



# A transcriptomic analysis of malignant transformation of human embryonic esophageal epithelial cells by HPV18 E6E7

Duo Tang<sup>1</sup>, Biqi Wang<sup>1</sup>, Sara Khodahemmati<sup>2</sup>, Jingtao Li<sup>1</sup>, Zhixiang Zhou<sup>1</sup>, Jingfeng Gao<sup>2</sup>, Wang Sheng<sup>1</sup>, Yi Zeng<sup>1</sup>

<sup>1</sup>College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100124, China; <sup>2</sup>College of Environmental and Energy Engineering, Beijing University of Technology, Beijing 100124, China

**Contributions:** (I) Conception and design: Z Zhou, D Tang; (II) Administrative support: Z Zhou, J Gao, WSheng, Y Zeng; (III) Provision of study materials or patients: J Li, Y Zeng; (IV) Collection and assembly of data: Z Zhou, D Tang, S Khodahemmati, B Wang; (V) Data analysis and interpretation: D Tang, B Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

**Correspondence to:** Dr. Zhixiang Zhou. College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100124, China. Email: zhouzhixiang@bjut.edu.cn.

**Background:** Esophageal cancer is one of the most common malignant tumours in humans. A series of esophageal cancer cell lines are accompanied by human papilloma virus (HPV) infection, but the mechanism behind HPV in cancer malignancy is not clear.

**Methods:** This research was conducted in different generations of HPV E6E7 gene-induced human foetal esophageal epithelial immortalised cells (Shantou Human Embryonic Esophageal Epithelial cell line; SHEE); the RNA sequencing transcriptomic analysis was performed to explore the mechanism of HPV infection in these cell lines.

**Results:** The results showed that there are 9,990 differential genes in late-stage cells compared with HPV18 E6E7-infected early foetal esophageal epithelial immortalised cells. Among these, 4,882 genes are upregulated, and 5,108 genes are downregulated. We used bioinformatics to analyze the expression and function of aberrantly expressed lncRNA, miRNA, mRNA and construct the competing endogenous RNA (ceRNA) network and protein protein interaction (PPI) network.

**Conclusions:** we predicted TP53TG1 promotes to malignant transformation of SHEEs by acting as a ceRNA to competitively bind to miR-6835 and regulate IGF2 expression. We also predicted IL6 serve as prognostic biomarkers and therapy target. With these results maybe provides new insights into the mechanisms of HPV carcinogenesis in esophageal cancer.

**Keywords:** Human papilloma virus (HPV); esophageal cancer; RNA-Seq; qRT-PCR

Submitted Sep 19, 2019. Accepted for publication Jan 18, 2020.

doi: 10.21037/tcr.2020.02.23

**View this article at:** <http://dx.doi.org/10.21037/tcr.2020.02.23>

## Introduction

Esophageal cancer is one of the most common malignant tumours in humans. Data from 2012 show that there were 455,800 of new cases of esophageal cancer worldwide, and of these cases, 400,200 died (1). One problem is that esophageal cancer lacks the typical early symptoms or is misdiagnosed as other diseases, such as cardiac spasms. It demonstrated that most advanced esophageal cancer

patients were unable to undergo surgery, meaning their survival rate is less than 10% in 5 years after diagnosis. However, once esophageal cancer is diagnosed early and treated promptly, the 10-year survival rate of patients with esophageal cancer can reach more than 95% (2). Therefore, it is particularly important to develop a new diagnostic strategy to improve the cure rate of esophageal cancer.

HPV is a double-stranded small DNA virus and has

been reported to be closely related to various tumours (3). In recent decades, the works of our laboratory and other international laboratories have shown that HPV is also an important cause of esophageal cancer (4-12). More than 80% of esophageal cancer patients in China's Shantou area are concomitant with HPV 16 and 18 infections. E6 and E7, two HPV early oncoproteins, are responsible for malignant transformation. The E6 protein can bind and degrade p53 and destroy the normal control of the cell cycle, while the E7 protein can bind with the pRb protein, hence, the pRb loses its ability to be an antioncogene. The effects of the E6 and E7 proteins lead to malignant cells to proliferate (13). Therefore, we established an immortalised esophageal epithelial cell line SHEEs with an adeno-associated virus containing HPV18 E6E7 to infect human foetal esophageal epithelial cells, these transformed cells are more likely to progress to anchorage-independent growth and tumorigenicity and may provide an *in vitro* model of carcinogenesis induced by HPV (14,15). However, it is still not clear if genetic and cellular mechanisms take place in the immortalisation and malignant transformation of human foetal esophageal epithelial cells induced by HPV18 E6E7.

