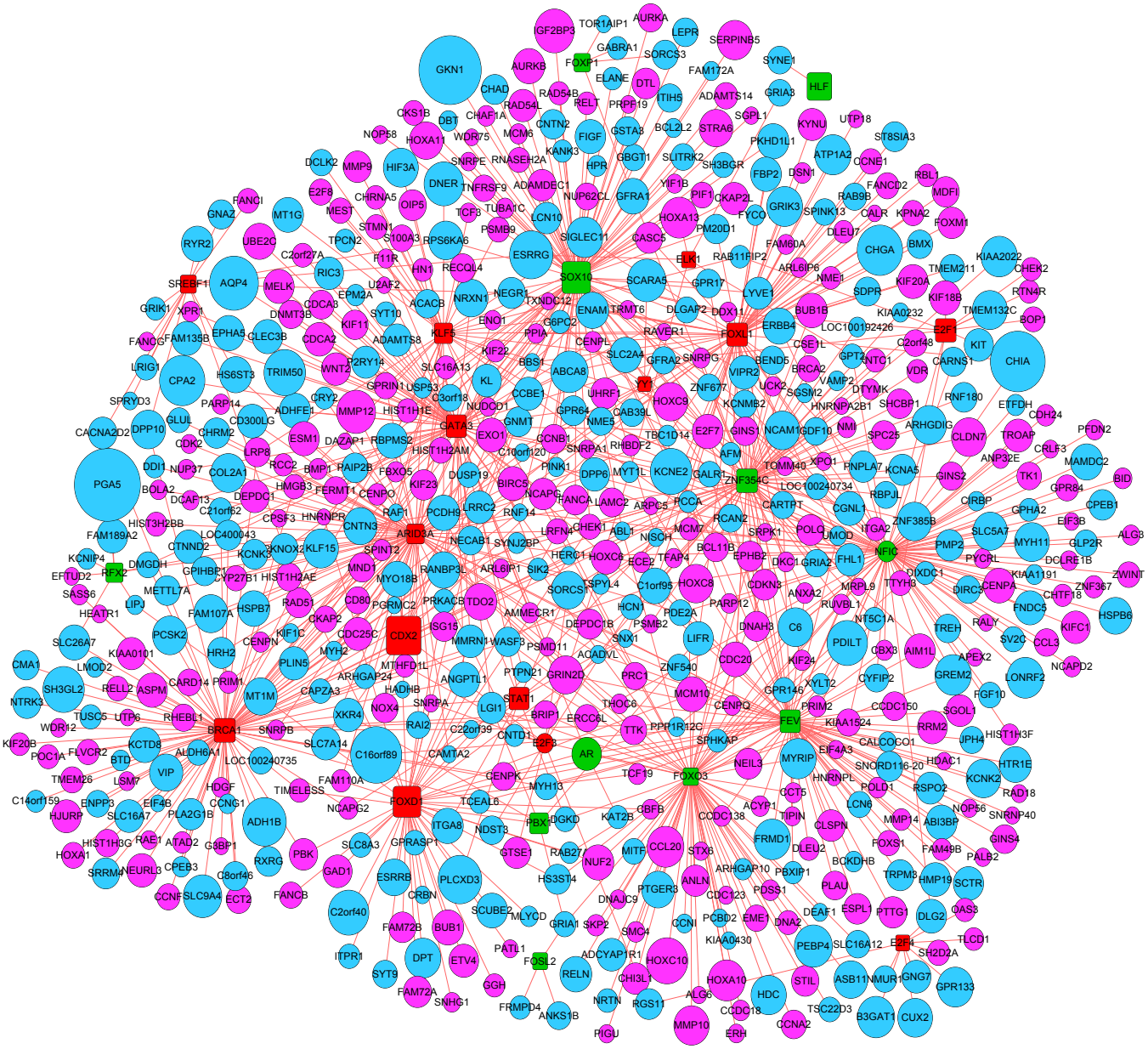

Figure 2 The TFs-genes network of ESCC. The red and green rectangles represented up- and down-regulated TFs in ESCC, respectively. The rose red and blue circular nodes represented up- and down-regulated DEGs in ESCC, respectively. The diameter of circles represented the $\log_2$ fold change. Solid lines indicated regulatory correlation between TFs and DEGs. TFs, transcription factors; ESCC, esophageal squamous cell carcinoma; DEGs, differentially expressed genes.

