

Construction of differential mRNA-lncRNA crosstalk networks based on ceRNA hypothesis uncover key roles of lncRNAs implicated in esophageal squamous cell carcinoma

Shuang Yang^{1,3,*}, Qianqian Ning^{1,3,*}, Guobin Zhang², Hong Sun⁴, Zhen Wang² and Yixue Li^{1,2,3,5}

¹ Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

² Key Lab of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

³ Shanghai Center for Bioinformation Technology, Shanghai, China

⁴ Biomedical Information Research Center, Children's Hospital of Shanghai

⁵ Collaborative Innovation Center for Genetics and Development, Fudan University, Shanghai, China

* These authors have contributed equally to this work

Correspondence to: Hong Sun, **email:** sunhong@scbif.org

Zhen Wang, **email:** zwang01@sibs.ac.cn

Yixue Li, **email:** yxli@sibs.ac.cn

Keywords: intrinsic subtype; competing endogenous RNA network; biomarker; lncRNA; ESCC; Pathology Section

Received: September 24, 2016

Accepted: November 18, 2016

Published: December 09, 2016

ABSTRACT

Increasing evidence has indicated that lncRNAs acting as competing endogenous RNAs (ceRNAs) play crucial roles in tumorigenesis, metastasis and diagnosis of cancer. However, the function of lncRNAs as ceRNAs involved in esophageal squamous cell carcinoma (ESCC) is still largely unknown. In this study, clinical implications of two intrinsic subtypes of ESCC were identified based on expression profiles of lncRNA and mRNA. ESCC subtype-specific differential co-expression networks between mRNAs and lncRNAs were constructed to reveal dynamic changes of their crosstalks mediated by miRNAs during tumorigenesis. Several well-known cancer-associated lncRNAs as the hubs of the two networks were firstly proposed in ESCC. Based on the ceRNA mechanism, we illustrated that the "loss" of miR-186-mediated *PVT1*-mRNA and miR-26b-mediated *LINC00240*-mRNA crosstalks were related to the two ESCC subtypes respectively. In addition, crosstalks between *LINC00152* and *EGFR*, *LINC00240* and *LOX* gene family were identified, which were associated with the function of "response to wounding" and "extracellular matrix-receptor interaction". Furthermore, functional cooperation of multiple lncRNAs was discovered in the two differential mRNA-lncRNA crosstalk networks. These together systematically uncovered the roles of lncRNAs as ceRNAs implicated in ESCC.

INTRODUCTION

In the past decade, research on the non-coding RNA has gained widespread attention. lncRNAs are a large class of non-coding RNAs, which are longer than 200 nucleotides and pervasively transcribed in the genome. Currently, about 15,767 lncRNAs have been annotated in the human genome (GENCODE 24). lncRNAs play important roles in chromatin remodeling, transcriptional

repression and post-transcriptional regulation [1]. Dysregulation of lncRNAs are associated with different types of cancers, and lncRNAs as reliable biomarkers for cancers are also identified [2-4]. However, due to functional diversity of lncRNAs, identification of cancer-related lncRNAs still faces challenges [5].

Esophageal squamous cell carcinoma (ESCC) as one of main forms of esophageal cancer is a highly aggressive solid tumor with poor prognosis and widely occurs in Asian countries [6]. Despite advances in the

various diagnosis and treatment, the 5-year survival rate of ESCC remains only approximately 30% for patients after surgery [7]. Recent whole-genome and whole-exome sequencing on ESCC patients revealed six well-known tumor-associated genes (*TP53*, *CDKN2A* etc.) and two novel oncogenes (*ADAM29* and *FAM135B*) [8]. In non-coding RNA research, miRNAs (*miR-20* and *miR-21* etc.) and lncRNAs (*HOTAIR* and *H19* etc.) as oncogenes or tumor suppressors were also studied in the development of ESCC [9-11]. In addition, using computational methods to predict lncRNA functions in ESCC has also been reported [12]. Even so, compared to coding genes and miRNAs, dysregulated lncRNAs implicated in ESCC remain little known.

Recently, the concept of competing endogenous RNAs (ceRNAs) indicates that RNA molecules harboring miRNA response elements (MREs) can communicate with each other by competing for common miRNA [13]. Especially, large-scale cross-linking immunoprecipitation (CLIP) experiments have identified thousands of miRNA-lncRNA interactions, which imply that miRNA-mediated crosstalks between mRNAs and lncRNAs widely exist in various biological processes [14]. Therefore, ceRNA hypothesis provides a new perspective to account for the function of as yet uncharacterized lncRNAs involved in ESCC.

A recent study discussed three scenarios of identifying cancer-related lncRNAs based on ceRNA network, including the “loss” and “gain” of miRNA-mediated mRNA-lncRNA crosstalks, and no significant change in mRNA-lncRNA crosstalks but their expression levels change in opposite directions [15]. However, most studies were related with the third class of dysregulation [16-19], and dynamic changes of the crosstalks between mRNAs and lncRNAs during tumorigenesis have never been studied. In addition, cancer subtype-specific ceRNA network was rarely considered before. In the present study, mRNA and lncRNA expression patterns of ESCC were identified, laying a foundation for discovering subtype-specific mRNA-lncRNA crosstalks. Furthermore, ESCC subtype-specific differential mRNA-lncRNA crosstalk networks mediated by miRNAs were constructed to compare lncRNA-related ceRNA networks in ESCC tissues and adjacent normal tissues. Our study uncovered ESCC-associated lncRNAs and shed light on mRNA-lncRNA crosstalks involved in tumorigenesis of ESCC.

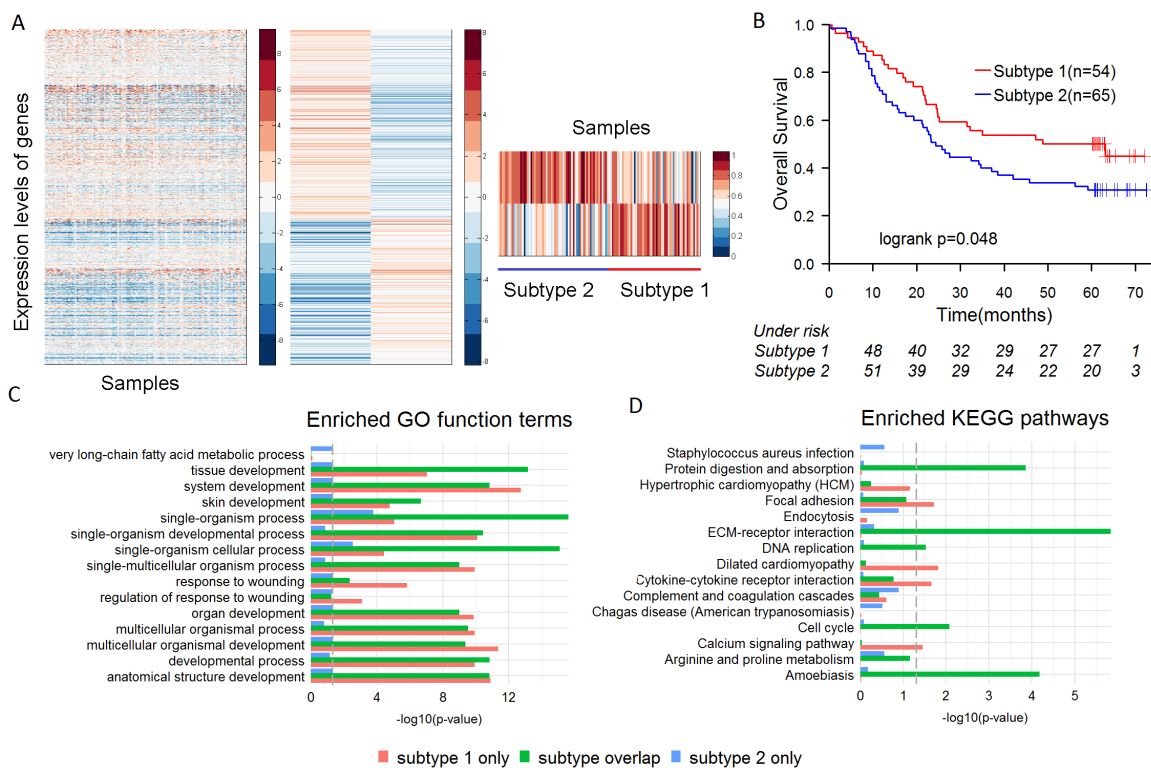


Figure 1: The subtypes of ESCC and their characteristics. A. semi-NMF clustering identifies two subtypes of ESCC based on expression levels of lncRNAs and mRNAs. B. Overall survival of different subtypes of ESCC by Kaplan-Meier plot. ESCC subtypes are highly correlated with overall survival outcomes. C.-D. Differential expressed genes (DEGs) in two subtypes were divided into three groups, including the DEGs only in the subtype 1, the DEGs only in the subtype 2 and the DEGs in both subtypes. Significantly enriched GO terms and KEGG pathways of each group are shown.