In the current work, we passaged the SHEE cell to 26th generation (P26) and 79th generations (P79), And then, the RNA-Seq analysis was used to compare the transcriptional differences between the two generation cells. Previous studies have shown, the SHEE cell from the 10th to the 85th passage were changed gradually from preimmortal, immortal to malignantly transformed stages (14). In our work, we passaged the SHEE cell to 26th generation (P26) which is an early immortalized cell and 79th generations (P79) which is a pre-malignant transformation cell. And then, the RNA-Seq analysis was used to compare the transcriptional differences between the two generation cells.

## Methods

### Cell culture

SHEE cells induced by the HPV 18 E6 and E7 genes were obtained from the Institute of Virology, Chinese Academy of Preventive Medicine. The SHEE cells were cultured in Dulbecco's modified Eagle's media (DMEM; Gibico, USA) supplemented with 10% fetal bovine serum (FBS; Gibico, USA), 100 U/mL penicillin and 100 mg/mL streptomycin (Hyclone, USA) and then incubated at 37 °C, 5% CO<sub>2</sub>. SHEE cells were taken from 26 passages (P26) and 79 passages (P79) for further analysis.

### Quantification and identification of RNA

RNA degradation and contamination were detected on a 1% agarose gel. RNA purity was detected using a NanoPhotometer<sup>®</sup> spectrophotometer (IMPLEN, CA, USA). The RNA concentration was measured using Qubit<sup>®</sup> RNA assay kit in Qubit<sup>®</sup> 2.0 Fluorometer (Life Technologies, CA, USA), and the RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

### Transcription groups

The total amount of each sample was 3 µg of RNA used as an experimental material for RNA sample preparation. A sequencing library was generated by using the NEBNext<sup>®</sup> UltraTM RNA Library Prep Kit for Illumina<sup>®</sup> (NEB, USA), per the instructions, and an index code was added to the attribute sequence for each sample. The mRNA was purified from the total RNA by using magnetic beads attached with poly-T oligo. In the NEBNext first-strand synthesis reaction buffer (5X), divalent cations were used for fragmentation as the temperature increased. The first-strand cDNA was synthesised using a random hexamer primer and M-MuLV reverse transcriptase (RNase H<sup>-</sup>). Second-strand cDNA synthesis was subsequently performed using DNA polymerase I and RNase H. The remaining overhangs were converted to blunt ends by exonuclease/polymerase activity. After being adenylated at the 3' end of the DNA fragment, a NEBNext adaptor with a hairpin loop structure was ligated to prepare for hybridisation. To select the cDNA fragments of, preferably, 150–200 bp in length, the library fragments were purified using the AMPure XP system (Beckman Coulter, Beverly, USA). The cDNA that was ligated with a size-selected adaptor was then used at 37 °C for 15 minutes using 3 µL USER enzyme (NEB, USA) and then at 95 °C for 5 minutes. A PCR was carried out using the Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) Primer. Finally, the PCR product (AMPure XP system) was purified, and the library quality was assessed using an Agilent Bioanalyzer 2100 system.

### Clustering and sequencing

Clustering of the samples was performed on the cBot Cluster Generation System using the TruSeq PE Cluster Kit v3-cBot-HS (Illumina). After clustering generation,

library preparations were sequenced on an Illumina HiSeq platform, generating paired-end reads of 125 bp/150 bp.

### *Quantification of gene expression levels*

HTSeq V0.6.0 was used to count each gene reading. The fragments per kilobase of transcript per million (FPKM) of each gene was then calculated based on the length of the gene, and the reading count was mapped to the gene. The FPKM takes into account the effect of sequencing depth and gene length using how many reads are used and is the most commonly used method for estimating gene expression levels.

### *Differential expression analysis*

Differential expression analysis of two conditions/groups (two biological replicates per condition) was performed using the DESeq2 R package (1.10.1). DESeq2 provide statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting P values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted P value <0.05 found by DESeq2 were assigned as differentially expressed.

Prior to differential gene expression analysis, for each sequenced library, the read counts were adjusted by edgeR program package through one scaling normalized factor. Differential expression analysis of two conditions was performed using the edgeR R package (3.12.1). The P values were adjusted using the Benjamini and Hochberg method. Corrected P value of 0.05 and absolute fold change of 2 were set as the threshold for significantly differential expression.