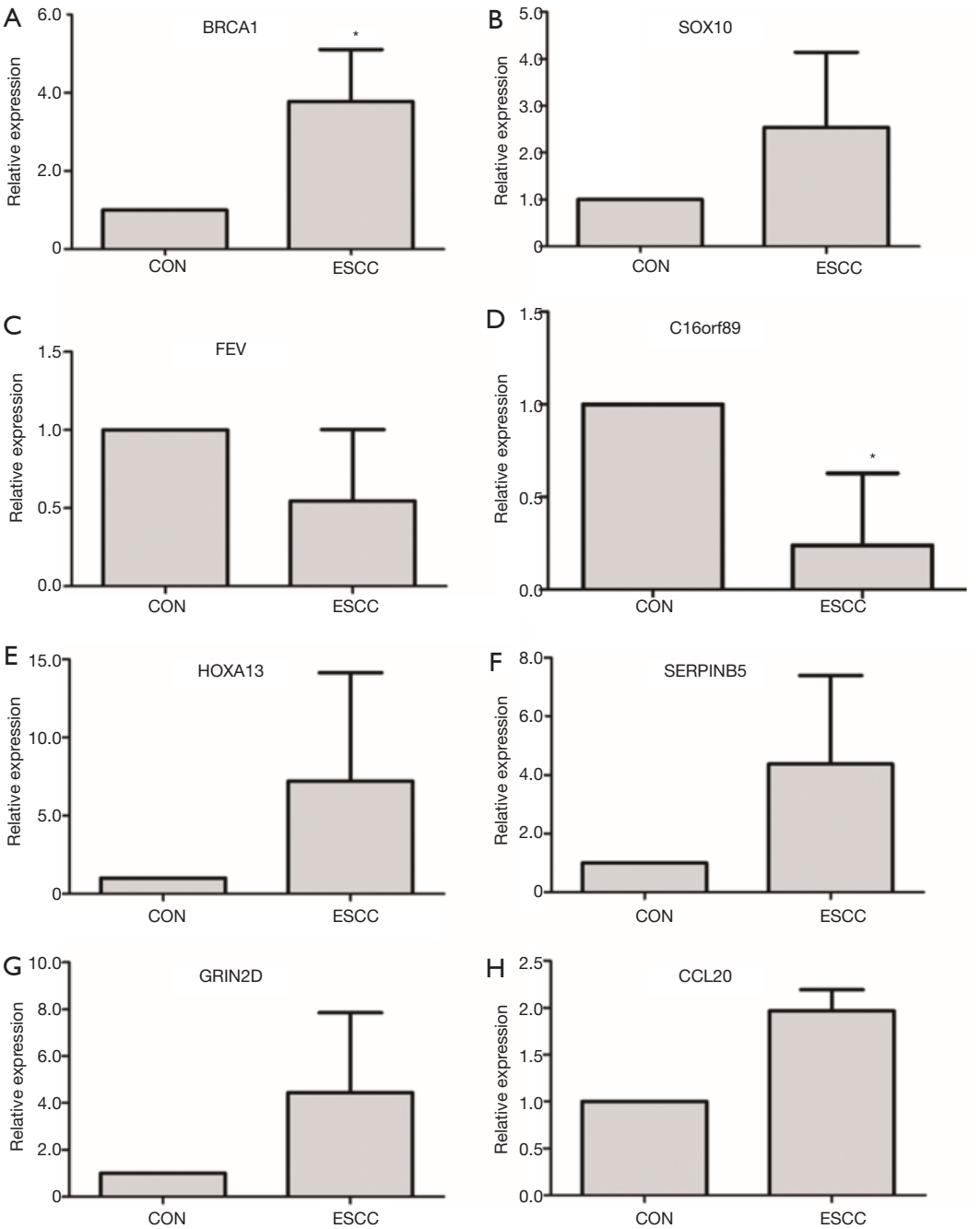

Figure 3 The qRT-PCR validation of the expression levels of TFs and DEGs in ESCC compared to adjacent non-tumor tissues. (A) The expression level of *BRCA1*; (B) the expression level of *SOX10*; (C) the expression level of *FEV*; (D) the expression level of *C16orf89*; (E) the expression level of *HOXA13*; (F) the expression level of *SERPINB5*; (G) the expression level of *GRIN2D*; (H) the expression level of *CCL20*. At least three independent experiments were performed for statistical evaluation. qRT-PCR experimental data were expressed as means ± SD. The statistical significance was evaluated using the Student's *t*-test and $P<0.05$ was considered as a significant difference. *, $P<0.05$. qRT-PCR, quantitative real-time polymerase chain reactions; TFs, transcription factors; ESCC, esophageal squamous cell carcinoma; DEGs, differentially expressed genes; CON, adjacent non-tumor tissues of ESCC.

