Bioinformatics Analyses of m-RNA Profiling Following Ezrin Knockdown in Esophageal Squamous Cell Carcinoma

Bingli Wu[1,†], Chunquan Li[1,2,†], Jianjun Xie[1], Zepeng Du[3], Liewei Luo[4], Jianyi Wu[1], Pixian Zhang[1], Liyan Xu[5,*] and Enmin Li[1,*]

[1]Department of Biochemistry and Molecular Biology, Shantou University Medical College, Shantou, Guangdong Province, China
[2]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang Province, China
[3]Department of Pathology, Shantou Central Hospital, Affiliated Shantou Hospital of Sun Yat-sen University, Shantou, Guangdong Province, China
[4]Department of Biochemistry and Molecular Biology, Guangdong Pharmaceutical University Guangzhou Higher Education Mega Center, Guangzhou, Guangdong Province, China
[5]Institute of Oncologic Pathology, Shantou University Medical College, Shantou, Guangdong Province, China
[†]These authors contributed equally to this work.

Abstract

Ezrin is involves in multiple cancer cell functions. In this study, differentially expressed genes (DEGs) identified from mRNA expression profiling following Ezrin knockdown in esophageal squamous cell carcinoma (ESCC), were analyzed by multiple bioinformatics methods for a comprehensive understanding. Gene Ontology enrichment found significant terms related to immunity, stimulus response, extracellular matrix-binding and signal transduction. Functional Annotation Chart showed except GO categories, these DEGs were annotated by 72 functional category terms. Subpathway analyses revealed the DEGs were involved in 36 subpathways, overwhelming the traditional pathway enrichment. Promoter analyses showed specific sequence patterns and transcription factors correspond to the co-downregulation and co-upregulation of DEGs. MNB1A, DOF2, PBF, YY1, PRRX2, UBX and ARNT-AHR co-regulate the downregulated genes, whereas GATA2, ZNF42, deltaEF1, MYCN and ARNT co-regulate the upregulated genes. This paper provides a workflow for a better understanding the roles of Ezrin knockdown in ESCC by the bioinformatics analyses of its DEGs.

Keywords: Ezrin; Bioinformatics; Differentially expressed genes; Esophageal squamous cell carcinoma

Abbreviations: DEG: Differentially Expressed Genes; ERM: Ezrin–Radixin–Moesin; ESCC: Esophageal Squamous Cell Carcinoma; GO: Gene Ontology; DAVID: Database for Annotation, Visualization and Integrated Discovery; KEGG: Kyoto Encyclopedia of Genes and Genomes; UCSC: University of California Santa Cruz

Introduction

Ezrin, also named VIL2, is a member of the Ezrin–radixin–moesin (ERM) protein family. Ezrin acts as a linker between the actin cytoskeleton and plasma membrane proteins, as well as a signal transducer involving in cytoskeletal remodeling [1]. Ezrin has been implicated in multiple aspects of cancer cell function, especially the metastasis of cancer cells [2-4]. Moreover, Ezrin translocates to the nucleus to regulate gene transcription in several cancer cell types [5-7]. Our previous researches show Ezrin is over-expressed in human esophageal squamous cell carcinoma (ESCC), which is associated with the invasive phenotype of malignantly transformed esophageal epithelial cells [8]. We subsequently found Ezrin translocates from the plasma membrane to the cytoplasm in the progression from normal epithelium to invasive carcinoma of esophageal epithelial cells [9].

The knockdown of Ezrin expression by RNA interference leads to suppression of growth, adhesion and invasiveness of ESCC cells *in vitro*, and tumorigenicity in nude mice also reveals that Ezrin regulates tumor formation *in vivo* [10]. Furthermore, mRNA expression profiling was obtained using Affymetrix GeneChip Human genome U133 plus 2.0 and indicated that certain proliferation- and invasiveness-related genes are involved in Ezrin function [10]. However, the biological meaning of this mRNA profile is not fully understood. For a comprehensive understanding of the roles of Ezrin, we re-analyzed this mRNA expression profile by multiple bioinformatic methods.

Materials and Methods

Identification of differentially expressed genes

The mRNA expression profiling following Ezrin knockdown by RNAi in ESCC EC109 cell line is available from GEO database (GSE6233) at http://www.ncbi.nlm.nih.gov/geo/. The raw data is normalized and log transformed. The differentially expressed genes (DEGs) are identified based on a threshold of 2-fold change.

Gene ontology enrichment and functional annotation

Gene Ontology (GO) enrichment analysis was performed using the WebGestalt webserver, which contains information from Gene Ontology Tree Machine software (http://bioinfo.vanderbilt.edu/webgestalt/) [11]. The whole human genome gene set was set as a reference list. The hypergeometric statistical method test was used, and only statistically enriched terms ($P < 0.05$) with at least 2 genes were selected.

DAVID (http://david.abcc.ncifcrf.gov/) consists of an integrated biological knowledge base and analytic tools aimed at systematically

extracting biological meaning from large gene/protein lists. Functional Annotation Chart in DAVID bioinformatics aims to identify over-represented biological terms associated with a given gene list [12]. Currently, Functional Annotation Chart provides over 40 category enrichments. Except for GO terms, it also includes protein functional domains, protein–protein interactions, sequence features, disease associations, pathways, homology, gene functional summaries and literature. The significantly enriched terms, from Functional Annotation Chart, with $P < 0.05$ were visualized by Enrichment Map plugin of Cytoscape [13].

KEGG pathway and subpathway analysis

The DEGs were enriched in the KEGG pathways by applying the bioconductor SubpathwayMiner package. Except for entire pathways, Subpathway Miner could also detect the local pathway structure (subpathway), which helps to understand the detailed information of the genes of interest [14]. This package adopts the k-clique concept to define a subpathway. All possible subpathways were obtained by searching all possible paths between start-points and end-points in the adjacency matrix generated by node relationships in an entire KEGG pathway. The k was set to 4 in this study, which means that the distance among all genes within the subpathways was no greater than 4.