### *Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of differentially expressed genes*

A GO enrichment analysis of differentially expressed genes (DEGs) was performed using the cluster Profiler R package, in which the gene length deviation was corrected. GO items with a P value of less than 0.05 correction were considered to be significantly enriched by DEGs. The KEGG is a database resource for understanding the advanced functions and effects of biological systems, such as cells, organisms and ecosystems, hence providing molecular-

level information, especially genome sequencing and other high-throughput, large-scale molecular data (<http://www.genome.jp/kegg/>). We used clusterProfiler R package to test the statistical enrichment of differential expression genes in KEGG pathways. KEGG enrichment is significantly enriched with a padj of less than 0.05.

### *Construction of the ceRNA network*

The putative lncRNA-miRNA-mRNA interactions were evaluated using the algorithms of Targetscan version 7.2 (<http://www.targetscan.org/>) and ENCORI (<http://starbase.sysu.edu.cn/>). The ceRNA network was constructed and visually displayed using the Cytoscape software V3.7.1 (<https://cytoscape.org/>).

### *Construction of the PPI network*

To study the PPI of malignant transformation of human embryonic esophageal epithelial cells by HPV18, Omicsbean (<http://www.omicsbean.cn/>) online tools were used to construct a PPI network for significantly differential mRNAs.

### *Survival analysis*

To analyse the prognostic significance of DEGs, survival analysis was conducted using clinical information from TCGA. Online tools GEPIA (<http://gepia.cancer-pku.cn/>) and OncoLnc (<http://www.oncolnc.org/>) were used to survival analysis of DEGs.

### *Real-time qPCR*

We selected seven DEGs involved in apoptosis, migration and inflammation for quantitative reverse transcriptase PCR (qRT-PCR) validation of the RNA-Seq results. The PCR primers were designed using Oligo7 software and are listed in *Table 1*. Total RNA was extracted from cells using trizol reagent (Invitrogen, USA). The 1 µg of total extracted RNA was reverse transcribed in a final volume of 20 µL to synthesise the first-strand cDNA using a PrimeScript RT Reagent Kit with gDNA Eraser (TaKaRa Biotechnology, Dalian, China) per the instructions of the manufacturer. Gene expression was detected by qRT-PCR using SYBR® Premix DimerEraser™ (TaKaRa Biotechnology, Dalian, China). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as an endogenous control to normalise

**Table 1** Primer sequence for real-time qPCR

Symbol	Oligonucleotide (5' to 3')
<i>MAGEA3</i>	Sense: TCACTCTTTGAAGCGAGCAG; antisense: TGCTGACGCTCATTCAACCA
<i>IGFBP5</i>	Sense: GCTCAACTTCGAAAATGGCAAC; antisense: TCTGTTTCATCCAACCTTGCCCT
<i>SPP1</i>	Sense: TATGCTGGTTGTAGACCCCAA; antisense: ATTGACCTCAGAAGATGCACT
<i>LINC00958</i>	Sense: CACGTTTTATTCTGACTGCT; antisense: AGTGGACTCATCTTTGCCT
<i>SPOCK1</i>	Sense: CCTTTGGCTTCGTAATAACCC; antisense: CCCAAAACAGAGTTTGCCTA
<i>IL6</i>	Sense: ACTCACCTCTTCAGAACGAATTG; antisense: CCATCTTTGGAAGGTTTCAGGTTG
<i>S100A7A</i>	Sense: TCCATAATAGGCATGATCGAC; antisense: TCAATCTTGCCATCACGTC
<i>GAPDH</i>	Sense: TGTGGCCATCAATGACCCCTTC; antisense: AGCATCGCCCCACTTGATTTTG

expression, and SHEE P26 was used as the experimental control. In total, 10  $\mu$ L of PCR reactions in quadruplicate were carried out by an initial denaturation at 95 °C for 30 s followed by 40 cycles, each consisting of 30 s at 95 °C, 30 s at 55 °C, 30 s at 72 °C and then one cycle for melting curve that consisted of 0 s at 95 °C, 15 s at 65 °C and 0 s at 95 °C. The  $2^{-\Delta\Delta C_t}$  method was used to calculate the relative gene expression levels.