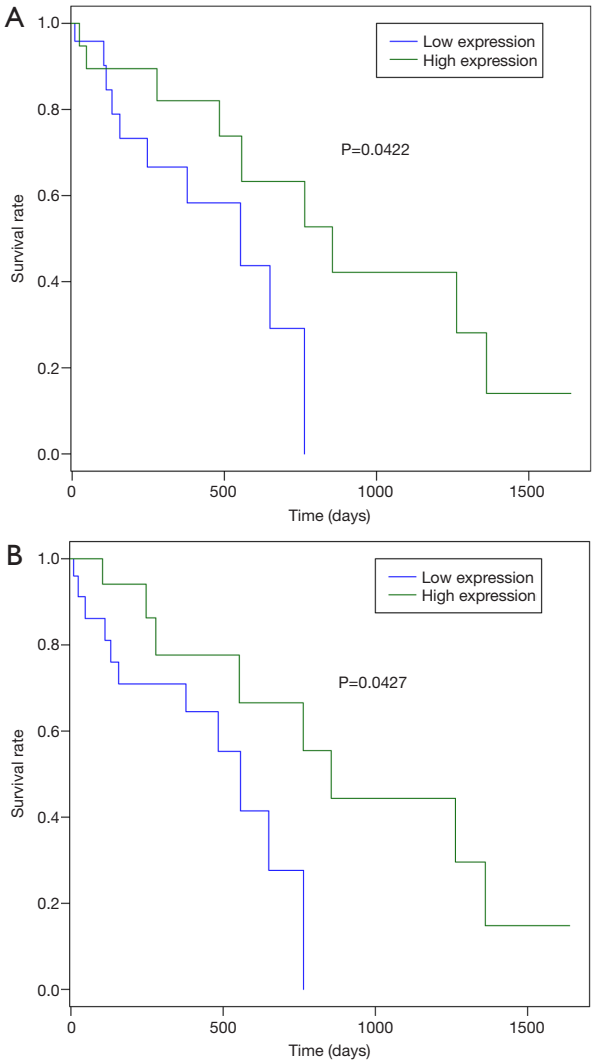

Figure 4 Kaplan-Meier survival curves between the expression of *TFAP2A* and *SREBF1* and the overall survival time of patients with ESCC. The x-axis represents the survival time (days) and the Y-axis represents cumulative survival rate. (A) Kaplan-Meier survival curves between the expression of *TFAP2A* and the overall survival time of patients with ESCC; (B) Kaplan-Meier survival curves between the expression of *SREBF1* and the overall survival time of patients with ESCC. ESCC, esophageal squamous cell carcinoma.

In the TFs-target genes regulatory network, *BRCA1*, *SOX10*, *ARID3A*, *ZNF354C* and *NFIC* had the high connectivity with DEGs, which regulated 92, 89 and 82, 79 and 78 DEGs, respectively.

BRCA1 was significantly up-regulated in ESCC compared to paired non-tumor tissues. *BRCA1* locates on chromosome 17 and encodes breast cancer 1. *BRCA1* is commonly known as mutations in this gene contribute to inherited breast cancers and ovarian cancers. In addition to, it is reported that high expression of *BRCA1* is responsible for chemotherapy resistance in ESCC treatment through experimental study *in vitro* and clinical observation. Low expression of *BRCA1* correlates with response rate of patients in advanced ESCC treated with cisplatin-/docetaxel-based chemoradiotherapy (20). Dasatinib enhances cisplatin sensitivity in human ESCC cells through suppression of PI3K/AKT and STAT3 pathways, and suppression of cisplatin-resistant molecules including *BRCA1* and *ERCC1* (21). In our study, *BRCA1* regulated *C16orf89*, which was significantly down-regulated in ESCC through microarray analysis and qRT-PCR verification (*Figure 3D*). *BRCA1* might be potential biomarkers for prognostic evaluation of chemotherapy treatment in ESCC patients.

SOX10 encodes SRY-box 10, a member of the RY-related HMG (high mobility group)-box family of TFs. This family is shared the highly conserved HMG sequences similarities. Our study demonstrated that *SOX10* regulated two of top 20 up-regulated DEGs including *HOXA13* and *SERPINB5*. The expression of these two DEGs had the up-regulated tendency in ESCC (*Figure 3E,F*) and *SERPINB5* was significantly enriched in p53 signaling pathway (*Table 4*). *SOX10* is broadly expressed in human normal adult tissues and fetal tissues. Previous studies present that *SOX10* is frequently silenced by promoter CpG methylation in colorectal cancer, gastric cancer and ESCC cell lines (22-24). The expression status of *SOX10* in ESCC in the report is consistent with our study that *SOX10* was down-regulated in ESCC compared to normal controls through bioinformatics analysis and qRT-PCR validation. The molecular function of *SOX10* in ESCC has been elucidated through experimental *in vitro*. Decreased mRNA expression of *SOX10* inhibits cell growth and metastasis of ESCC via suppressing the Wnt/β-catenin pathway (22). The experiment *in vivo* of biological function of *SOX10* in ESCC need to be further explored.

NFIC encodes nuclear factor I/C (CCAAT-binding TF), belongs to CTF/NF-I family. *NFIC* was down-regulated and regulated 78DEGs in the TFs-genes regulatory network of ESCC (*Figure 2*). The bioinformatics analysis of gene expression profiling reveals *NFIC* was dysregulated in gastric cancer and might contribute to the pathogenesis (25). NFI-C2, a dominating NFIC protein, it is lost during mammary tumor progression and