RESULTS

Characteristics of subtypes of ESCC

We performed an unsupervised hierarchical clustering analysis based on the expression of mRNAs and lncRNAs of 119 ESCC patients. Semi-NMF analysis identified 2 intrinsic subtypes of ESCC, which contained 54 and 65 samples respectively (Figure 1A). Analysis of the survival prognosis after surgery showed that the overall survival time of patients in the subtype 1 was significantly longer than patients in the subtype 2 (median survival 41.4 vs. 33.5 months, $p = 0.048$) (Figure 1B). Fisher's exact test revealed the subtypes are significantly correlated with tumor grade ($p = 3.55e-05$), but not with other clinical features (Table S1). Patients with poorly differentiated ESCC were more prevalent in the subtype 2 than 1 (43.1% vs. 7.4%) (Table S2). In addition, cox

regression analysis indicated that the subtypes of ESCC (HR = 1.601, 95% CI 0.998 to 2.569, $p = 0.048$) were more significantly correlated with overall survival of the patients than tumor grade (HR = 1.162, 95% CI 0.6264 to 2.157, $p = 0.222$).

Then, 3,036 and 3,551 differentially expressed coding genes were respectively identified in the subtype 1 and subtype 2. The differentially expressed genes (DEGs) were divided into three groups, including the DEGs only in the subtype 1 (849 genes), the DEGs only in the subtype 2 (1,364 genes) and the DEGs in both subtypes (2,187 genes). Enrichment analysis on each group of DEGs was performed based on GO terms and KEGG pathways. The DEGs in both subtypes were more significantly enriched for cancer-related pathways (such as "Cell cycle", "DNA replication" and "ECM-receptor interaction" etc.) than the other two groups (Figure 1C, 1D). Nevertheless, the DEGs in the subtype 1 were more significantly enriched for the function of response to wounding ("response to wounding" and "regulation of response to wounding")

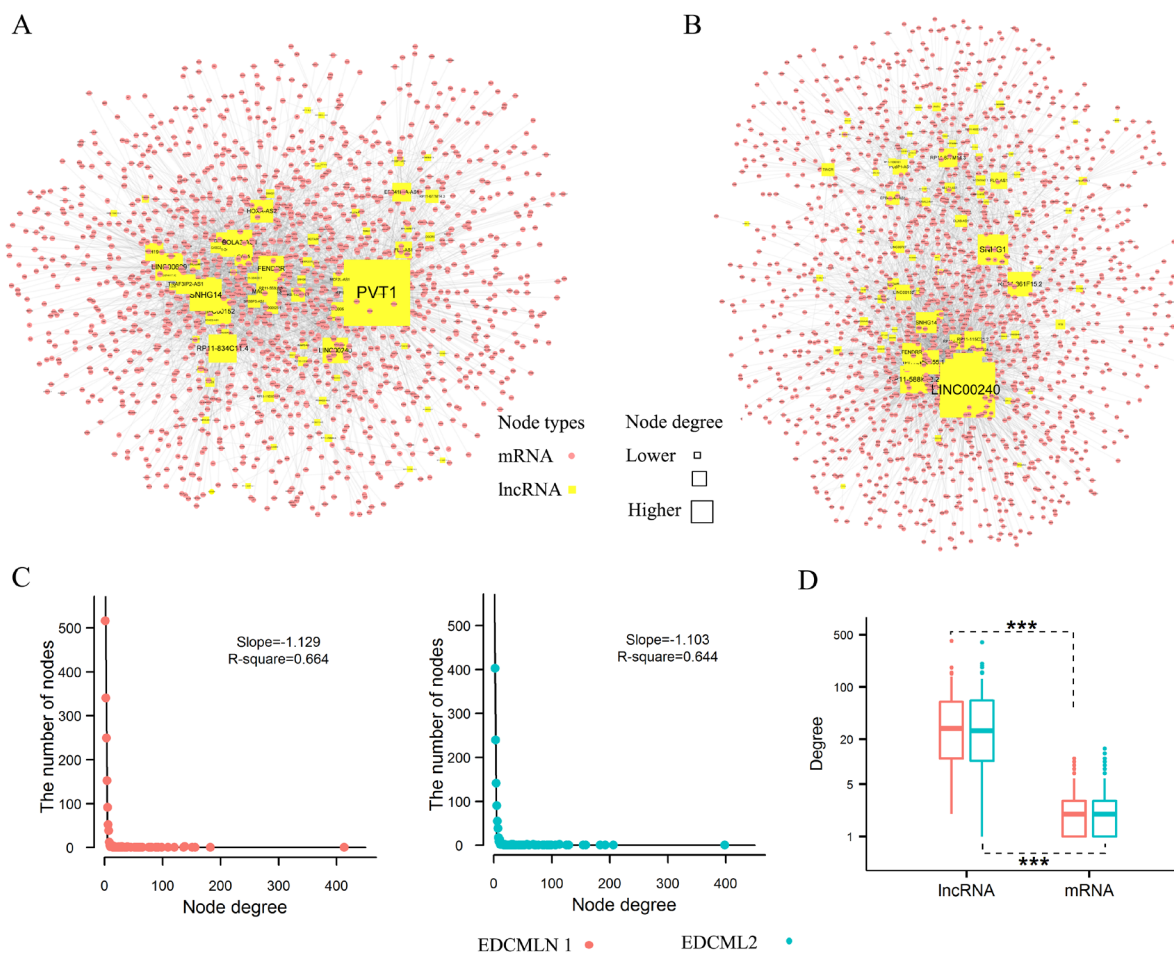


Figure 2: ESCC subtype-specific differential mRNA-lncRNA crosstalk networks based on ceRNA hypothesis and their properties. **A.** Global view of the differential mRNA-lncRNA crosstalk network 1, which consists of 1,524 nodes and 3,699 links. **B.** Global view of the differential mRNA-lncRNA crosstalk network 2, which consists of 1,774 nodes and 4,131 links. **C.** Degree distribution of two differential mRNA-lncRNA crosstalk networks. **D.** The difference between degree of lncRNAs and mRNAs nodes. Wilcoxon test assessed the different significance.

Table 1: The top 10 genes in betweenness, degree, closeness and DCLs enrichment of differential mRNA-lncRNA crosstalk networks

	Betweenness	Degree	Closeness	DCLs enrichment	Overlaps
Network 1	PVTI	PVTI	PVTI	PVTI	PVTI
	<i>SNHG14</i>	<i>SNHG14</i>	<i>SNHG14</i>	LINC00152	<i>SNHG14</i>
	<i>RP11-834C11.4</i>	<i>RP11-834C11.4</i>	<i>SLC16A14</i>	<i>LINC00689</i>	<i>RP11-834C11.4</i>
	<i>LINC00240</i>	<i>LINC00689</i>	<i>RP11-834C11.4</i>	<i>RP11-834C11.4</i>	<i>LINC00240</i>
	LINC00152	LINC00152	<i>LINC00240</i>	<i>SNHG14</i>	LINC00152
	<i>BOLA3-AS1</i>	FENDRR	LINC00152	FENDRR	FENDRR
	<i>LINC00689</i>	<i>LINC00240</i>	<i>PPFIA1</i>	<i>BOLA3-AS1</i>	<i>BOLA3-AS1</i>
	FENDRR	<i>BOLA3-AS1</i>	FENDRR	HOXA-AS2	
	HOXA-AS2	HOXA-AS2	<i>BOLA3-AS1</i>	H19	
<i>EPB41L4A-AS1</i>	<i>MAGI2-AS3</i>	<i>IKBIP</i>	<i>LINC00240</i>		
Network 2	<i>LINC00240</i>	<i>LINC00240</i>	<i>LINC00240</i>	<i>LINC00240</i>	<i>LINC00240</i>
	SNHG1	SNHG1	<i>SEMA6D</i>	<i>AP000525.9</i>	SNHG1
	<i>RP11-361F15.2</i>	<i>AP000525.9</i>	SNHG1	<i>AC156455.1</i>	<i>RP11-361F15.2</i>
	<i>AP000525.9</i>	<i>RP11-588K22.2</i>	<i>RP11-588K22.2</i>	<i>RP11-588K22.2</i>	<i>AP000525.9</i>
	<i>RP11-588K22.2</i>	<i>RP11-361F15.2</i>	<i>PCDH19</i>	<i>RP11-677M14.3</i>	<i>RP11-588K22.2</i>
	<i>RP11-677M14.3</i>	<i>AC156455.1</i>	<i>AP000525.9</i>	<i>RP11-361F15.2</i>	<i>AC156455.1</i>
	<i>SNHG14</i>	<i>SNHG14</i>	<i>PTPDC1</i>	SNHG1	
	<i>AC156455.1</i>	<i>AC009948.5</i>	<i>AC156455.1</i>	<i>FLG-AS1</i>	
	<i>FLG-AS1</i>	FENDRR	<i>RP11-361F15.2</i>	<i>RP11-115C21.2</i>	
<i>AC009948.5</i>	<i>RP11-677M14.3</i>	<i>MET</i>	<i>TINCR</i>		