Promoter sequence patterns and potential transcription factors of DEGs

The 2000 bp promoter sequence of the top 20 most changed downregulated and upregulated genes were downloaded from the UCSC genome database (http://genome.ucsc.edu), respectively. Promoter sequence patterns over-represented or down-represented in these two sequence sets were analyzed by the POCO program (http://ekhidna. biocenter.helsinki.fi/poxo/poco/poco). POCO is tool to find distinctly represented regulatory patterns from either one or two sequence
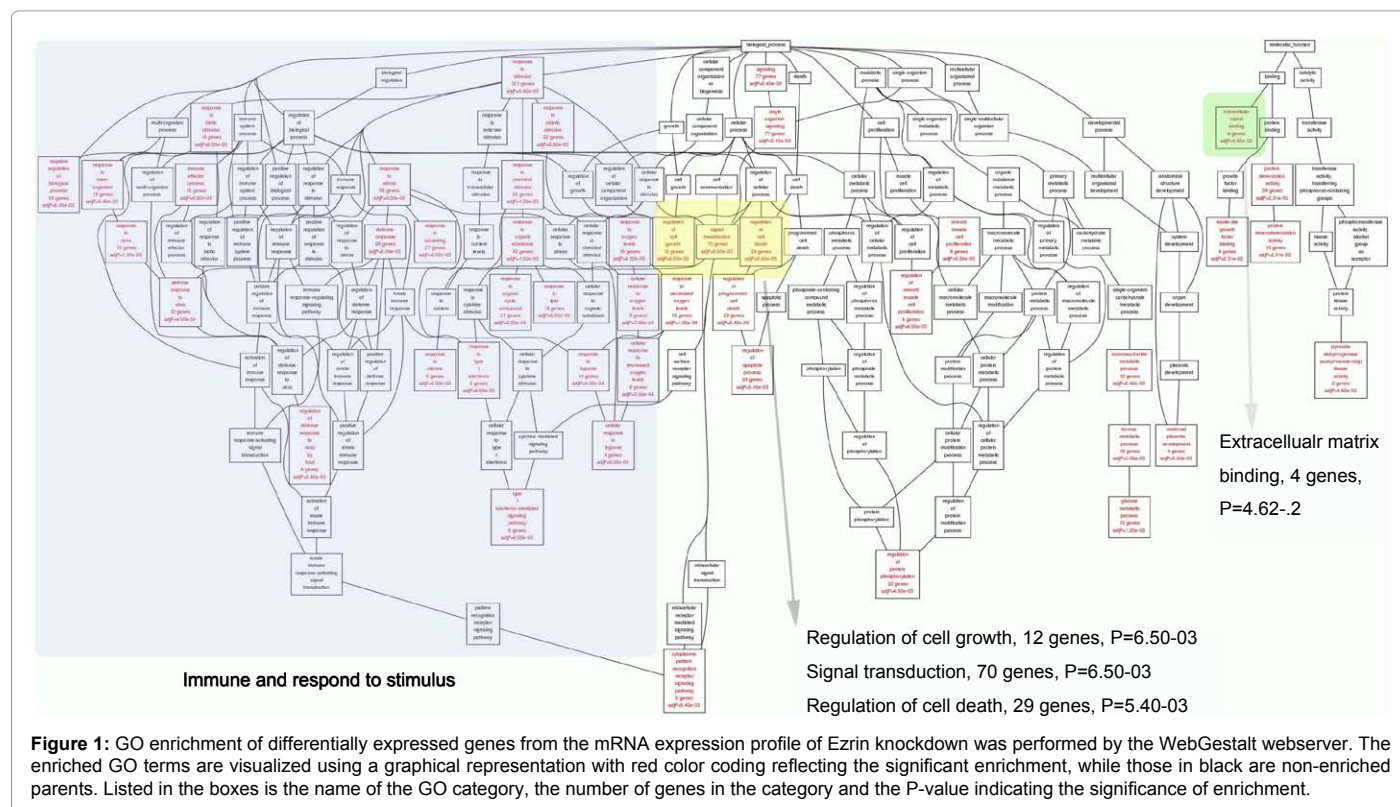
sets. The background organism was set as homo_sapiens_clean, and the longest pattern length was set at 8 in this study. Subsequently, significant sequence patterns were screened in the JASPAR database to identify recognized transcription factors (similarity index >0.70) [15].

Results

GO enrichment and functional annotation

We obtained 244 DEGs in total, including 199 upregulated genes and 45 downregulated genes, using a 2-fold change as the threshold (Supplementary File 1). To understand their functional categories, GO enrichment analysis was performed on these DEGs using hypergeometric test and multiple test adjustment method BH through WebGestalt (Figure 1). The result is showed in the tree view to illustrate the enriched GO structure. To our surprise, we found twelve Biological process terms related to immunity and response to stimuli, such as "response to stress" (65 genes, $P= 3.20$E-05), "response to biotic stimulus" (18 genes, $P= 6.50$E-03), and "defense response to virus" (12 genes, $P= 4.00$e-04). These are indicated in the left portion of the result graph (Figure 1). Ezrin interacts with both the plasma membrane and filamentous actin. GO terms associated with cytoskeleton organization were found. In the Cellular component enrichment, the term "extracellular matrix-binding" contains four genes (CYR61, DCN, SMOC1 and OLFML2B) with significantly altered expression ($P= 4.62$e-02), suggesting these 4 genes participate in extracellular matrix re-modeling. Moreover, Ezrin might affect signal transduction between cells. As many as 70 genes are enriched in "Signal transduction", suggesting intracellular signal transduction was significantly impacted by Ezrin knockdown. Dozens of the genes regulating cell growth and cell death are also enriched.

The DEGs were also clustered by Functional Annotation Chart in DAVID bioinformatics, and the enrichment result was visualized