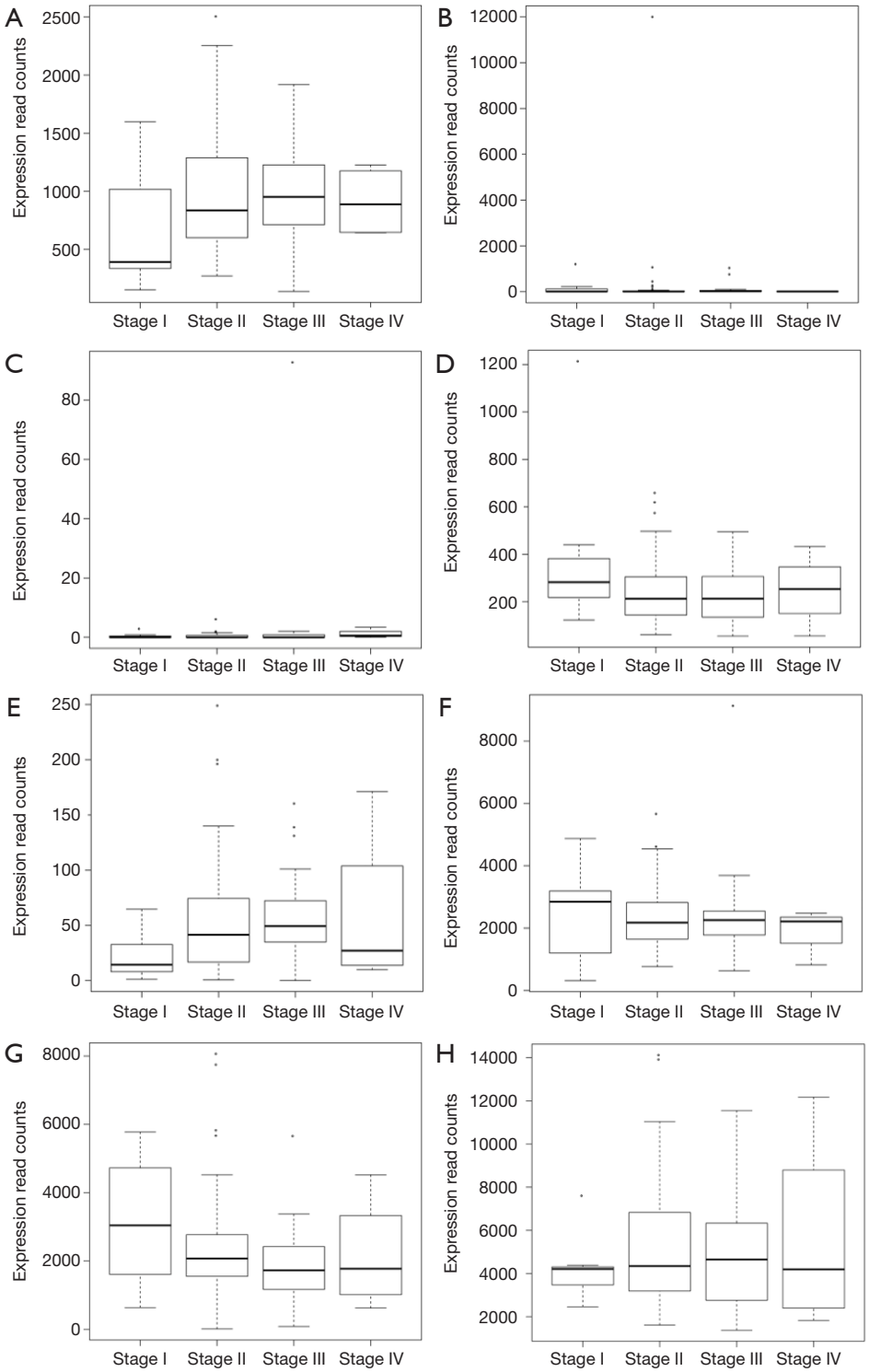

Figure 5 The expression levels of selected TFs in stage I–IV of patients with ESCC based on the TCGA illumina HiSeq RNA-Seq data. The X-axis represents the stages of ESCC and the Y-axis represents expression read counts of TCGA. (A) *BRCA1*; (B) *SOX10*; (C) *FEV*; (D) *ARID3A*; (E) *ZNF354C*; (F) *NFIC*; (G) *TFAP2A*; (H) *SREBF1*. *, P<0.05. TFs, transcription factors; ESCC, esophageal squamous cell carcinoma; TCGA, The Cancer Genome Atlas.

is almost invariably absent from lymph node metastases. Enhanced expression of NFI-C2 abolishes tumorigenicity and inhibits EMT and invasiveness of breast cancer through regulating the expression of *FOxF1 in vitro* and in nude mice (26). In addition to, *NFIC* suppresses EMT, migration, and invasion in breast cancer cells through regulating the expression level of *E-cadherin* and *KLF4* (27). To our knowledge, *NFIC* was a novel TF, which have not been reported to be related to ESCC before. *NFIC* was a key TF in the regulatory network (*Figure 2*). Based on the abovementioned, *NFIC* might play essential roles in the carcinogenesis and development of ESCC. The explicit biological functions of NFIC in ESCC need to be further elucidated through experimental exploration *in vitro* and *in vivo*.

FEV was one of down-regulated TFs in ESCC (*Figure 3C*). *C16orf89* and *GIN2D* were top 20 down-regulated and top 20 up-regulated DEGs of ESCC in our study. Both of them were targeted by *FEV*. *C16orf89*, a common putative target of *BRCA1* and *FEV*, is significantly down-regulated (*Figure 3D*) and *GIN2D* had the up-regulated tendency (*Figure 3G*) in ESCC; *CCL20* was the top 10 up-regulated DEGs and had the up-regulated tendency in ESCC compared to adjacent non-tumor tissues. In a word, the results of qRT-PCR verification of DEGs and TFs were consistent with our microarray analysis; it suggested that our integrated analysis was acceptable.

The other two TFs, *ARID3A* and *ZNF354C*, was identified as up- and down-regulated in ESCC, respectively. *ARID3A*, encodes AT-rich interaction domain 3A, is a member of the ARID family of DNA-binding proteins. Previous study indicates that *ARID3A* is dysregulated in colorectal cancer (CRC) and might be potentially prognostic biomarker of CRC (28). In addition, *ARID3A* is over-expressed in stomach cancer (29). *ZNF354C* encodes zinc finger protein 354C. *ZNF354C* is a key TF in gastric cardia adenocarcinoma (GCA) that dysregulated genes in GCA compared with adjacent normal tissues tend to be bound by *ZNF354C* (30). *ARID3A* and *ZNF354C* were novel TFs, which have not been reported to be related to ESCC before. The expression status of *ARID3A* and *ZNF354C* need to be validated through large sample size of ESCC specimens; the molecular functions need to be explored through experimental observation.