Overlaps contains the lncRNAs that appeared in each dimension. Cancer-related lncRNAs were highlighted in bold font.

than the other two groups (Figure 1C). Many studies have shown that biological process of response to wounding was closely related to ESCC [20, 21]. High-level expression of epidermal growth factor receptor (*EGFR*), as the main positive regulator involved in wound re-epithelialization, is associated with well-differentiated ESCC [21, 22]. Our study revealed that *EGFR* was up-regulated (fold-change = 2.14) in the subtype 1 of ESCC, but not obviously up-regulated (fold-change = 1.59) in the subtype 2, which indicated that *EGFR* might be a potential molecular marker to identify the subtypes of ESCC.

Overall of ESCC subtype-specific differential mRNA-lncRNA crosstalk networks

Based on ceRNA hypothesis, we respectively constructed ESCC subtype-specific differential mRNA-lncRNA crosstalk networks. The networks were constructed by integrating prior knowledge of miRNA-lncRNA interactions [14], miRNA-mRNA interactions [14, 23] and expression profiles, and searching the most significant changes of mRNA-lncRNA crosstalks between ESCC and adjacent normal tissues. There were 1,449 coding genes, 75 lncRNA genes and 3,699 edges (4,786 mRNA-lncRNA differentially co-expressed links (DCLs)) in the network 1 of subtype 1 (Figure 2A), and 1,688 coding genes, 86 lncRNA genes and 4,132 edges (4,868 mRNA-lncRNA DCLs) in the network 2 of subtype

2 (Figures 2B). As observed, the change of crosstalks between mRNAs and lncRNAs formed a scale-free structure typical of transcriptional regulatory networks (Figure 2C). In the two networks, degree of lncRNAs were significantly more than coding genes ($p < 2.2e-16$, Wilcoxon test) (Figure 2D). Moreover, crosstalks between mRNAs and lncRNAs tended to disappear rather than appear in ESCC. 77.1% and 83% DCLs were separately found as the “loss” of crosstalks in the subtype 1 and subtype 2, whereas, only 20.9% and 14.7% DCLs were identified as the “gain” of crosstalks in the two subtypes (Table S3).

In addition, compared to high proportion of overlapping coding genes and lncRNAs, few overlapping edges (mRNA-lncRNA DCLs) in the two networks (Table S3) were found. Further analysis revealed that the difference of crosstalks between mRNAs and lncRNAs also existed in adjacent normal tissues of the two subtypes of ESCC (Table S4). Recent study showed that histologically-normal cancer-adjacent tissue, which influences recurrence risk, could reflect the intrinsic tumor subtypes of breast cancer [24]. So the difference of mRNA-lncRNA crosstalks in adjacent normal tissues might result in different local recurrence rates and prognosis after surgery of two subtypes of ESCC.

The hub nodes of differential mRNA-lncRNA crosstalk networks

Identification of important nodes of differential mRNA-lncRNA crosstalk networks was based on four topological properties (detail in Methods). lncRNAs almost occupied the top 10 of each property and the top-ranked lncRNAs had striking difference between the two subtypes (Table 1). Several cancer-associated lncRNAs were found, which have rarely been reported in ESCC (Table 1). Except that *H19* could promote ESCC cell proliferation and metastasis [25], *PVT1* as a rising star among oncogenic lncRNAs has been reported in plenty of cancers [26-28]. *LINC00152* could directly bind to *EGFR* and activate *EGFR* signaling pathway [29] and *FENDER* could suppress cell invasion and migration by downregulating *FNI* and *MMP2/MMP9* expression

in gastric cancer [30]. *HOXA-AS2* acts as an apoptosis repressor in promyelocytic leukemia cells [31], and *SNHG1* has also been suggested as a new biomarker for lung cancer and prostate cancer [32]. Moreover, cancer-associated lncRNAs (collected by Lnc2Cancer [33]) were mapped to the two networks. We found cancer-associated nodes had significantly higher degrees, betweenness and closeness centrality than other nodes in the network 1 (Figure S1). Thus, all these lncRNAs as the hub nodes of the differential mRNA-lncRNA crosstalk networks may be essential for ESCC development and progression.

To uncover the hub miRNAs likely involved in the two networks, miRNAs were ranked according to their frequency of mediating mRNA-lncRNA crosstalks. 90% of the top 10 miRNAs were disease-associated miRNAs (collected by miR2Disease [34]) (Table S5). Surprisingly, we found well-known *miR-15a/16-1* cluster appeared in

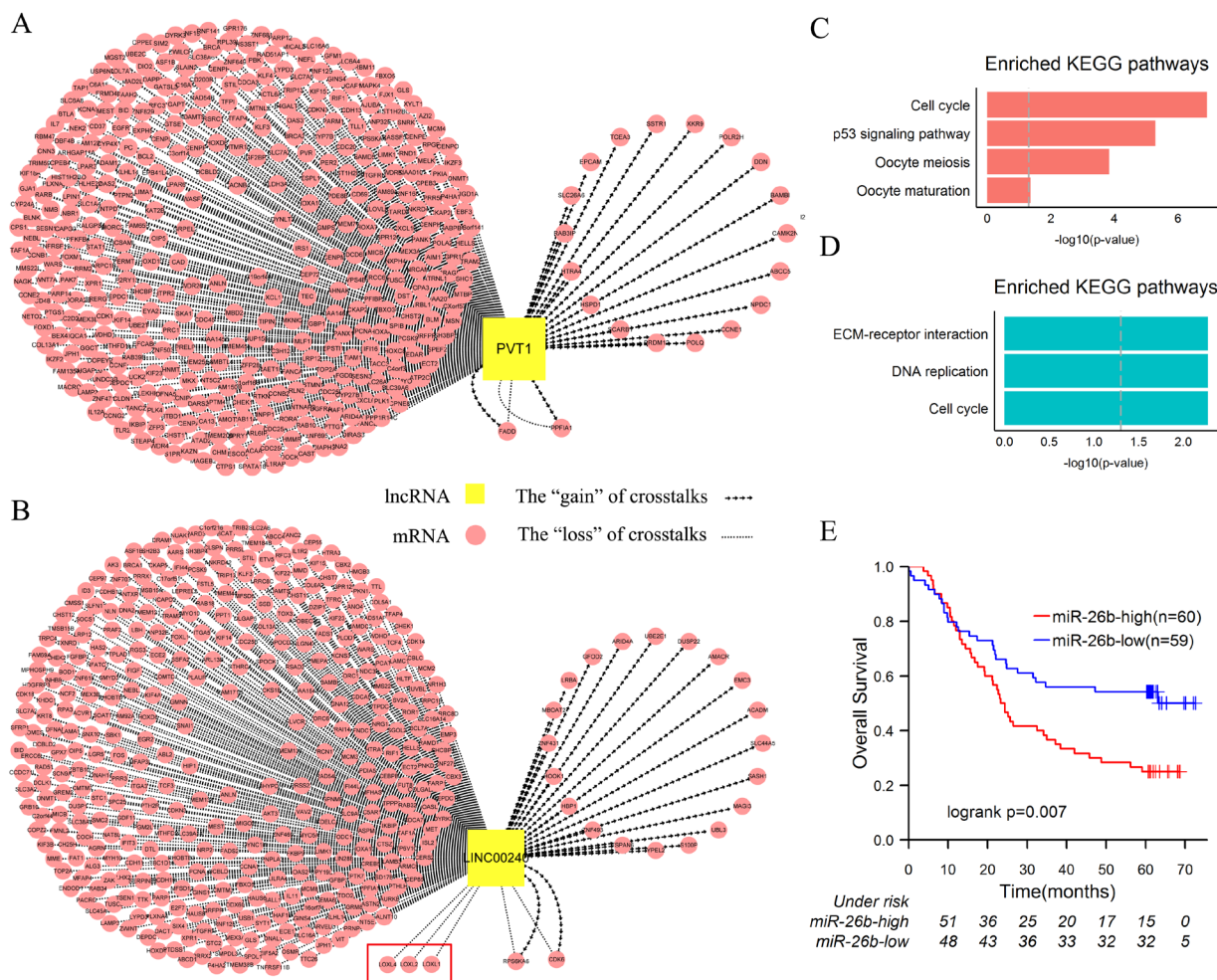


Figure 3: Dysregulation of PVT1 and LINC00240 as ceRNAs in ESCC. A. The sub-network of network 1, including *PVT1* and its first neighbor mRNAs. B. The sub-network of network 2, including *LINC00240* and its first neighbor mRNAs. C. KEGG enrichment analysis of the first neighbor mRNAs of *PVT1*. D. KEGG enrichment analysis of the first neighbor mRNAs of *LINC00240*. E. In the independent dataset, patients were divided into two groups according to median expression level of *miR-26b* (high *miR-26b* expression and low *miR-26b* expression). Kaplan-Meier survival analysis of two groups of patients shows the significantly different clinical outcomes. P values were determined by the log rank test.