Extracellualr matrix binding, 4 genes, P=4.62-.2

Regulation of cell growth, 12 genes, P=6.50-03

Signal transduction, 70 genes, P=6.50-03

Regulation of cell death, 29 genes, P=5.40-03

Immune and respond to stimulus

Figure 1: GO enrichment of differentially expressed genes from the mRNA expression profile of Ezrin knockdown was performed by the WebGestalt webserver. The enriched GO terms are visualized using a graphical representation with red color coding reflecting the significant enrichment, while those in black are non-enriched parents. Listed in the boxes is the name of the GO category, the number of genes in the category and the P-value indicating the significance of enrichment.
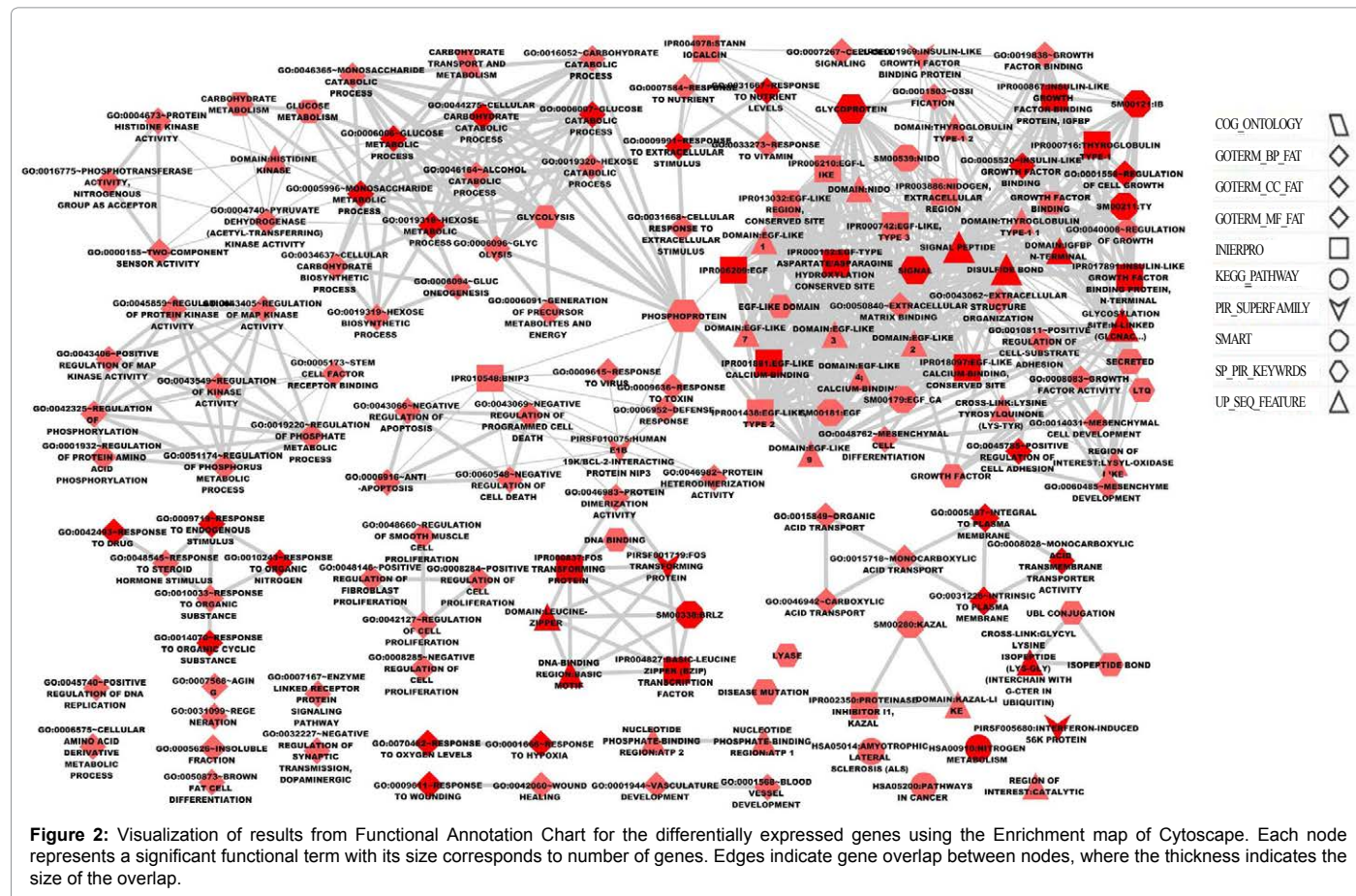
by the Enrichment Map in Cytoscape (Figure 2). Each functional category is represented by a node, and the node size represents the number of enriched genes (Figure 2). Nodes of the same functional category are indicated by the same shape. The edge between nodes is established as overlapping genes exist between these two nodes, and the width is defined as the overlap coefficient between these two categories. In total, 158 Functional Annotation Chart enrichments are identified, excluding 86 terms from three GO categories, but including 72 terms from the following annotation categories, 17 INTERPRO, 7 SMART, 17 SP_PIR_KEYWORDS, 23 UP_SEQ_FEATURE, 4 PIR_SUPERFAMILY, 1 COG_ONTOLOGY and 3 KEGG_PATHWAY. In the category of SP_PIR_KEYWORDS, as many as 52 genes are enriched in the term of "Signal", which suggests an important signaling transduction role for Ezrin. Ezrin knockdown greatly influenced the activities of other signaling molecules, with nine genes (MAFF, FOS, FOSL2, ATF3, HIST1H2BC, HIST1H2AG, POU5F1, PRDM1 and TNFAIP3) are enriched in the "DNA binding" term in the category of SP_PIR_KEYWORDS. On the other hand, the "DNA-binding region:Basic motif" term, in the category of UP_SEQ_FEATURE, contains 8 genes (MAFF, FOS, FOSL2, ATF3, SREBF1, MGA, NFIL3 and MXI1). A total of 13 unique genes, characterized by DNA-binding, appear to be critical downstream effectors of Ezrin, suggesting Ezrin exerts a large biological impact through gene transcription regulation. The top enrichment term in KEGG-PATHWAY, is "Pathways in cancer", which consists of 10 genes (FOS, CDKN1A, FZD10, BMP2, PTGS2, CASP9, PDGFA, ITGAV, EGLN3 and KITLG). These chart results would provide an overview of the biological impact of Ezrin knockdown in ESCC.

Pathway and subpathway enrichments

To identify how the downstream genes from Ezrin knockdown influence cell signal transduction, the DEGs were mapped to KEGG pathways by traditional entire KEGG pathway enrichment. DEGs are found enriched only in 10 entire pathways, of which only two pathways (Pathways in cancer and p53 signaling pathway) are related to carcinoma progression (Table 1). A subpathway is limited to local area of an entire pathway, which helps to understand how the interested genes affect the pathway locally. Using SubpathwayMiner analysis, our DEGs are found significantly enriched in 36 subpathways that corresponded to 21 entire pathways (Table 2). Two subpathways that derived from the entire Pathways in caner (path: 05200) are showed, path: 05200_4 involved 5 genes (CDKN1A, CASP9, PDGFA, ITGAV and KITLG), while path: 05200_20 contained 3 genes (CASP9, CDKN1A and FZD10) (Figure 3). Except for the 10 entire pathways found by the traditional pathway enrichment, dozens of subpathways derived from many other entire pathways are found which were not detected by traditional KEGG pathway enrichment because of the statistical threshold (Table 2).