Taken together, these five TFs (*BRCA1*, *SOX10*, *ARID3A*, *ZNF354C* and *NFIC*) mentioned above were identified key TFs in ESCC. To better research the molecular mechanism of these five key TFs in ESCC. We obtained the common targeted genes of these five key TFs including *CDC25C*, *MCM6*, *ABL1*, *CHEK1*, *CCNB1*, *MCM7*, *TTK*, *CDK2*, *CDC20* and *BUB1B* (underlined genes in *Tables S5* and *4*). Moreover, all these ten genes were significantly enriched in cell cycle pathway based on the KEGG enrichment analysis. Deregulation of the cell cycle is one of the most frequent alterations during tumor development. Cancer originates from the deregulation of regulators of cell cycle (31). We speculated that these five key TFs may play important roles in the development in ESCC by modulating the cell cycle of ESCC cancer cells through regulating these ten cell cycle-related DEGs. Further research was needed to explore the precise roles of these TFs and DEGs in ESCC.

After evaluating the correlation between TFs and survival in patients with ESCC, we found that high expression of *SREBF1* and *TFAP2A* were significantly correlated with the longer overall survival time of patients with ESCC. These two TFs may also play important roles in ESCC which may be potential prognostic biomarkers of ESCC.

In conclusion, our integrated analysis identified key TFs in ESCC which provides clues for exploring the mechanism of ESCC. However, there are limitations in our study. Firstly, the small size of the clinical samples for qRT-PCR validation was small. Secondly, only little clinical information of samples of these ten datasets used in our study can be obtained by GEO. Hence, it is difficult to analyze the confounding effects of age, gender, TNM staging, and histological differentiation on gene expression in ESCC. Thirdly, there were radical differences of samples used between our integrated analysis and qRT-PCR results. Fourthly, whether these TF-gene sets are only specific to ESCC or just related to esophageal cancer or cancer was uncertain. All these four limitations may be the reasons why most of qRT-PCR results were not significant. Lastly, the five TFs of *FEV*, *ARID3A*, *ZNF354C*, *SREBF1* and *TFAP2A* were identified as the novel TFs in ESCC, but the biological functions of the three TFs were not explored in our work. Thus, the large sample size of ESCC specimens will be collected and validated through qRT-PCR verification; the confounding effects of clinical characteristics such as age, gender, histological differentiation and race on gene expression in ESCC need to be analyzed in future research; further investigation of these TF-gene sets in EAC and other cancers was needed to identify the specificity of these TF-gene sets in ESCC, further studies are needed to investigate the precise role of TFs and target-genes in ESCC in the future work as well.

Acknowledgements

Funding: The work was supported by Major Medical Scientific Research Subject of Hebei Province (zd2013044) and Medical Application Tracking Project of Hebei Province (GL201637).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: Our study was approved by the Ethics Committee of The Fourth Hospital of Hebei Medical University (No. 2015MEC073). All of patients signed the consent form and our study compiled with the Declaration of Helsinki.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J Clin 2012;62:10-29.
2. Rustgi AK, El-Serag HB. Esophageal Carcinoma. New Engl J Med 2014;371:2499-509.
3. Kumagai N, Wakai T, Akazawa K, et al. Heavy alcohol intake is a risk factor for esophageal squamous cell carcinoma among middle-aged men: A case-control and simulation study. Mol Clin Oncol 2013;1:811-6.
4. Rafiq R, Shah IA, Bhat GA, et al. Secondhand Smoking and the Risk of Esophageal Squamous Cell Carcinoma in a High Incidence Region, Kashmir, India: A Case-control–observational Study. Medicine (Baltimore) 2016;95:e2340.
5. Cao F, Han H, Zhang F, et al. HPV infection in esophageal squamous cell carcinoma and its relationship to the prognosis of patients in northern China. Scientific World Journal 2014;2014:804738.
6. Bhat GA, Shah IA, Makhdoomi MA, et al. CYP1A1 and CYP2E1 genotypes and risk of esophageal squamous cell carcinoma in a high-incidence region, Kashmir. Tumour Biol 2014;35:5323-30.
7. Yin J, Wang X, Zheng L, et al. Hsa-miR-34b/c rs4938723 T>C and hsa-miR-423 rs6505162 C>A polymorphisms are associated with the risk of esophageal cancer in a Chinese population. PLoS One 2013;8:e80570.
8. Shen F, Chen J, Guo S, et al. Genetic variants in miR-196a2 and miR-499 are associated with susceptibility to esophageal squamous cell carcinoma in Chinese Han population. Tumour Biol 2016;37:4777-84.
9. Yang X, Wang Q, Dai W, et al. Overexpression of zinc finger E-box binding homeobox factor 1 promotes tumor invasiveness and confers unfavorable prognosis in esophageal squamous cell carcinoma. Tumour Biol 2014;35:11977-84.
10. Gao H, Teng C, Huang W, et al. SOX2 Promotes the Epithelial to Mesenchymal Transition of Esophageal Squamous Cells by Modulating Slug Expression through the Activation of STAT3/HIF-α Signaling. Int J Mol Sci 2015;16:21643-57.
11. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30:207-10.
12. Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey VJ, Huber W, et al. editiors. Bioinformatics and computational biology solutions using R and Bioconductor. Berlin: Springer, 2005:397-420.
13. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995;57:289-300.
14. Eden E, Navon R, Steinfeld I, et al. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 2009;10:48.
15. Carmona-Saez P, Chagoyen M, Tirado F, et al. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. Genome Biol 2007;8:R3.
16. Matys V, Fricke E, Geffers R, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res 2003;31:374-8.
17. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498-504.
18. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc 2008;3:1101-8.
19. Su P, Wen S, Zhang Y, et al. Identification of the Key Genes and Pathways in Esophageal Carcinoma. Gastroenterol Res Pract 2016;2016:2968106.
20. Gao Y, Zhu J, Zhang X, et al. BRCA1 mRNA expression as a predictive and prognostic marker in advanced esophageal squamous cell carcinoma treated with cisplatin- or docetaxel-based chemotherapy/chemoradiotherapy. PLoS One 2013;8:e52589.
21. Chen J, Lan T, Zhang W, et al. Dasatinib enhances cisplatin sensitivity in human esophageal squamous cell carcinoma (ESCC) cells via suppression of PI3K/AKT and