the both subtypes. *MiR-15a/16-1* cluster within 0.5 kb at chromosome position 13q14 target several oncogenes (including *BCL2*, *CCNE1* and *CCNB1*, which also existed in the two networks), to suppress cell cycle progression and proliferation in several malignant tumors [35-37]. In addition, we found 2 members (*miR-17-5p*, *miR-20a*) of *miR-17-92* cluster only appeared in the subtype 1. *MiR-17-92* cluster which locates on the *C13orf25* gene is overexpressed in ESCC and usually promote cancer cell proliferation [38]. Then we validated the expression of these two clusters in an independent data set (detail in Methods), and found *miR-17-5p*, *miR-20a* and *miR-16-5p* were upregulated in ESCC. So, dysregulated *miR-17-92* and *miR-15a/16-1* clusters might result in the change of crosstalks between mRNAs and lncRNAs in the pathogenesis of ESCC.

Dysregulation of lncRNAs as ceRNAs in the two subtypes

In order to explicitly explore dysregulation of lncRNAs as ceRNAs involved in the two subtypes of ESCC, we selected *PVT1* and *LINC00240*—the foremost lncRNAs of two subtypes according to topological properties for research (Table 1). Enrichment analysis was firstly performed based on their first neighbor mRNAs in the two networks (Figure 3A, 3B). Two gene sets were significantly enriched for the function of cell cycle and cancer-associated pathways such as “P53 signaling pathway”, “Cell cycle” and “Pathway in cancer” etc. (Figure 3C, 3D, Figure S2). Moreover, 95.2% and 94.5% genes of the two gene sets with *PVT1* and *LINC00240* occurred as the “loss” of crosstalks respectively (Figure 3A, 3B).

Then, we ranked all the miRNAs according to their frequency of mediating the “loss” of crosstalks. In the subtype 1 of ESCC, 38.2% “loss” of *PVT1*-associated crosstalks were likely mediated by *miR-186-5p* (Figure S3). It has been suggested that *miR-186* could inhibit the cell proliferation, migration and invasion of non-small cell lung cancer by modulating pituitary tumor-transforming gene-1 (*PTTG1*) [39]. In ESCC, *miR-186* and *PTTG1* that function as tumor suppressor and oncogene always are downregulated and upregulated respectively [40, 41]. Meanwhile, *miR-186-PVT1* interaction has been verified in several cell lines [14, 42], which implied that *PVT1* function as *miR-186* sponge and crosstalks between *PVT1* and *miRNA-186* target genes exist in various biological processes. In our study, although *PVT1* and *PTTG1* were both overexpressed in ESCC, positive correlation between them disappeared during the development of ESCC (Figure S4A, S4C). Therefore, we speculated that *miRNA-186*-mediated crosstalk between *PVT1* and *PTTG1* existed in adjacent tissues and maintained normal functions, and the “loss” of *PVT1-PTTG1* crosstalk might act as a

driver event implicated in development of the subtype 1 of ESCC. Similar results—the “loss” of *miR-200* family mediated crosstalks between *PVT1*-mRNAs have also been proposed in breast cancer [43].

In the subtype 2 of ESCC, 49.2% “loss” of *LINC00240*-associated crosstalks were possibly mediated by *miR-26b-5p* (Figure S3). Downregulation of *miR-26b-5p* could inhibit proliferation and induce cell-cycle arrest in ESCC [44]. In the independent microarray dataset, we found that expression of *miR-26b-5p* in well-differentiated cancer cells was significantly lower than poorly differentiated ($p = 0.004$, t -test). Furthermore, the patients were divided into two groups according to median expression level of *miR-26b-5p* and survival analysis was performed on the two groups. Patients with the higher expression level of *miR-26b-5p* had significantly shorter overall survival rate than those with the lower expression level (5-year survival rate: 25.0% vs. 52.0%, $p = 0.008$) (Figure 3E). Meanwhile, in the subtype 2, the “loss” of *miR-26b-5p*-mediated *LINC00240*-associated crosstalks accompanied the bad prognosis. So we proposed a bold hypothesis that the increased *LINC00240* are not necessarily binding to *miR-26b-5p* and accordingly overexpressed *miR-26b-5p* as an oncogene result in poor clinical outcome of the subtype 2. In order to verify this assumption, we compared the expression levels of target genes of *miR-26b-5p* in the two subtypes of ESCC samples. Kruppel-Like Factor 3 (Basic) (*KLF3*), which is an important transcriptional repressor involved in cell growth, apoptosis, and angiogenesis [45, 46], was significantly downregulated in the subtype 2 (fold-change = 0.7, q -value = $4.3e-4$). Previous study suggested that low expression of *KLF3* was associated with bad prognosis of uterine cervical cancer [46]. Therefore, the “loss” of *miR-26b-5p*-mediated *LINC00240-KLF3* crosstalk was probably implicated in tumorigenesis of the subtype 2 of ESCC.

When comparing enriched pathways based on neighbor mRNAs of *PVT1* and *LINC00240*, extracellular matrix (ECM)-receptor interaction was the most significantly enriched pathway in the subtype 2 (Figure 3C, 3D). Upregulated *LOX*-mediated ECM remodeling is a poor prognostic marker in breast cancer and head and neck cancer [47, 48]. In ESCC, high *LOXL2* expression is a poor prognostic marker and high level of *LOXL4* is closely correlated with the poor differentiation [49]. In present study, the “loss” of crosstalks between *LINC00240* and 3 members of *LOX* gene family (including *LOXL1*, *LOXL2* and *LOXL4*) were found in the subtype 2 of ESCC (Figure 3B). Experimental data have shown that *LINC00240* and *LOXL2* possibly compete for *miR-26b-5p* binding [14], and overexpression of *miR-26b-5p* served as a poor prognostic marker in the subtype 2 of ESCC. Likewise, *LINC00240* and *LOXL1* or *LOXL4* possibly compete for *miR-124-3p* binding [14], which also has a significant difference in degree of differentiation of ESCC [50]. Thus,

Table 2: Correlations between any two lncRNAs within the modules in the different cell types

	Pairs	Adjacent tissues		Tumor tissues	
		CC	q-value	CC	q-value
Module 1	<i>BOLA3-AS1-TRAF3IP2-AS1</i>	0.32	0.016	-0.27	0.15
	<i>TRAF3IP2-AS1-LINC00689</i>	0.66	1.81e-07	0.23	0.18
	<i>LINC00689-BOLA3-AS1</i>	0.37	0.011	-0.02	0.87
Module 2	<i>LINC00240-AP000525.9</i>	0.61	3.74e-07	0.28	0.11
	<i>LINC00240-RP11-588K22.2</i>	0.52	2.09e-05	0.17	0.55
	<i>LINC00240-AC156455.1</i>	0.61	2.95e-07	0.35	0.03
	<i>AP000525.9-RP11-588K22.2</i>	0.49	8.17e-05	0.14	0.55
	<i>AP000525.9-AC156455.1</i>	0.47	8.17e-05	0.19	0.49
	<i>RP11-588K22.2-AC156455.1</i>	0.76	1.78e-12	0.15	0.55

CC represents the Pearson correlation coefficient.

in the subtype 2, the “loss” of miRNA-mediated crosstalks between *LINC00240* and *LOX* gene family might result from *LOX*-mediated dysregulation of ECM.