Subpathways, such as Cytokine-cytokine receptor interaction and ECM-receptor interaction, are expected based on existing knowledge of Ezrin biology. Interestingly, several metabolism-related subpathways are unexpectedly over-represented in this enrichment analysis, for example, pathways of "Glycolysis/Gluconeogenesis", "Pentose phosphate pathway", "Fructose and mannose metabolism",

Figure 2: Visualization of results from Functional Annotation Chart for the differentially expressed genes using the Enrichment map of Cytoscape. Each node represents a significant functional term with its size corresponds to number of genes. Edges indicate gene overlap between nodes, where the thickness indicates the size of the overlap.

pathwayId	pathwayName	annMolecule Ratio#	p-value
path:00910	Nitrogen metabolism	4/287	0.0002
path:05014	Amyotrophic lateral sclerosis (ALS)	4/287	0.0052
path:00051	Fructose and mannose metabolism	3/287	0.0100
path:00010	Glycolysis/Gluconeogenesis	4/287	0.0101
path:05200	Pathways in cancer	10/287	0.0107
path:04115	p53 signaling pathway	4/287	0.0118
path:04360	Axon guidance	5/287	0.0258
path:04710	Circadian rhythm - mammal	2/287	0.0336
path:03018	RNA degradation	3/287	0.0393
path:00030	Pentose phosphate pathway	2/287	0.0489

#: the first number refers to the annotated DEGs in the pathway; the second number is total number of pathways.

Table 1: Enriched KEGG pathways of the DEGs.

Entire pathway ID	Entire pathway Name	subpathway ID	p-value
path:00010	Glycolysis/Gluconeogenesis	path:00010_1	0.0019
		path:00010_2	0.0022
		path:00010_5	0.0230
		path:00010_6	0.0308
		path:00010_7	0.0255
path:00030	Pentose phosphate pathway	path:00030_3	0.0489
		path:00030_6	0.0390
path:00051	Fructose and mannose metabolism	path:00051_1	0.0058
		path:00051_11	0.0047
		path:00051_2	0.0058
		path:00051_3	0.0230
		path:00051_7	0.0042
		path:00051_9	0.0255
path:00230	Purine metabolism	path:00230_14	0.0359
path:00910	Nitrogen metabolism	path:00910_1	0.0073
		path:00910_2	0.0002
		path:00910_3	0.0073
path:04060	Cytokine-cytokine receptor interaction	path:04060_39	0.0262
path:04115	p53 signaling pathway	path:04115_1	0.0484
path:04512	ECM-receptor interaction	path:04512_23	0.0390
path:04610	Complement and coagulation cascades	path:04610_5	0.0489
path:04620	Toll-like receptor signaling pathway	path:04620_15	0.0489
path:04622	RIG-I-like receptor signaling pathway	path:04622_1	0.0033
path:04672	Intestinal immune network for IgA production	path:04672_8	0.0262
path:04710	Circadian rhythm - mammal	path:04710_1	0.0308
		path:04710_2	0.0206
path:04912	GnRH signaling pathway	path:04912_5	0.0073
path:05012	Parkinson's disease	path:05012_3	0.0122
path:05014	Amyotrophic lateral sclerosis (ALS)	path:05014_2	0.0230
path:05160	Hepatitis C	path:05160_11	0.0184
		path:05160_8	0.0390
path:05200	Pathways in cancer	path:05200_20	0.0178
		path:05200_4	0.0033
path:05215	Prostate cancer	path:05215_9	0.0178
path:05222	Small cell lung cancer	path:05222_6	0.0390
path:05414	Dilated cardiomyopathy	path:05414_2	0.0336

Table 2. Enriched KEGG subpathways of the DEGs.

"Purine metabolism" and "Nitrogen metabolism". These pathways are not widely acknowledged, but are highly over-represented in our analysis. Of these five pathways, "Pentose phosphate pathway" and "Purine metabolism" are discovered by subpathway analysis.