Stat3 pathways. Arch Biochem Biophys 2015;575:38-45.
22. Tong X, Li L, Li X, et al. SOX10, a novel HMG-box-containing tumor suppressor, inhibits growth and metastasis of digestive cancers by suppressing the Wnt/β-catenin pathway. Oncotarget 2014;5:10571-83.
23. Ying J, Li H, Yu J, et al. WNT5A exhibits tumor-suppressive activity through antagonizing the Wnt/β-catenin signaling, and is frequently methylated in colorectal cancer. Clin Cancer Res 2008;14:55-61.
24. Wang J, Xia Y, Li L, et al. Double restriction-enzyme digestion improves the coverage and accuracy of genome-wide CpG methylation profiling by reduced representation bisulfite sequencing. BMC Genomics 2013;14:11.
25. Jiang HB, Yang TJ, Lu P, et al. Gene expression profiling of gastric cancer. Eur Rev Med Pharmacol Sci 2014;18:2109-15.
26. Nilsson J, Helou K, Kovács A, et al. Nuclear Janus-activated kinase 2/nuclear factor 1-C2 suppresses tumorigenesis and epithelial-to-mesenchymal transition by repressing Forkhead box F1. Cancer Res 2010;70:2020-9.
27. Lee HK, Lee DS, Park JC. Nuclear factor I-C regulates E-cadherin via control of KLF4 in breast cancer. BMC Cancer 2015;15:113.
28. Song M, Kim H, Kim WK, et al. High expression of AT-rich interactive domain 3A (ARID3A) is associated with good prognosis in colorectal carcinoma. Ann Surg Oncol 2014;21 Suppl 4:S481-9.
29. Samyesudhas SJ, Roy L, Cowden Dahl KD. Differential expression of ARID3B in normal adult tissue and carcinomas. Gene 2014;543:174-80.
30. Ju F, Lu L, Zhao QY, et al. Systematic analysis of gene expression and molecular interactions in cardiac and non-cardiac gastric carcinomas. Hepatogastroenterology 2014;61:1835-42.
31. Park MT, Lee SJ. Cell cycle and cancer. J Biochem Mol Biol.

Cite this article as: Zhang Y, Xu Y, Li Z, Zhu Y, Wen S, Wang M, Lv H, Zhang F, Tian Z. Identification of the key transcription factors in esophageal squamous cell carcinoma. J Thorac Dis 2018;10(1):148-161. doi: 10.21037/jtd.2017.12.27

Supplementary

Table S1 Basic information of five ESCC patients for qRT-PCR validation

Patient No.	Age (years)	Gender	Histological type	TNM staging
1	69	Male	ESCC	T3N0M0
2	71	Male	ESCC	T3N2M0
3	61	Male	ESCC	T3N0M0
4	48	Female	ESCC	T3N1M0
5	62	Male	ESCC	T3N0M0

No., number; ESCC, esophageal squamous cell carcinoma; TNM, tumor, node and metastasis; qRT-PCR, quantitative real time polymerase chain reaction.

Table S2 The primers of genes for qRT-PCR detection

Gene symbol	Primers (5' to 3')	Sequence of genes (bp)
BRCA1	Forward-AAGGAGCTTTCATCATTCACCC	151
	Reverse-CTACACTGTCCAACACCCACTCTC	
SOX10	Forward-AGGCGGACGATGACAAGTTC	126
	Reverse-CGTGCGGCTTGCTTTTG	
FEV	Forward-AGAGCAAGCCCAACATGAACTA	117
	Reverse-CCTGGAAGTCGAAGCGGTAG	
FOXO3	Forward-TCGCAATGATCCGATGATGTC	244
	Reverse-CAAGGAGGAGCCTGAGAGAGAGT	
C16orf89	Forward-CACACAGGGACCACTCCAACAG	152
	Reverse-AGCCGCCCATTCCACAGA	
GKN1	Forward-ACTCCTCTGTCCACTGCTTTCG	155
	Reverse-ACTCACTGACTGCTGCCCACTT	
HOXA13	Forward-CCAAATGTACTGCCCCAAAGA	229
	Reverse-CTCAGAGAGATTCGTCGTGGCT	
SERPINB5	Forward-TCAACAAGACAGACACCAAACCA	258
	Reverse-TGGAGAGTTTGACCTTGGCATT	
GRIN2D	Forward-CCCTGAAGTTTGGGACCGT	134
	Reverse-CCCTGCCTTGAGCTGAGTGA	
CCL20	Forward-GTGTGCGCAAATCCAAAACAGACT	173
	Reverse-CTAAACCCTCCATGATGTGCAAGTG	
β-*actin*	Forward-CTGAAGTACCCCATCGAGCAC	223
	Reverse-ATAGCACAGCCTGGATAGCAAC	

qRT-PCR, quantitative real-time polymerase chain reactions.