Module analysis of lncRNAs as ceRNAs in the two classes

To further investigate the cooperative function of multiple lncRNAs as miRNAs sponges, bidirectional hierarchical clustering on the two networks was respectively performed. We found two DCL modules in the heat map of the two subtypes. The first module of the subtype 1 contained 4 lncRNAs (*LINC00689*, *BOLA3-AS1*, *H19* and *TRAF3IP2-AS1*) and 126 mRNAs (Figure 4A), and the second module of the subtype 2 contained 5 lncRNAs (*LINC00240*, *SNHG1*, *AP000525.9*, *RP11-588K22.2*, *AC156455.1*) and 172 mRNAs (Figure 4B).

The coding genes in the first module were significantly enriched for GO terms related to neuron development (Figure 4C). We ranked these coding genes according to the degree in the first module and found 6 hub coding genes (Table S6). Among them, neuronal growth regulator 1 (*NEGR1*) and basic helix-loop-helix family member e22 (*BHLHE22*) are crucial transcription factors associated with neuron development, which as tumor suppressors were downregulated in the subtype 1 of ESCC [51, 52]. Among 4 lncRNAs, *LINC00689*, *BOLA3-AS1* and *TRAF3IP2-AS1*, were as the most frequent combination in crosstalks (Table S6). Any two lncRNAs within them were significantly positively correlated in adjacent tissues, whereas lost their correlations in tumor tissues (Table 2).

In the second module, the coding genes were significantly enriched for GO terms related to mitotic cell cycle (Figure 4D). Similarly, we found 11 hub

coding genes (Table S6), including apoptosis-regulatory gene (*BID*) and DNA double-strand break repair gene (*RAD51AP1*) etc., which are closely related to cancer development [53, 54]. The most frequent lncRNA combination of module 2 including *LINC00240*, *AP000525.9*, *RP11-588K22.2* and *AC156455.1* also lost their correlations during the development of ESCC (Table 2). Moreover, the “loss” of crosstalks between the hub coding genes and lncRNAs were found in both modules as well. These together suggested that collaborative crosstalks between these lncRNAs and the key coding genes within the functional modules existed in normal tissue and the “loss” of them might participate in ESCC development.

DISCUSSION

The ceRNA hypothesis suggests that miRNA-mediated crosstalks between mRNAs and lncRNAs are well-organized. However, molecular classification of ESCC based on mRNA-lncRNA expression data has never been reported. Our study identified two biologically and clinically relevant subtypes of ESCC based on expression profiles of lncRNA and mRNA. The functional characteristic of subtypes 1 was closely associated with “response to wounding”. *EGFR* is a crucial regulator in wound healing and tumor growth, and overexpressed *EGFR* was related to the process of ESCC infiltration in the early stages of carcinogenesis [55]. In present study, the hub lncRNA—*LINC00152* was upregulated in the subtype 1 of ESCC as well as *EGFR*. *LINC00152* can directly bind to *EGFR*, resulting in activation of *PI3K/AKT* signaling pathway in gastric cancer [29]. Intriguingly, the “gain” of miRNA-mediated crosstalk between *LINC00152* and *EGFR* was also found in the subtype 1 of ESCC.

It implied that these two genes might form a positive feedback loop to enhance *EGFR* downstream signaling pathway in tumor cell. Currently, anti-*EGFR* inhibitors have been evaluated in ESCC [56]. Thus, inhibition of activity of *LINC00152* may provide a new opportunity for therapeutic strategy.

Differential co-expression analysis (DCEA) is proposed to understand the roles of genes interconnection in complex human diseases as a complementary technique to the traditional differential expression analysis (DEA)

[57]. Combination of DCEA and DEA, a great many “loss” of miRNA-mediated crosstalks between mRNAs and lncRNAs were detected during the tumorigenesis of ESCC. The most “loss” of crosstalks between *PVT1*-mRNAs in the subtype 1 of ESCC were mediated by *miR-186*, which was inconsistent with previous report in breast cancer [43]. One possible explanation for this difference is that experimentally supported instead of predicted miRNA-lncRNA interactions was used. Another potential reason is that tissue-specific genes express in different

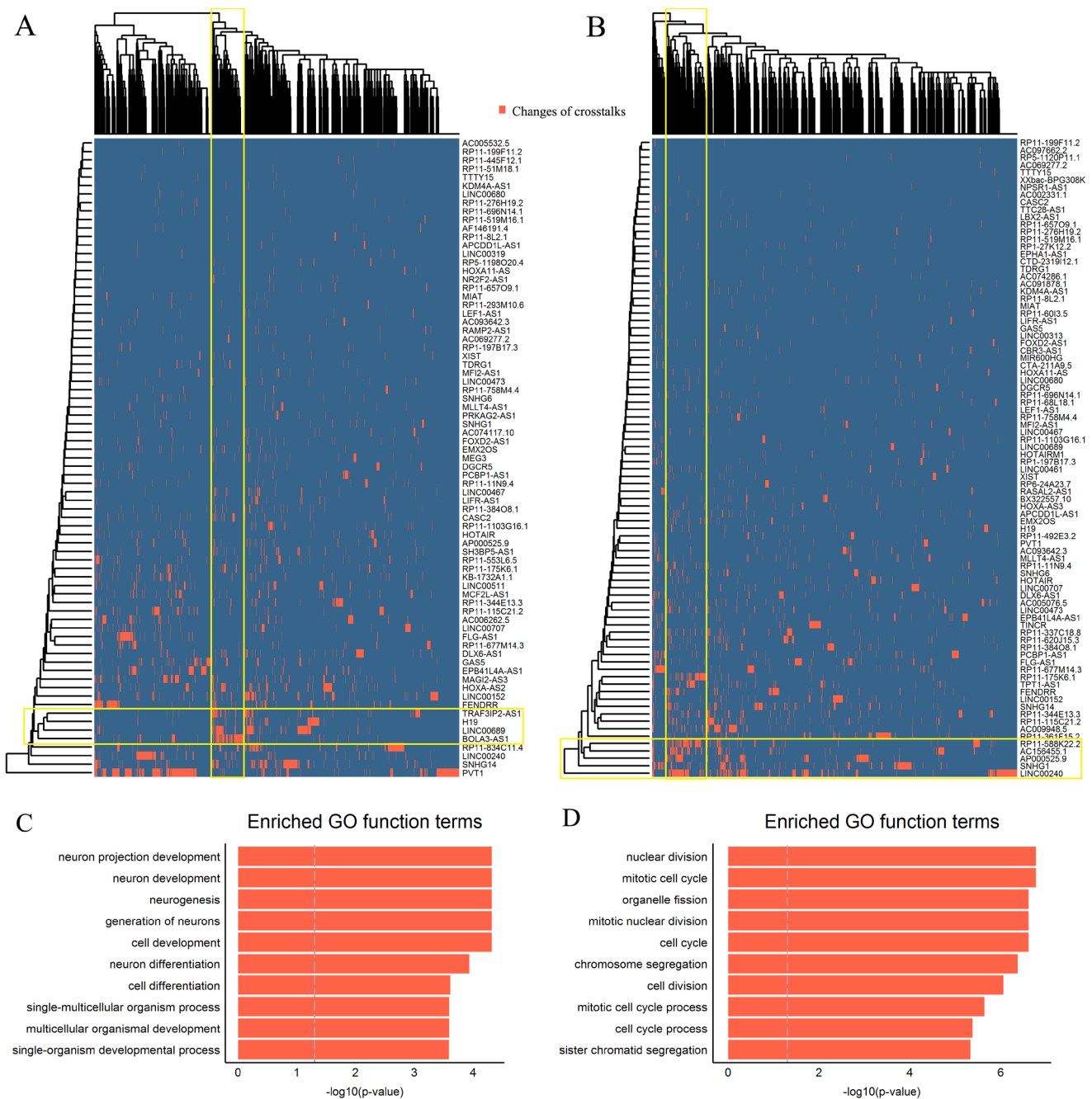


Figure 4: ESCC subtype-specific modules of differential mRNA-lncRNA crosstalks. A.-B. The bidirectional hierarchical heat maps of two differential mRNA-lncRNA crosstalk networks. The overlaps of yellow rectangles represent functional modules. **C.-D.** Top 10 significantly enriched GO terms of two modules.

adjacent normal tissues. A recent review tried to account for this “loss” of crosstalks in cancer, including lncRNA isoforms losing the specific MREs and preferential expressing the lncRNA isoforms without specific miRNA binding site [58]. Moreover, a single crosstalk between mRNA and lncRNA is mediated by several miRNAs, and dysregulation of any miRNAs may disrupt the crosstalk. Although these reasons require further confirmation, our results still discovered a number of meaningful “loss” of crosstalks between mRNAs and lncRNAs, which may serve as driver events in ESCC development.