Promoter sequence patterns in upregulated and downregulated genes

We assumed that in the promoters of upregulated and downregulated genes, there would be certain sequence patterns corresponding to transcriptional regulation for gene co-expression following Ezrin knockdown. POCO was applied to find over-represented and under-represented regulatory patterns between the upregulated and downregulated gene promoter sequence sets. We found 68 significant sequence patterns that are over-represented in the downregulated genes, but under-represented in the upregulated genes, while 64 significant sequence patterns are over-represented in the upregulated genes. The top 20 patterns presented in downregulated or upregulated genes are shown in Tables 3 and 4, respectively. All these patterns appeared in all 20 most downregulated or upregulated genes, respectively.
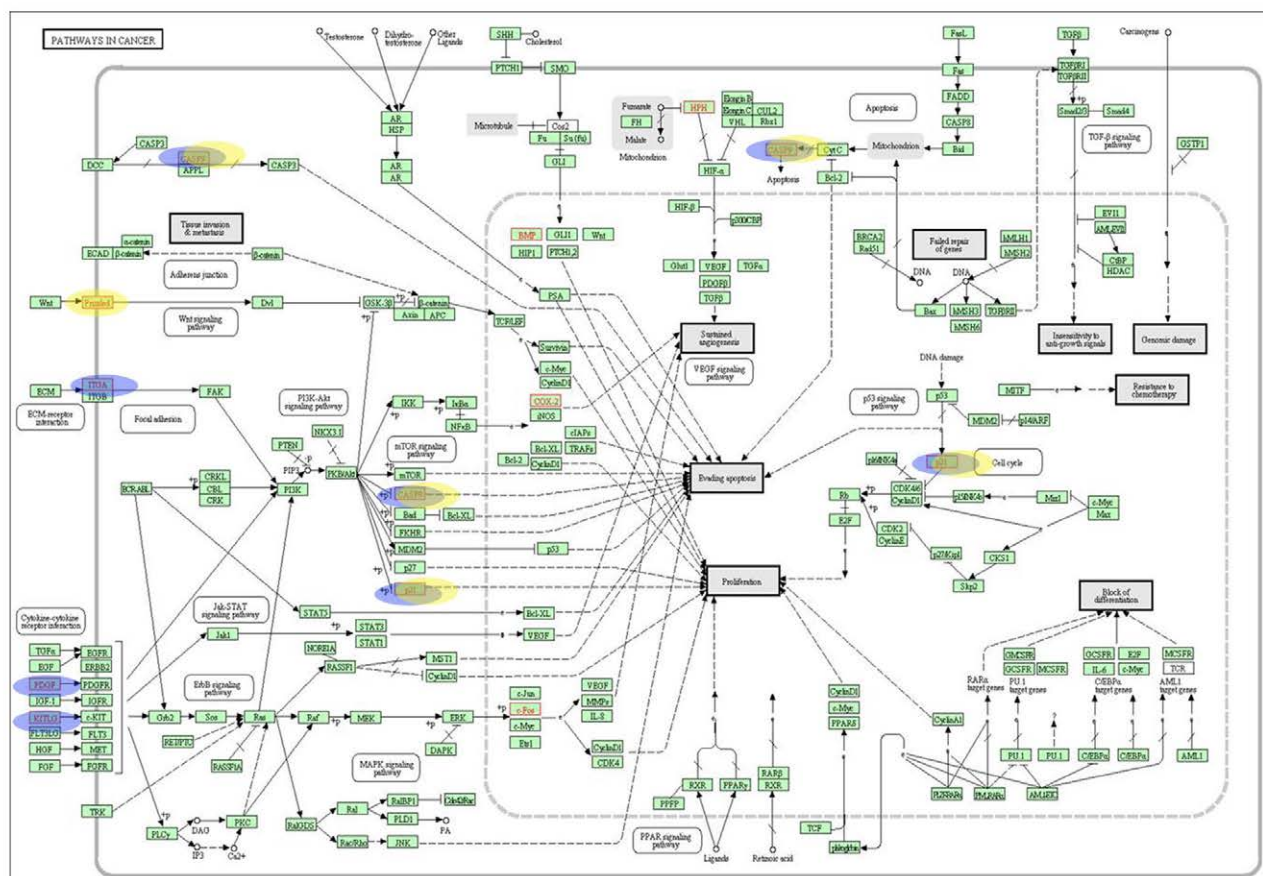
Subsequently, these significant patterns were then screened using the JASPAR database to identify potential transcription factors (TFs). Five unique patterns, corresponding to seven transcription factors (MNB1A, DOF2, PBF, YY1, PRRX2, UBX and ARNT-AHR), were found in the downregulated gene set (Figure 4). On the other hand, two unique patterns corresponding to five transcription factors (GATA2, ZNF42, deltaEF1, MYCN and ARNT) were found in the upregulated gene set (Figure 4). These results suggest these transcription factors might be responsible for the transcriptional regulation of the downregulated and upregulated DEGs.
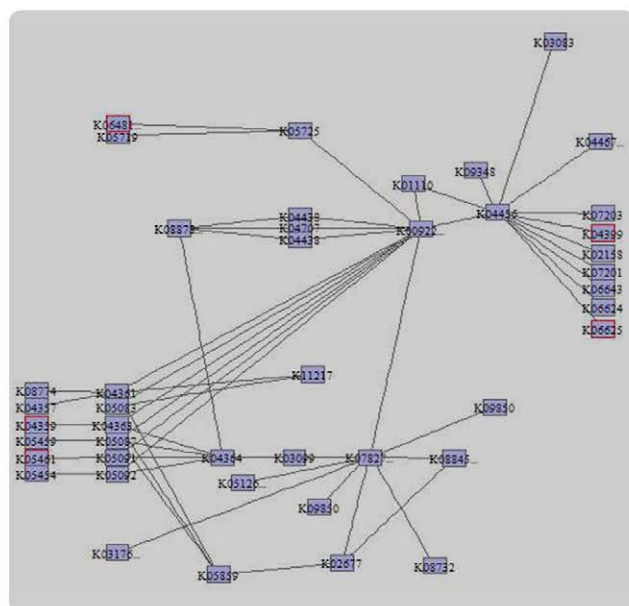
Discussion

Esophageal cancer is the eighth most common cancer and the sixth most common cause of cancer deaths worldwide, with esophageal squamous cell carcinoma (ESCC) being the dominant type in East Asia [16]. Previous research indicates Ezrin plays critical roles in the invasion and metastasis of ESCC [8-10]. In this study, we present a comprehensive analysis of mRNA profile following Ezrin knockdown in ESCC EC109 cell line by using various bioinformatic methods to reveal the potential mechanisms of Ezrin action.

The 244 DEGs were initially analyzed using GO enrichment tools. Ezrin protein interacts with both the plasma membrane and filamentous actin. Its change might cause the rearrangement of cytoskeleton, which might explain that many significant Biological process terms related to immunity and response to stimulus are found. Cellular component enrichment identified four genes involved in extracellular matrix binding (CYR61, DCN, SMOC1 and OLFML2B). These genes have been reported to play important roles in cancer cell migration and metastasis. CYR61 is a potential target downstream of the TGF-β/p38 MAPK pathway to regulate cell migration that plays a key role in the metastasis of osteosarcoma cells [17,18]. Decorin (DCN) is a candidate biomarker in primary breast cancer patients and plays important roles in the regulation of the tumor microenvironment and pathways related to tumorigenesis [19]. On the other hand, as many as 70 signal transduction genes and other genes regulating cell growth and cell death are also enriched, suggesting the intracellular signal transduction, growth, and death are significantly impacted after Ezrin knockdown. These results are consistent with our previous report [10].
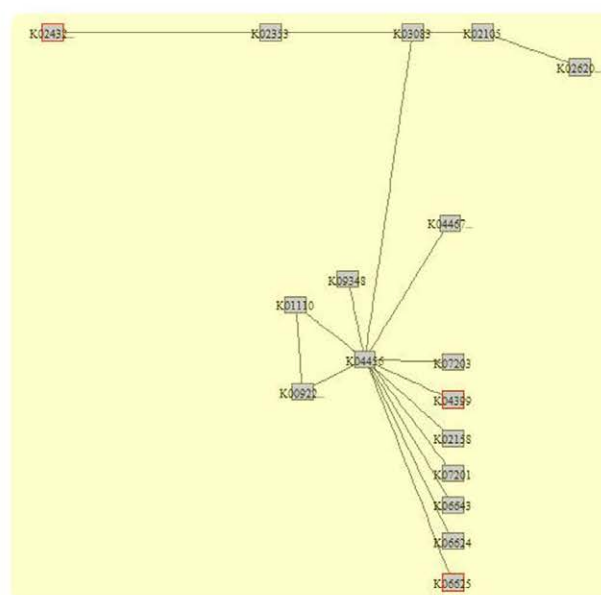
To understand the functional classification of these DEGs to a greater extern, the DEGs were also clustered by Functional Annotation Chart and presented by visualization. Fifty-two genes are enriched in the term of "Signal", which is consistent with the GO enrichment result. Several findings indicate that Ezrin alters gene transcription. The