Table S3 The full list of identified transcription factors in ESCC

TFs	Log_2FC	P value	FDR
AR	–3.28003	7.54^{-12}	4.33^{-10}
SOX10	–3.06912	1.26^{-5}	0.0001353
HLF	–2.39581	0.0001442	0.0010635
ZNF354C	–1.73515	0.0006923	0.0039241
FEV	–1.62243	0.0015651	0.0076703
NFIC	–1.35645	1.21^{-8}	3.29^{-7}
PBX1	–1.33592	0.0017248	0.0083301
RFX2	–1.08686	0.0001603	0.001158
FOXP1	–0.90709	0.0002061	0.0014285
FOXO3	–0.7663	3.83^{-6}	4.91^{-5}
FOSL2	–0.69547	0.0015903	0.0077752
YY1	0.360324	0.0012472	0.0063631
E2F4	0.518079	0.0005682	0.0033263
ELK1	0.59021	0.0001143	0.0008751
SREBF1	0.876235	0.0003207	0.0020626
ARID3A	1.126115	0.0016557	0.0080472
E2F3	1.134516	8.44^{-8}	1.78^{-6}
KLF5	1.246651	0.0004792	0.0028951
STAT1	1.56075	1.15^{-8}	3.16^{-7}
GATA3	1.59008	0.0016735	0.0081201
FOXL1	1.683068	0.0003848	0.0024079
E2F1	1.701482	5.22^{-12}	3.13^{-10}
BRCA1	1.786722	3.03^{-12}	1.90^{-10}
FOXD1	2.931021	2.44^{-7}	4.56^{-6}
TFAP2A	3.549122	1.05^{-9}	3.79^{-8}
CDX2	4.15098	0.0001603	0.001158

TFs, transcription factors; FC, fold change; ESCC, esophageal squamous cell carcinoma; FDR, false discovery rate.