In addition, the cooperation of multiple lncRNAs deserves our attention. More accurate prediction of patient survival has been acquired through combining lncRNA signatures in ESCC [6]. lncRNA-lncRNA synergistic networks were also discussed [33]. Our findings suggested that ceRNA network may partially account for lncRNA-lncRNA synergistic effects. In summary, our study depicted mRNA-lncRNA molecular portraits of ESCC. Analysis of subtype-specific differential mRNA-lncRNA crosstalk networks provided multiple perspectives on roles of lncRNAs involved in ESCC, and facilitated research in personalized medicine and potential new therapeutic targets for ESCC.

MATERIALS AND METHODS

Gene expression profile

lncRNA and mRNA expression profiles of cancer and adjacent normal tissues from 119 ESCC patients, and clinical data were downloaded from the GEO database under accession number of GSE53624 [6]. miRNA expression profiles of ESCC were downloaded from the GEO database under accession number of GSE43732 [59].

Probe re-annotation

We have re-annotated probes from Agilent-038314 CBC Homo sapiens lncRNA + mRNA microarray V2.0 platforms (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL18109>). Human protein-coding transcript sequences and long non-coding transcript sequences were downloaded from the NCBI Reference Sequence Database (Refseq) and GENCODE database. The microarray probes were re-annotated as follow:

- (1) The probes named as “CB*” and “RNA*” were aligned to protein-coding transcript sequences and long non-coding transcript sequences by BLASTn tools respectively.
- (2) The probes that only perfectly matched to one transcript or multiple transcripts from same gene were

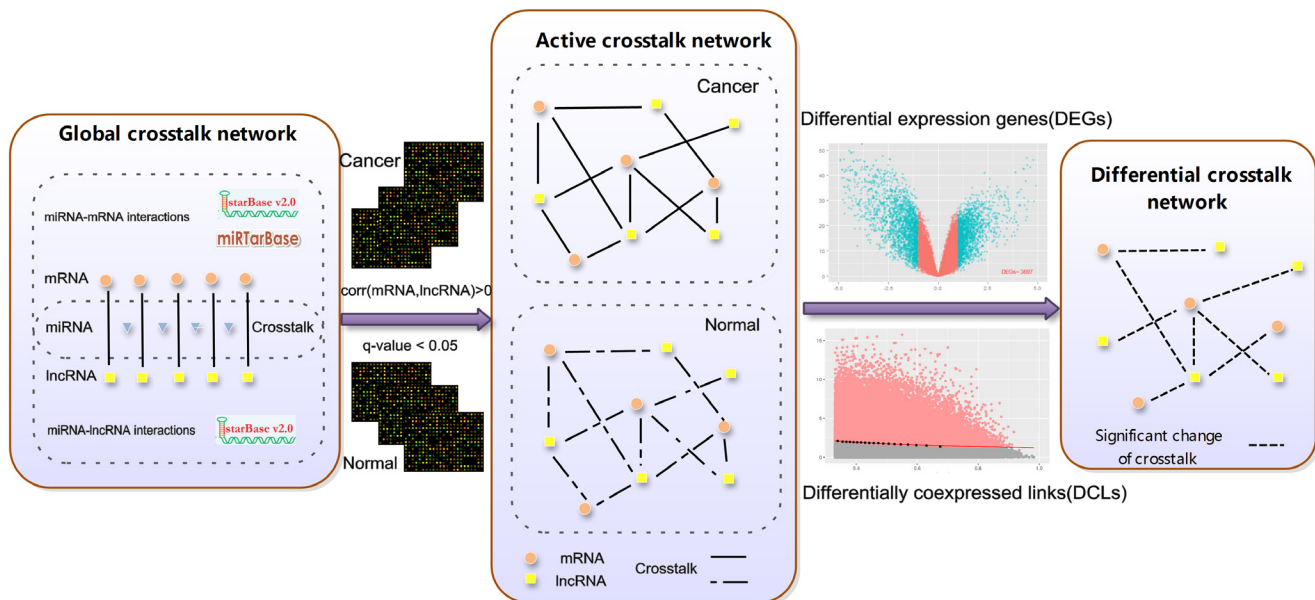


Figure 5: An integrative pipeline for construction of differential mRNA-lncRNA crosstalk networks based on ceRNA hypothesis. Experimental-supported miRNA-lncRNA and miRNAs-mRNA interactions were downloaded from starBase and miRTarBase. miRNA-lncRNA and miRNA-mRNA pairs sharing the same miRNA form the global crosstalk network. If the Pearson correlation coefficient between competing lncRNA-mRNA pairs > 0 and q value < 0.05 in ESCC tumor or adjacent tissues, the competing lncRNA-mRNA pairs were retained and formed the active crosstalk network. Finally, subtype-specific differential mRNA-lncRNA crosstalk networks based on ceRNA hypothesis are respectively constructed by integrating DEGs and DCLs.

reserved.

(3) The average expression values of multiple probes sets that mapped to the same gene or transcript were calculated.

Then, 18,755 mRNAs and 7,171 lncRNAs were annotated. The average expression levels of mRNAs and lncRNAs were 10.25 and 6.78 (log 2 transformation, Figure S5).

Differentially expressed genes (DEGs)

Moderated paired *t*-test within the linear models of R package Limma was used to assess differential expression between tumor and adjacent normal tissues [60]. DEGs were identified if the absolute value of fold-change > 2 and adjust *p*-value < 0.05 (Benjamini-Hochberg method to control the false discovery). 3,170 and 765 differentially expressed mRNAs and lncRNAs were retained in the subtype 1 of ESCC, meanwhile 3,697 and 878 differentially expressed mRNAs and lncRNAs were retained in the subtype 2 (Figure S6).

ESCC subtypes classification based on mRNA-lncRNA expression profile

Considering the influence of heterogeneity among different patients, relative expression levels between tumor and normal tissues were used. Semi-non-negative matrix factorization (semi-NMF) consensus clustering was used to identify intrinsic subtypes of ESCC based on the top 30% variable mRNAs and lncRNAs. The best number of clusters was selected based on dispersion coefficients of semi-NMF (range of cluster *k* values from 1 to $\sqrt{\text{sample size}}$, 30 times rerun of each cluster *k* and 200 iterations of each run). The value of the fitting residual of consensus clustering terminated at 1e-6 and consensus clustering iterated 1,000 times. All above were implemented by the NMF toolbox in MATLAB [61].

Construction of differential mRNA-lncRNA crosstalk network based on ceRNA hypothesis

Global mRNA-lncRNA crosstalk network and active mRNA-lncRNA crosstalk network were built before constructing the differential mRNA-lncRNA crosstalk network (Figure 5). First 668,645 human experimentally validated miRNA-mRNA interactions (2,650 miRNAs and 15,353 mRNAs) were downloaded from the starbase V2.0 database [14] and miRTarBase database [23], and the CLIP-Seq experimentally supported 10,212 miRNA-lncRNA interactions (227 miRNAs and 1,127 lncRNAs) were gotten from the starbase V2.0 database. 1,812,628 lncRNA-mRNA pairs (445 lncRNAs and 13,805 mRNAs) competition for common miRNAs binding constituted the

global mRNA-lncRNA crosstalk network.

The competing lncRNA-mRNA pairs were retained if $\text{corr}(\text{mRNA}, \text{lncRNA}) > 0$ and $q\text{-value} < 0.05$ in ESCC tumor or adjacent tissues, which formed the active mRNA-lncRNA crosstalk network (q -value was estimated from the p -value of Pearson correlation coefficients using the Benjamini-Hochberg method). Identifying the significant changes of crosstalks between mRNAs and lncRNAs during the development of ESCC was simplified into searching the differentially co-expressed links (DCLs) between ESCC tumor and adjacent tissues. The DCLs were identified by the limit fold change model (LFC) [62]. LFC defined a fraction δ of links with highest log fold change (y) between maximum co-expression (x), and links lying above the fitted curve $y = a + (b/x)$ are considered as DCLs. The top 20% significant changes of crosstalks between mRNAs and lncRNAs were retained ($\delta = 0.2$, δ setting was tested in supplementary). Then we filtered the DCLs in which lncRNAs and mRNAs were not DEGs. Finally, the differential mRNA-lncRNA crosstalk network was constituted by merging multiple DCLs into one edge if they connected same coding genes and lncRNAs.

In addition, the DCLs were divided into two groups. Competing lncRNA-mRNA pairs had higher positive correlation in adjacent tissues than tumor tissues, which were defined as the “loss” of crosstalks. On the contrary, they were regarded as the “gain” of crosstalks.

Topological features selected

Four topological features were selected to assess importance of node in the differential mRNA-lncRNA crosstalk networks.

(1) *Degree*. The number of edges linking to the given node.

(2) *Betweenness*. The number of shortest paths between any two vertexes that pass through the given node.