path 05200

path 05200_4

path 05200_20

Figure 3: Enriched differentially expressed gene subpathways from Pathway in cancer (path:05200). The two significant subpathways, path: 05200_4 and 05200_20, are indicated by the blue and yellow shapes, respectively. The subpathway structures of path: 05200_4 and path: 05200_20 generated by Subpathway package are also shown, with the differentially expressed genes indicated by a red frame.

PATTERN	OCC1 (#PRO / #TOT)	OCC2 (#PRO / #TOT)	FSCORE	p-value
TNCNCNGA	75(20/20)	35(18/20)	3894.25	1.76E-04
TNNTCNGA	87(20/20)	45(17/20)	3414.27	5.71E-04
CCCNNTA	70(20/20)	34(18/20)	3100.98	1.23E-03
ATANNNGG	58(20/20)	27(16/20)	3047.55	1.40E-03
GNNTGTNT	203(20/20)	66(19/20)	2728.31	3.07E-03
GCCNTNT	82(20/20)	45(19/20)	2719.77	3.13E-03
TNCNNTGT	100(20/20)	58(19/20)	2659.46	3.63E-03
GNNTGNNT	434(20/20)	277(20/20)	2480.68	5.63E-03
ATNNNGCA	76(20/20)	41(15/20)	2397.1	6.92E-03
AANGCC	78(20/20)	44(20/20)	2368.59	7.42E-03
TANNNCCC	62(20/20)	32(17/20)	2329.17	8.17E-03
ANNNATCT	113(20/20)	62(18/20)	2266.08	9.54E-03
ATNNGGC	61(20/20)	33(15/20)	2256.65	9.77E-03
TNNGNNTG	409(20/20)	267(20/20)	2226.36	1.05E-02
TGCTGNNG	39(20/20)	18(12/20)	2125.04	1.35E-02
ANGNTTNA	100(20/20)	52(15/20)	2120.31	1.37E-02
ACANNNNC	432(20/20)	284(20/20)	2106.46	1.41E-02
CNNACNCA	173(20/20)	60(20/20)	2082.1	1.50E-02
GGTNGNNA	73(20/20)	41(17/20)	2075.39	1.53E-02
GATNGG	61(20/20)	30(11/20)	2065.88	1.56E-02

Note:
1. OCC: total number of patterns within the sequences of the corresponding cluster (OCC1 represents the downregulated DEG promoter sequence set; OCC2 represents the upregulated DEG promoter sequence set)
2. PRO: total number of sequences with the pattern in the corresponding cluster
3. TOT: total number of sequences in the corresponding cluster
4. FSCORE: results of the ANOVA between the two input clusters and the background sequence set)

Table 3. Sequence patterns over-represented in the downregulated genes, but under-represented in the upregulated genes.

PATTERN	OCC1 (#PRO / #TOT)	OCC2 (#PRO / #TOT)	FSCORE	p-value
CNTNTGNA	53(19/20)	102(20/20)	3267.99	1.67E-03
TTCTNCT	45(18/20)	78(20/20)	3257.9	1.70E-03
CTTNNGNT	57(20/20)	96(20/20)	3102.72	2.41E-03
CAGNAGNA	37(17/20)	68(20/20)	2951.51	3.38E-03
CNAANNGC	59(20/20)	90(20/20)	2810.07	4.64E-03
CTNNNCAA	85(20/20)	125(20/20)	2784.7	4.91E-03
TNAGCNA	55(16/20)	104(20/20)	2772.61	5.04E-03
CTTNTG	84(19/20)	131(20/20)	2747.18	5.34E-03
GNAACNNA	56(20/20)	94(20/20)	2706.67	5.85E-03
TANCCA	59(18/20)	97(20/20)	2635.22	6.86E-03
TNGNNNAG	268(20/20)	340(20/20)	2623.85	7.04E-03
CNGNCTA	34(15/20)	70(20/20)	2605.81	7.33E-03
TNCNNCT	373(20/20)	465(20/20)	2564.07	8.04E-03
TNNGCNA	211(20/20)	292(20/20)	2525.51	8.77E-03
AGANGNA	102(20/20)	150(20/20)	2474.9	9.82E-03
CATTTG	21(12/20)	43(20/20)	2408.01	1.14E-02
CACNTT	88(20/20)	127(20/20)	2392.7	1.18E-02
GNNCTTNC	62(19/20)	100(20/20)	2370.31	1.24E-02
TNCNTNTG	80(20/20)	126(20/20)	2339.02	1.33E-02
TAGNNNGA	42(18/20)	70(20/20)	2333	1.35E-02

Note:
1. OCC: total number of patterns within the sequences of the corresponding cluster (OCC1 represents the downregulated DEG promoter sequence set; OCC2 represents the upregulated DEG promoter sequence set)
2. PRO: total number of sequences with the pattern in the corresponding cluster
3. TOT: total number of sequences in the corresponding cluster
4. FSCORE: result of the ANOVA between the two input clusters and the background sequence set)

Table 4. Sequence patterns over-represented in the upregulated genes, but under-represented in the downregulation genes.