Table S4 The top 5 TFs had the high connectivity with DEGs in ESCC

TFs	Log_2FC	Up/down	Count	Genes
BRCA1	1.79	Up	92	*ARL6IP1*, *RAE1*, *C8orf46*, *KIF20B*, *SASS6*, *TCEAL6*, *KIAA0101*, *PAIP2B*, *HIST1H3G*, *RNF14*, *ENPP3*, *XKR4*, *SH3GL2*, *PLA2G1B*, *KCTD8*, *CENPN*, *C16orf89*, *ISG15*, *CDC25C*, *ASPM*, *RHEBL1*, *PKNOX2*, *CCNF*, *MTHFD1L*, *CMA1*, *HIST1H1E*, *WDR12*, *BTD*, *TUSC5*, *MYO18B*, *CENPK*, *LOC100240735*, *CAPZA3*, *ATAD2*, *EIF4B*, *NEURL3*, *TDO2*, *TMEM26*, *SYNJ2BP*, *RELL2*, *SPINT2, PRIM1*, *DUSP19*, *LSM7*, *HJURP*, *PBK*, *RANBP3L*, *CKAP2*, *NCAPG2*, *RXRG*, *HOXA1, PINK1*, *KIF23*, *LRP8*, *C22orf39*, *HMGB3*, *SLC16A7*, *LMOD2*, *ERCC6L*, *VIP*, *RAF1*, *C1orf95*, *BRIP1*, *UTP6*, *CACNA2D2*, *ALDH6A1, CCNG1*, *FLVCR2*, *CARD14*, *LOC400043*, *G3BP1*, *HDGF, DMGDH*, *C21orf62*, *HIST1H2AM*, *C14orf159*, *ANGPTL1*, *TIMELESS*, *SLC9A4*, *SRRM4*, *DPP10*, *ADH1B*, *WASF3*, *HERC1*, *SLC26A7*, *PCSK2*, *HEATR1*, *NTRK3*, *POC1A*, *CPEB3*, *MYH13*, *ECT2*
SOX10	–3.07	Down	89	*AURKA*, *NRXN1, NECAB1, RAD54B*, *FANCA*, *SLC2A4*, *CHAD*, *VIPR2*, *STRA6*, *DDX11*, *AURKB*, *LCN10*, *GNMT*, *HPR*, *ITIH5*, *ADAMDEC1*, *FIGF*, *KANK3*, *ELANE*, *PINK1*, *NFIC*, *PAIP2B*, *E2F7*, *PRPF19*, *NOP58*, *PIF1*, *ENO1*, *ERBB4*, *RNASEH2A*, *GSTA3*, *DBT*, *PCCA*, *CHAF1A*, *MCM6*, *HIST1H2AM*, *DTL*, *CKS1B*, *C3orf18*, *GBGT1*, *USP53*, *BCL11B*, *WDR75*, *HOXA11*, *IGF2BP3*, *KCNMB2*, *GKN1*, *SLITRK2*, *TNFRSF9*, *SGPL1*, *KL*, *RAD54L*, *KIF23*, *KIF22*, *SNRPE*, *YIF1B*, *PSMB9*, *LYVE1*, *SNRPG*, *ADAMTS14*, *CKAP2L*, *TUBA1C*, *GALR1*, *HOXA13*, *RNF14*, *GPR17*, *BCL2L2*, *NUP62CL*, *HIST1H1E*, *PKHD1L1*, *FAM172A*, *AFM*, *PM20D1*, *BBS1*, *ABL1*, *ZNF677*, *SLC16A13*, *RPS6KA6*, *SERPINB5*, *CNTN2*, *RELT*, *LRRC2*, *NISCH*, *CHEK1*, *CCNB1*, *LEPR*, *NUDCD1*, *SH3BGR*, *NCAPG*, *HNRNPR*
ARID3A	1.13	Up	82	*MND1*, *RAD51*, *GNMT*, *ARHGAP24*, *ACADVL*, *SNRPB*, *PLIN5*, *CHRM2*, *CAPZA3*, *BMP1*, *MMP12*, *KCNK3*, *GPIHBP1*, *RCC2*, *HIST3H2BB*, *NOX4*, *NECAB1*, *HRH2*, *MT1M*, *MYH2*, *HIST1H2AE*, *BBS1*, *CNTD1*, *GPRIN1*, *KIF22*, *KLF15*, *CDC25C*, *BIRC5*, *PRIM1*, *CD300LG*, *BOLA2*, *FAM107A*, *HNRNPR*, *LRRC2*, *C3orf18*, *CD80*, *KIF1C*, *KIF23*, *RBPMS2*, *FAM110A*, *CTNND2*, *USP53*, *DAZAP1*, *COL2A1*, *KL*, *TTK*, *METTL7A*, *DPP6*, *SLC16A13*, *GLUL*, *DEPDC1B*, *LAMC2*, *RNF14*, *ZNF540*, *AMMECR1*, *SNRPA*, *HSPB7*, *PCSK2*, *TBC1D14*, *ESM1*, *MCM7*, *C10orf120*, *SLC7A14*, *CDK2*, *HADHB*, *DDI1*, *CENPN*, *HS6ST3*, *DEPDC1*, *THOC6*, *PGRMC2*, *CAMTA2*, *ECE2*, *TSPYL4*, *FERMT1*, *CPSF3*, *RAI2*, *GRIN2D*, *PDE2A*, *CRY2*, *CCNB1*, *CYP27B1*
ZNF354C	–1.74	Down	79	*GRIA2*, *ANXA2*, *ZNF677*, *HNRNPA2B1*, *RBPJL*, *SGSM2*, *CGNL1*, *SLC2A4*, *GPR17*, *BEND5*, *VAMP2*, *GREM2*, *UCK2*, *TDO2*, *CHEK1*, *NCAM1*, *ISG15*, *KIF24*, *ZNF385B*, *LAMC2*, *FHL1*, *HCN1*, *GPR64*, *G6PC2*, *TSPYL4*, *ARHGDIG*, *CDC20*, *NMI*, *LOC100240734*, *DNAH3*, *NFIC*, *HOXC8*, *GFRA1*, *SPC25*, *TRMT6*, *SNRPG*, *BCL11B*, *KCNA5*, *GINS2*, *LYVE1*, *LOC100192426*, *SNRPA1*, *DPP6*, *NME5*, *GPR146*, *WASF3*, *KIAA0232*, *RANBP3L*, *KIAA2022*, *TOMM40*, *SORCS1*, *SYNJ2BP*, *VDR*, *CSE1L*, *RUVBL1*, *BRCA2*, *DDX11*, *CDKN3*, *ARL6IP1*, *KNTC1*, *C1orf95*, *POLQ*, *CENPQ*, *DKC1*, *XPO1*, *DTYMK*, *PSMB2*, *CLDN7*, *RNF180*, *GDF10*, *MRPL9*, *TFAP4*, *BRIP1*, *BUB1B*, *C2orf48*, *SHCBP1*, *NT5C1A*, *DIXDC1*, *GINS1*
NFIC	–1.36	Down	78	*DLG2*, *CRLF3*, *ETFDH*, *XYLT2, C6*, *CHTF18*, *SV2C*, *CIRBP*, *PMP2*, *CPEB1*, *NT5C1A*, *PNPLA7*, *CARNS1*, *ARPC5*, *DIRC3*, *ZWINT*, *KCNMB2*, *SLC5A7*, *RNF14*, *NISCH*, *SPHKAP*, *KIFC1*, *GPT2*, *LONRF2*, *HOXC9*, *PDILT*, *ZNF367*, *BID*, *RHBDF2*, *KIAA1191*, *ERBB4*, *CCL3*, *EPHB2*, *GPHA2*, *KCNE2*, *CHIA*, *ALG3*, *HIST1H3F*, *MYH11*, *MCM10*, *EIF3B*, *TREH*, *CCNB1*, *CDH24*, *PDE2A*, *C1orf95*, *MAMDC2*, *PINK1*, *UHRF1*, *CYFIP2*, *AIM1L*, *PFDN2*, *PPP1R12C*, *HERC1*, *GPR84*, *ZNF677*, *ERCC6L*, *VIPR2*, *PARP12*, *MYRIP*, *MCM7*, *PYCRL*, *GLP2R*, *TK1*, *FGF10*, *CBX3*, *FNDC5*, *TROAP*, *CENPA*, *GREM2*, *HSPB6*, *APEX2*, *DIXDC1*, *RALY*, *ANP32E*, *DCLRE1B*, *NCAPD2*, *KIAA1524*

TFs, transcription factors; DEGs, differentially expressed genes; ESCC, esophageal squamous cell carcinoma; FC, fold change.