(3) *Closeness*. The average length of paths from given node to all other reachable nodes.

(4) *DCLs enrichment*. The DCL edges linking to the given node in the differential crosstalk network enriched the edges linking to the given node in the active crosstalk network. *DCLs enrichment* reflected the importance of nodes in differential regulation network [63]. It was measured by hypergeometric test as:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$$

where, *N* represented the total number of edges in the active crosstalk network, *n* represented the total number of DCL edges in the differential crosstalk network, *M* represented the number of edges linking to the given node in the active crosstalk network and *m* represented the number of DCL edges linking to the given node in the

differential crosstalk network.

Degree, Betweenness and Closeness were calculated by Cytoscape and *DCLs enrichment* was implemented by R.

Survival and cox regression analysis

Kaplan-Meier survival analysis was performed for the patients after surgery and statistical significance was assessed by the log-rank test. All analysis was implemented by R using the survival package.

Enrichment analysis

Gene Ontology and KEGG pathways enrichment analysis were implemented by R using annotation data packages (org.Hs.egGO, org.Hs.egPATH). Biology process terms and KEGG pathways with q-value < 0.05 were statistically significant (Benjamini-Hochberg method adjusted).

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest to disclose.

FUNDINGS

This work was supported by the National High Technology Research and Development Program of China (2015AA020104), the National Key Research and Development Program on Precision Medicine (2016YFC0901700), the National Basic Research Program of China (2011CB910204, 2011CB510102, and 2010CB529200), the National Key Technology Support Program (2013BAI101B09), the National Key Scientific Instrument and Equipment Development Project (2012YQ03026108), Medical-Engineering Cross Project of Shanghai Jiao Tong University (YG2016MS33), the National Grand Program on Key Infectious Diseases (2015ZX10004801) and the Youth Innovation Promotion Association CAS.

REFERENCES

1. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research*. 2012; 22: 1775-89.
2. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*. 2010; 465: 1033-8. doi: 10.1038/nature09309.
3. Zhou M, Zhong L, Xu W, Sun Y, Zhang Z, Zhao H, Yang L, Sun J. Discovery of potential prognostic long non-coding RNA biomarkers for predicting the risk of tumor recurrence of breast cancer patients. *Sci Rep*. 2016; 6: 31038. doi: 10.1038/srep31038.
4. Zhou M, Xu W, Yue X, Zhao H, Wang Z, Shi H, Cheng L, Sun J. Relapse-related long non-coding RNA signature to improve prognosis prediction of lung adenocarcinoma. *Oncotarget*. 2016; 7: 29720-38. doi: 10.18632/oncotarget.8825.
5. Baker M. Long noncoding RNAs: the search for function. *Nature Methods*. 2011; 8.
6. Li J, Chen Z, Tian L, Zhou C, He MY, Gao Y, Wang S, Zhou F, Shi S, Feng X. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. *Gut*. 2014: gutjnl-2013-305806.
7. Rouvelas I, Zeng W, Lindblad M, Viklund P, Ye W, Lagergren J. Survival after surgery for oesophageal cancer: a population-based study. *The lancet oncology*. 2005; 6: 864-70.
8. Song Y, Li L, Ou Y, Gao Z, Li E, Li X, Zhang W, Wang J, Xu L, Zhou Y. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature*. 2014; 509: 91-5.
9. Kimura S, Naganuma S, Susuki D, Hirono Y, Yamaguchi A, Fujieda S, Sano K, Itoh H. Expression of microRNAs in squamous cell carcinoma of human head and neck and the esophagus: miR-205 and miR-21 are specific markers for HNSCC and ESCC. *Oncology reports*. 2010; 23: 1625.
10. Ge XÄ, Ma HÄ, Zheng XÄ, Ruan HÄ, Liao XÄ, Xue WÄ, Chen YÄ, Zhang Y, Jia WÄ. HOTAIR, a prognostic factor in esophageal squamous cell carcinoma, inhibits WIF,ÄÄ1 expression and activates Wnt pathway. *Cancer science*. 2013; 104: 1675-82.
11. Tan D, Wu Y, Hu L, He P, Xiong G, Bai Y, Yang K. Long noncoding RNA H19 is up,ÄÄregulated in esophageal squamous cell carcinoma and promotes cell proliferation and metastasis. *Diseases of the Esophagus*. 2016.
12. Xie J, Guo B, Ding Z, Kang J, Deng X, Wu B, Fan Y. Microarray Analysis of lncRNAs and mRNAs Co-Expression Network and lncRNA Function as ceRNA in Papillary Thyroid Carcinoma. *Journal of Biomaterials and Tissue Engineering*. 2015; 5: 872-80.
13. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*. 2011; 146: 353-8.
14. Li J-H, Liu S, Zhou H, Qu L-H, Yang J-H. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic acids research*. 2013.
15. Wang P, Ning S, Zhang Y, Li R, Ye J, Zhao Z, Zhi H, Wang T, Guo Z, Li X. Identification of lncRNA-associated competing triplets reveals global patterns and prognostic markers for cancer. *Nucleic acids research*. 2015.

16. Zhou M, Wang X, Shi H, Cheng L, Wang Z, Zhao H, Yang L, Sun J. Characterization of long non-coding RNA-associated ceRNA network to reveal potential prognostic lncRNA biomarkers in human ovarian cancer. *Oncotarget*. 2016; 7: 12598-611. doi: 10.18632/oncotarget.7181.
17. Zhang J, Fan D, Jian Z, Chen GG, Lai PB. Cancer specific long noncoding RNAs show differential expression patterns and competing endogenous RNA potential in hepatocellular carcinoma. *PLoS one*. 2015; 10: e0141042.
18. Huang M, Zhong Z, Lv M, Shu J, Tian Q, Chen J. Comprehensive analysis of differentially expressed profiles of lncRNAs and circRNAs with associated co-expression and ceRNA networks in bladder carcinoma. *Oncotarget*. 2016; 7:47186-47200. doi: 10.18632/oncotarget.9706.
19. Song C, Zhang J, Liu Y, Pan H, Qi H, Cao Y, Zhao J, Li S, Guo J, Sun H. Construction and analysis of cardiac hypertrophy-associated lncRNA-mRNA network based on competitive endogenous RNA reveal functional lncRNAs in cardiac hypertrophy. *Oncotarget*. 2016; 7: 10827-40. doi: 10.18632/oncotarget.7312.
20. Hansen LV, Lærum OD, Illemann M, Nielsen BS, Ploug M. Altered expression of the urokinase receptor homologue, C4. 4A, in invasive areas of human esophageal squamous cell carcinoma. *International journal of cancer*. 2008; 122: 734-41.
21. Bagherani N. The role of wounding therapies in preventing squamous cell carcinoma. *Dermatologic therapy*. 2014; 27: 312. doi: 10.1111/dth.12156.
22. Sunpaweravong P, Sunpaweravong S, Puttawibul P, Mitarnun W, Zeng C, Baron AE, Franklin W, Said S, Varella-Garcia M. Epidermal growth factor receptor and cyclin D1 are independently amplified and overexpressed in esophageal squamous cell carcinoma. *Journal of cancer research and clinical oncology*. 2005; 131: 111-9.
23. Chou C-H, Chang N-W, Shrestha S, Hsu S-D, Lin Y-L, Lee W-H, Yang C-D, Hong H-C, Wei T-Y, Tu S-J. miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic acids research*. 2016; 44: D239-D47
24. Casbas-Hernandez P, Sun X, Roman-Perez E, D'arcy M, Sandhu R, Hishida A, McNaughton KK, Yang XR, Makowski L, Sherman ME. Tumor intrinsic subtype is reflected in cancer-adjacent tissue. *Cancer Epidemiology Biomarkers & Prevention*. 2015; 24: 406-14.
25. Tan D, Wu Y, Hu L, He P, Xiong G, Bai Y, Yang K. Long noncoding RNA H19 is up-regulated in esophageal squamous cell carcinoma and promotes cell proliferation and metastasis. *Diseases of the Esophagus*. 2016.
26. Guan Y, Kuo W-L, Stilwell JL, Takano H, Lapuk AV, Fridlyand J, Mao J-H, Yu M, Miller MA, Santos JL. Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. *Clinical Cancer Research*. 2007; 13: 5745-55.
27. Meyer KB, Maia A-T, O'Reilly M, Ghousaini M, Prathalingam R, Porter-Gill P, Ambs S, Prokunina-Olsson L, Carroll J, Ponder BA. A functional variant at a prostate cancer predisposition locus at 8q24 is associated with PVT1 expression. *PLoS Genet*. 2011; 7: e1002165.
28. Ding J, Li D, Gong M, Wang J, Huang X, Wu T, Wang C. Expression and clinical significance of the long non-coding RNA PVT1 in human gastric cancer. *OncoTargets & Therapy*. 2014; 7.
29. Zhou J, Zhi X, Wang L, Wang W, Li Z, Tang J, Wang J, Zhang Q, Xu Z. Linc00152 promotes proliferation in gastric cancer through the EGFR-dependent pathway. *Journal of Experimental & Clinical Cancer Research*. 2015; 34: 1.
30. Xu T-p, Xia R, Liu X-x, Sun M, Yin L, Chen W-m, Han L, Zhang E-b, Kong R, De W. Decreased expression of the long non-coding RNA FENDRR is associated with poor prognosis in gastric cancer and FENDRR regulates gastric cancer cell metastasis by affecting fibronectin1 expression. *Journal of hematology & oncology*. 2014; 7: 1.
31. Zhao H, Zhang X, Frazão JB, Condino-Neto A, Newburger PE. HOX antisense lincRNA HOXA-AS2 is an apoptosis repressor in all Trans retinoic acid treated NB4 promyelocytic leukemia cells. *Journal of cellular biochemistry*. 2013; 114: 2375-83.
32. You J, Fang N, Gu J, Zhang Y, Li X, Zu L, Zhou Q. Noncoding RNA small nucleolar RNA host gene 1 promote cell proliferation in nonsmall cell lung cancer. *Indian journal of cancer*. 2014; 51: 99.
33. Li Y, Chen J, Zhang J, Wang Z, Shao T, Jiang C, Xu J, Li X. Construction and analysis of lncRNA-lncRNA synergistic networks to reveal clinically relevant lncRNAs in cancer. *Oncotarget*. 2015; 6: 25003-16. doi: 10.18632/oncotarget.4660.
34. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, Li M, Wang G, Liu Y. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic acids research*. 2009; 37: D98-D104.
35. Pekarsky Y, Croce CM. Role of miR-15/16 in CLL. *Cell Death & Differentiation*. 2015; 22: 6-11.
36. Dai L, Wang W, Zhang S, Jiang Q, Wang R, Dai L, Cheng L, Yang Y, Wei YQ, Deng HX. Vector-based miR-15a/16-1 plasmid inhibits colon cancer growth *in vivo*. *Cell biology international*. 2012; 36: 765-70.
37. 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. doi: 10.1093/bib/bbw076.
38. Liu M, Wang Z, Yang S, Zhang W, He S, Hu C, Zhu H, Quan L, Bai J, Xu N. TNF- α is a novel target of miR-19a. *International journal of oncology*. 2011; 38: 1013.
39. Li H, Yin C, Zhang B, Sun Y, Shi L, Liu N, Liang S, Lu S, Liu Y, Zhang J. PTTG1 promotes migration and invasion of human non-small cell lung cancer cells and is modulated by miR-186. *Carcinogenesis*. 2013; 34: 2145-55.