enriched terms of "DNA binding" and "DNA-binding region: Basic motif" containing 13 unique genes indicate that Ezrin regulates gene transcription following the alteration of signal transduction. Of these, adenylate kinase-4 is a marker for poor clinical outcome and promotes metastasis in lung cancer by downregulating the transcription factor ATF3 [20]. Though Ezrin is a membrane-actin cross-linking protein, its translocation to cytosol has been found in multiple cancers, which is considered to be involved in cell motility, invasion, and carcinoma metastasis [21-23]. These relevant findings strongly indicate the ability of Ezrin in the regulation of gene expression. Consistent with these
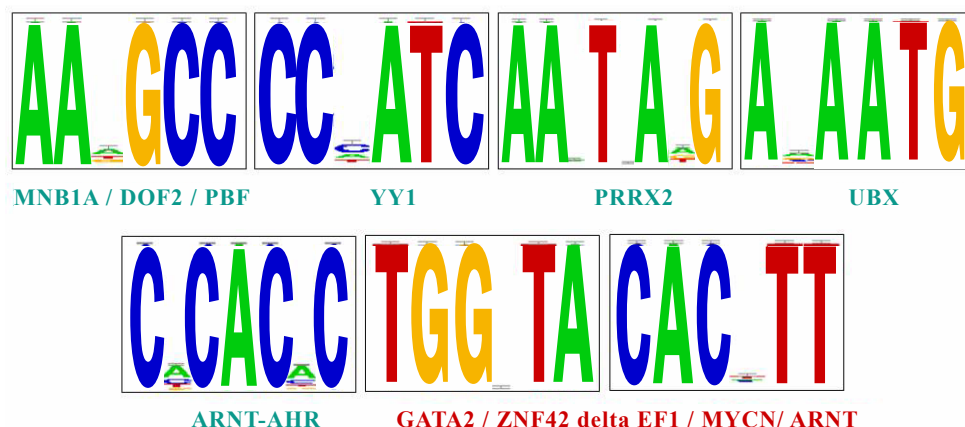
Figure 4: Motif logos of the predicted transcription factors. Potential transcription factors were identified by sequence patterns from searching the JASPAR database. Potential transcription factors regulating downregulated genes are labeled in blue, while transcription factors regulating downregulated genes are labeled in red.

findings, our microarray results also suggest the great impact of Ezrin on the gene expression regulation by the regulatory cascade from the signal transduction to DNA binding. These chart results would provide an overview of the biological impact of Ezrin knockdown in ESCC.

The subpathway mining algorithm in SubpathwayMiner has been successful in identifying disease-related subpathways of biological significance [24,25]. One of the advantages of the subpathway method is that it identifies more pathways than the traditional pathway enrichment. Using SubpathwayMiner, we found several metabolism-related subpathways for the DEGs in this study, supporting recent findings about the relationship between Ezrin and cellular metabolism [26]. Ren et al. found differentially expressed genes that obtained from mutant Ezrin overexpression in osteosarcoma were functionally linked to carbohydrate and amino acid metabolism. In particular, cells expressing closed, inactive Ezrin exhibited reduced lactate production and basal or ATP-dependent oxygen consumption [26]. It suggests that the correlation between Ezrin and metabolism might be a common phenomenon in different cell type of cancers. The second advantage of the subpathway method is it provides more detailed knowledge about the local structure of interested pathways. In this study, two subpathways from Pathway in Cancer indicate the information flow of DEGs in the pathway. The relationship between Ezrin and DEGs in the subpathways shown in Figure 3 is interesting. Ezrin is novel interacting protein of CASP9 in lung cancer [27]. PDGF is able to induce phosphorylation of Ezrin protein [28].

To find biologically significant putative regulatory patterns for the co-expressed DEGs from Ezrin knockdown, the promoters of upregulated and downregulated DEGs were analyzed using the POCO program to find distinguishing sequence patterns and potential transcription factors. The results show MNB1A, DOF2, PBF, YY1, PRRX2, UBX and ARNT-AHR are unique for the downregulated DEGs promoters, while GATA2, ZNF42, deltaEF1, MYCN and ARNT are unique for the upregulated DEG promoters. The transcription factor Yin Yang (YY) 1 is overexpressed in most cancers, regulating multiple cancer-related proteins and signaling pathways, and overexpression of YY1 both at mRNA and protein levels in ESCC tissue has been confirmed by Luo et al. [29-31]. For other transcription factors that might respond to the upregulated DEGs, overexpression of Snail occurs in ESCC and correlates positively with lymphovascular invasion and was associated with worse overall survival [32]. GATA factors are kind of zinc finger DNA binding proteins, regulating the

specification and differentiation of numerous tissues by activating or repressing transcription [33]. GATA2 promotes breast cancer cell growth by directly repressing transcription of PTEN tumor suppressor gene [34]. DeltaEF1, also called ZEB1, is focally expressed in invasive ESCC [35]. These suggests these transcription factors are not only critical in the regulation of upregulated and downregulated DEGs by Ezrin knockdown, but might participate in the invasion or metastasis mediated by Ezrin.

In summary, a comprehensive understanding the molecular mechanisms underlying Ezrin in ESCC after its knockdown was obtained by bioinformatics analyses, especially by the means of subpathway and sequence patterns for co-expression, which provide more clues than the traditional methods. These analytical methods are suited for other kinds of high-throughput data that are applicable to search novel functional genes and pathways related to gene(s) of interest.

Acknowledgements

Competing interests

The authors declare that they have no competing interest.

Submission declaration

This paper has not been published or submitted elsewhere.

References

1. Moleirinho S, Tilston-Lunel A, Angus L, Gunn-Moore F, Reynolds PA (2013) The expanding family of FERM proteins. Biochem J 452: 183-193.

2. Khanna C, Wan X, Bose S, Cassaday R, Olomu O, et al. (2004) The membrane-cytoskeleton linker ezrin is necessary for osteosarcoma metastasis. Nat Med 10: 182-186.

3. Yu Y, Khan J, Khanna C, Helman L, Meltzer PS, et al. (2004) Expression profiling identifies the cytoskeletal organizer ezrin and the developmental homeoprotein Six-1 as key metastatic regulators. Nat Med 10: 175-181.

4. Yu H, Zhang Y, Ye L, Jiang WG (2011) The FERM family proteins in cancer invasion and metastasis. Front Biosci (Landmark Ed) 16: 1536-1550.

5. Batchelor CL, Woodward AM, Crouch DH (2004) Nuclear ERM (ezrin, moesin) proteins: regulation by cell density and nuclear import. Exp Cell Res 296: 208-222.

6. Kishore R, Qin G, Luedemann C, Bord E, Hanley A, et al. (2005) The cytoskeletal protein ezrin regulates EC proliferation and angiogenesis via TNF-alpha-induced transcriptional repression of cyclin A. J Clin Invest 115: 1785-1796.

7. Adyshev DM, Moldobaeva NK, Elangovan VR, Garcia JG, Dudek SM (2011) Differential involvement of ezrin/radixin/moesin proteins in sphingosine 1-phosphate-induced human pulmonary endothelial cell barrier enhancement. Cell Signal 23: 2086-2096.

8. Shen ZY, Xu LY, Li EM, Li JT, Chen MH, et al. (2003) Ezrin, actin and cytoskeleton in apoptosis of esophageal epithelial cells induced by arsenic trioxide. Int J Mol Med 12: 341-347.

9. Zeng H, Xu L, Xiao D, Zhang H, Wu X, et al. (2006) Altered expression of ezrin in esophageal squamous cell carcinoma. J Histochem Cytochem 54: 889-896.

10. Xie JJ, Xu LY, Xie YM, Zhang HH, Cai WJ, et al. (2009) Roles of ezrin in the growth and invasiveness of esophageal squamous carcinoma cells. Int J Cancer 124: 2549-2558.

11. Zhang B, Kirov S, Snoddy J (2005) WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res 33: W741-748.

12. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57.

13. Merico D, Isserlin R, Stueker O, Emili A, Bader GD (2010) Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One 5: e13984.

14. Li C, Li X, Miao Y, Wang Q, Jiang W, et al. (2009) SubpathwayMiner: a software package for flexible identification of pathways. Nucleic Acids Res 37: e131.

15. Kankainen M, Holm L (2005) POCO: discovery of regulatory patterns from promoters of oppositely expressed gene sets. Nucleic Acids Res 33: W427-431.

16. Jemal A, Bray F, Center MM, Ferlay J, Ward E, et al. (2011) Global cancer statistics. CA Cancer J Clin 61: 69-90.

17. Chen J, Song Y, Yang J, Gong L, Zhao P, et al. (2013) The up-regulation of cysteine-rich protein 61 induced by transforming growth factor beta enhances osteosarcoma cell migration. Mol Cell Biochem 384: 269-277.

18. Fromigue O, Hamidouche Z, Vaudin P, Lecanda F, Patino A, et al. (2011) CYR61 downregulation reduces osteosarcoma cell invasion, migration, and metastasis. J Bone Miner Res 26: 1533-1542.

19. Cawthorn TR, Moreno JC, Dharsee M, Tran-Thanh D, Ackloo S, et al. (2012) Proteomic analyses reveal high expression of decorin and endoplasmin (HSP90B1) are associated with breast cancer metastasis and decreased survival. PLoS One 7: e30992.

20. Jan YH, Tsai HY, Yang CJ, Huang MS, Yang YF, et al. (2012) Adenylate kinase-4 is a marker of poor clinical outcomes that promotes metastasis of lung cancer by downregulating the transcription factor ATF3. Cancer Res 72: 5119-5129.

21. Jiang WG, Hiscox S, Singhrao SK, Puntis MC, Nakamura T, et al. (1995) Induction of tyrosine phosphorylation and translocation of ezrin by hepatocyte growth factor/scatter factor. Biochem Biophys Res Commun 217: 1062-1069.

22. Chen Z, Fadiel A, Feng Y, Ohtani K, Rutherford T, et al. (2001) Ovarian epithelial carcinoma tyrosine phosphorylation, cell proliferation, and ezrin translocation are stimulated by interleukin 1alpha and epidermal growth factor. Cancer 92: 3068-3075.

23. Briggs JW, Ren L, Nguyen R, Chakrabarti K, Cassavaugh J, et al. (2012) The ezrin metastatic phenotype is associated with the initiation of protein translation. Neoplasia 14: 297-310.

24. Qiao WJ, Cheng HY, Li CQ, Jin H, Yang SS, et al. (2011) Identification of pathways involved in paclitaxel activity in cervical cancer. Asian Pac J Cancer Prev 12: 99-102.

25. Li C, Han J, Shang D, Li J, Wang Y, et al. (2012) Identifying disease related sub-pathways for analysis of genome-wide association studies. Gene 503: 101-109.

26. Ren L, Hong SH, Chen QR, Briggs J, Cassavaugh J, et al. (2012) Dysregulation of ezrin phosphorylation prevents metastasis and alters cellular metabolism in osteosarcoma. Cancer Res 72: 1001-1012.

27. CheciÅ„ska A, Giaccone G, Rodriguez JA, Kruyt FA, Jimenez CR (2009) Comparative proteomics analysis of caspase-9-protein complexes in untreated and cytochrome c/dATP stimulated lysates of NSCLC cells. J Proteomics 72: 575-585.

28. Baumgartner M, Sillman AL, Blackwood EM, Srivastava J, Madson N, et al. (2006) The Nck-interacting kinase phosphorylates ERM proteins for formation of lamellipodium by growth factors. Proc Natl Acad Sci U S A 103: 13391-13396.

29. Nicholson S, Whitehouse H, Naidoo K, Byers RJ (2011) Yin Yang 1 in human cancer. Crit Rev Oncog 16: 245-260.

30. Zhang Q, Stovall DB, Inoue K, Sui G (2011) The oncogenic role of Yin Yang 1. Crit Rev Oncog 16: 163-197.

31. Luo J, Zhou X, Ge X, Liu P, Cao J, et al. (2013) Upregulation of Ying Yang 1 (YY1) suppresses esophageal squamous cell carcinoma development through heme oxygenase-1. Cancer Sci 104: 1544-1551.

32. Kuo KT, Chou TY, Hsu HS, Chen WL, Wang LS (2012) Prognostic significance of NBS1 and Snail expression in esophageal squamous cell carcinoma. Ann Surg Oncol 19 Suppl 3: S549-557.

33. Zheng R, Blobel GA (2010) GATA Transcription Factors and Cancer. Genes Cancer 1: 1178-1188.

34. Wang Y, He X, Ngeow J, Eng C (2012) GATA2 negatively regulates PTEN by preventing nuclear translocation of androgen receptor and by androgen-independent suppression of PTEN transcription in breast cancer. Hum Mol Genet 21: 569-576.

35. Ohashi S, Natsuizaka M, Naganuma S, Kagawa S, Kimura S, et al. (2011) A NOTCH3-mediated squamous cell differentiation program limits expansion of EMT-competent cells that express the ZEB transcription factors. Cancer Res 71: 6836-6847.