40. Ito T, Kan T, David S, Abraham J, Mori Y, Cheng Y, Paun B, Jin Z, Oлару A, Agarwal R. Downregulation of PTTG1 by siRNA suppresses cell proliferation of esophageal squamous cell carcinoma. *Cancer Research*. 2007; 67: 2106.
41. He W, Feng J, Zhang Y, Wang Y, Zang W, Zhao G. microRNA-186 inhibits cell proliferation and induces apoptosis in human esophageal squamous cell carcinoma by targeting SKP2. *Laboratory Investigation*. 2016; 96: 317-24.
42. Paraskevopoulou MD, Vlachos IS, Karagkouni D, Georgakilas G, Kanellos I, Vergoulis T, Zagganas K, Tsanakas P, Floros E, Dalamagas T. DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic acids research*. 2016; 44: D231-D8.
43. Paci P, Colombo T, Farina L. Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. *BMC systems biology*. 2014; 8: 83.
44. Chen Z, Zhao L, Zhao F, Yang G, Wang J. MicroRNA-26b regulates cancer proliferation migration and cell cycle transition by suppressing TRAF5 in esophageal squamous cell carcinoma. *Am J Transl Res*. 2016; 8: 1957-70.
45. Tetreault M-P, Yang Y, Katz JP. Kruppel-like factors in cancer. *Nature Reviews Cancer*. 2013; 13: 701-13.
46. Lyng H, Brøvig RS, Svendsrud DH, Holm R, Kaalhus O, Knutstad K, Oksefjell H, SundfØr K, Kristensen GB, Stokke T. Gene expressions and copy numbers associated with metastatic phenotypes of uterine cervical cancer. *BMC genomics*. 2006; 7: 1.
47. Le Q-T, Harris J, Magliocco AM, Kong CS, Diaz R, Shin B, Cao H, Trotti A, Erler JT, Chung CH. Validation of lysyl oxidase as a prognostic marker for metastasis and survival in head and neck squamous cell carcinoma: Radiation Therapy Oncology Group trial 90-03. *Journal of Clinical Oncology*. 2009; 27: 4281-6.
48. Barker HE, Chang J, Cox TR, Lang G, Bird D, Nicolau M, Evans HR, Gartland A, Erler JT. LOXL2-mediated matrix remodeling in metastasis and mammary gland involution. *Cancer research*. 2011; 71: 1561-72.
49. Li T-Y, Xu L-Y, Wu Z-Y, Liao L-D, Shen J-H, Xu X-E, Du Z-P, Zhao Q, Li E-M. Reduced nuclear and ectopic cytoplasmic expression of lysyl oxidase-like 2 is associated with lymph node metastasis and poor prognosis in esophageal squamous cell carcinoma. *Human pathology*. 2012; 43: 1068-76.
50. Cheng Y, Li Y, Nian Y, Liu D, Dai F, Zhang J. STAT3 is involved in miR-124-mediated suppressive effects on esophageal cancer cells. *BMC cancer*. 2015; 15: 1.
51. Kim H, Hwang J-S, Lee B, Hong J, Lee S. Newly identified cancer-associated role of human neuronal growth regulator 1 (NEGR1). *Journal of Cancer*. 2014; 5: 598.
52. Ross SE, McCord AE, Jung C, Atan D, Mok SI, Hemberg M, Kim T-K, Salogiannis J, Hu L, Cohen S. Bhlhb5 and Prdm8 form a repressor complex involved in neuronal circuit assembly. *Neuron*. 2012; 73: 292-303.
53. Orzechowska EJ, Kozłowska E, Czuby A, Kozłowski P, Staron K, Trzcinska-Danielewicz J. Controlled delivery of BID protein fused with TAT peptide sensitizes cancer cells to apoptosis. *BMC cancer*. 2014; 14: 771.
54. Miles GD, Seiler M, Rodriguez L, Rajagopal G, Bhanot G. Identifying microRNA/mRNA dysregulations in ovarian cancer. *BMC research notes*. 2012; 5: 1.
55. Hanawa M, Suzuki S, Dobashi Y, Yamane T, Kono K, Enomoto N, Ooi A. EGFR protein overexpression and gene amplification in squamous cell carcinomas of the esophagus. *International journal of cancer*. 2006; 118: 1173-80.
56. Lacouture ME. Mechanisms of cutaneous toxicities to EGFR inhibitors. *Nature Reviews Cancer*. 2006; 6: 803-12.
57. de la Fuente A. From 'differential expression' to 'differential networking'-identification of dysfunctional regulatory networks in diseases. *Trends in genetics*. 2010; 26: 326-33.
58. Colombo T, Farina L, Macino G, Paci P. PVT1: a rising star among oncogenic long noncoding RNAs. *BioMed research international*. 2015; 2015.
59. Chen Z, Li J, Tian L, Zhou C, Gao Y, Zhou F, Shi S, Feng X, Sun N, Yao R. MiRNA expression profile reveals a prognostic signature for esophageal squamous cell carcinoma. *Cancer letters*. 2014; 350: 34-42.
60. Smyth GK. (2005). *Limma: linear models for microarray data*. *Bioinformatics and computational biology solutions using R and Bioconductor*: Springer, pp. 397-420.
61. Li Y, Ngom A. The non-negative matrix factorization toolbox for biological data mining. *Source code for biology and medicine*. 2013; 8: 1.
62. Yu H, Liu B-H, Ye Z-Q, Li C, Li Y-X, Li Y-Y. Link-based quantitative methods to identify differentially coexpressed genes and gene pairs. *BMC bioinformatics*. 2011; 12: 1.
63. Yang J, Yu H, Liu B-H, Zhao Z, Liu L, Ma L-X, Li Y-X, Li Y-Y. DCGL v2. 0: an R package for unveiling differential regulation from differential co-expression. *PLoS One*. 2013; 8: e79729.