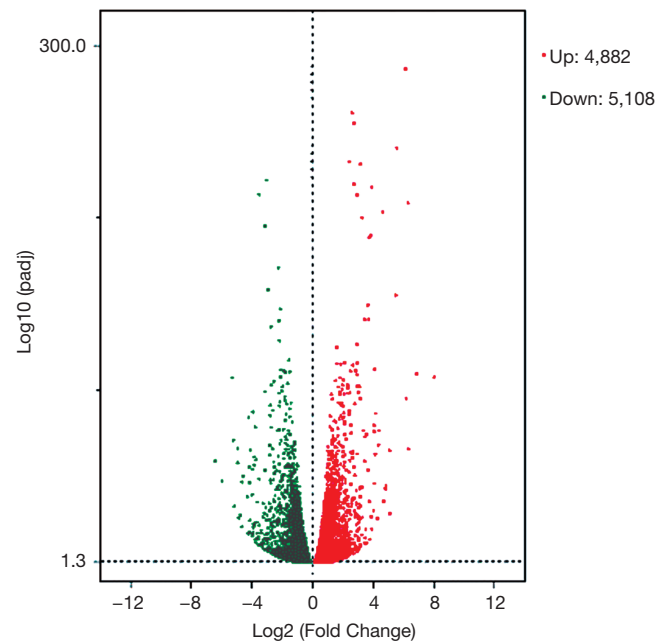
### Statistical analysis

For the obtained experimental data, the mean ( $\pm$  standard deviation) of at least three independent experiments was calculated. Using an analysis of variance (ANOVA), the statistical differences were considered to be significant when the P value was less than 0.05.

## Results

### Analysis of DEGs between P26 and P79 cells

Using DESeq2 software to perform a differential expression analysis, the padj was set to less than 0.05 as the difference significance criterion, and a total of 9,990 differential genes were obtained, including 4,882 upregulated genes and 5,108 downregulated genes (Figure 1). We focused on genes with



**Figure 1** Differential gene volcano map. The volcano map can visualize the overall distribution of genes with the significant difference, and the horizontal axis represents the fold change in the expression of the gene in different samples (Log2Fold Change), and the vertical axis indicates the significant difference in expression ( $-\text{Log}_{10}\text{Padj}$ ). The upward gene represented by the red dot and the downward gene indicated by a green dot.

an absolute value  $>5$  of expression fold change ( $\log_2\text{Fold Change}$ ,  $\log_2\text{FC}$ ), that is, genes with expression *MAGEA3*, *CAPN6*, *S100A7A*, *SPP1*, *SLCO2B1*, *IGFBP5*, *MT1E*, *S100A7*, *RIBC2*, *IQGAP2* were significantly up-regulated. It is worth noting that *MAGEA3* is most significant up-regulated gene. Six DEGs were down-regulated, they are *LINC00958*, *IL6*, *KRT6A*, *ELFN2*, *C6*, *SPOCK1*. *LINC00958* is most significant down-regulated gene in all differential genes (Table 2, Figure 2).

### Differentially expressed lncRNAs between P26 and P79 cells

There are 82 lncRNAs were identified in differential expression analysis (Figure 3). Thirteen lncRNAs were found to be aberrantly expression, an absolute value  $>5$  of  $\log_2\text{FC}$  and the padj was set to less than 0.05. Six lncRNAs were downregulated and 7 lncRNAs were upregulated (Table 3, Figure 4).

**Table 2** Differential genetic classification of prominent type between SHEE P26 and P79, as measured by RNA-Seq. Log2FC >5.0, Padj ≤0.05

Gene name	Gene type	Log2FC	Gene description
Up-regulated			
<i>MAGEA3</i>	Protein coding	8.027788	MAGE family member A3
<i>CAPN6</i>	Protein coding	6.882745	Calpain 6
<i>S100A7A</i>	Protein coding	6.309254	S100 calcium binding protein A7A
<i>SPP1</i>	Protein coding	6.285962	Secreted phosphoprotein 1
<i>SLCO2B1</i>	Protein coding	6.177653	Solute carrier organic anion transporter family member 2B1
<i>IGFBP5</i>	Protein coding	6.128139	Insulin like growth factor binding protein 5
<i>MT1E</i>	Protein coding	5.542693	Metallothionein 1E
<i>S100A7</i>	Protein coding	5.487781	S100 calcium binding protein A7
<i>RIBC2</i>	Protein coding	5.077229	RIB43A domain with coiled-coils 2
<i>IQGAP2</i>	Protein coding	5.076827	IQ motif containing GTPase activating protein 2
Down-regulated			
<i>LINC00958</i>	lncRNA	-6.44381	Long intergenic non-protein coding RNA 958
<i>IL6</i>	Protein coding	-5.9908	Interleukin 6
<i>KRT6A</i>	Protein coding	-5.32287	Keratin 6A type II
<i>ELFN2</i>	Protein coding	-5.24081	Extracellular leucine-rich repeat and fibronectin type III domain containing 2
<i>C6</i>	Protein coding	-5.17227	Complement component 6
<i>SPOCK1</i>	Protein coding	-5.07011	Sparc/osteonectin cwcx and kazal-like domains proteoglycan (testican) 1

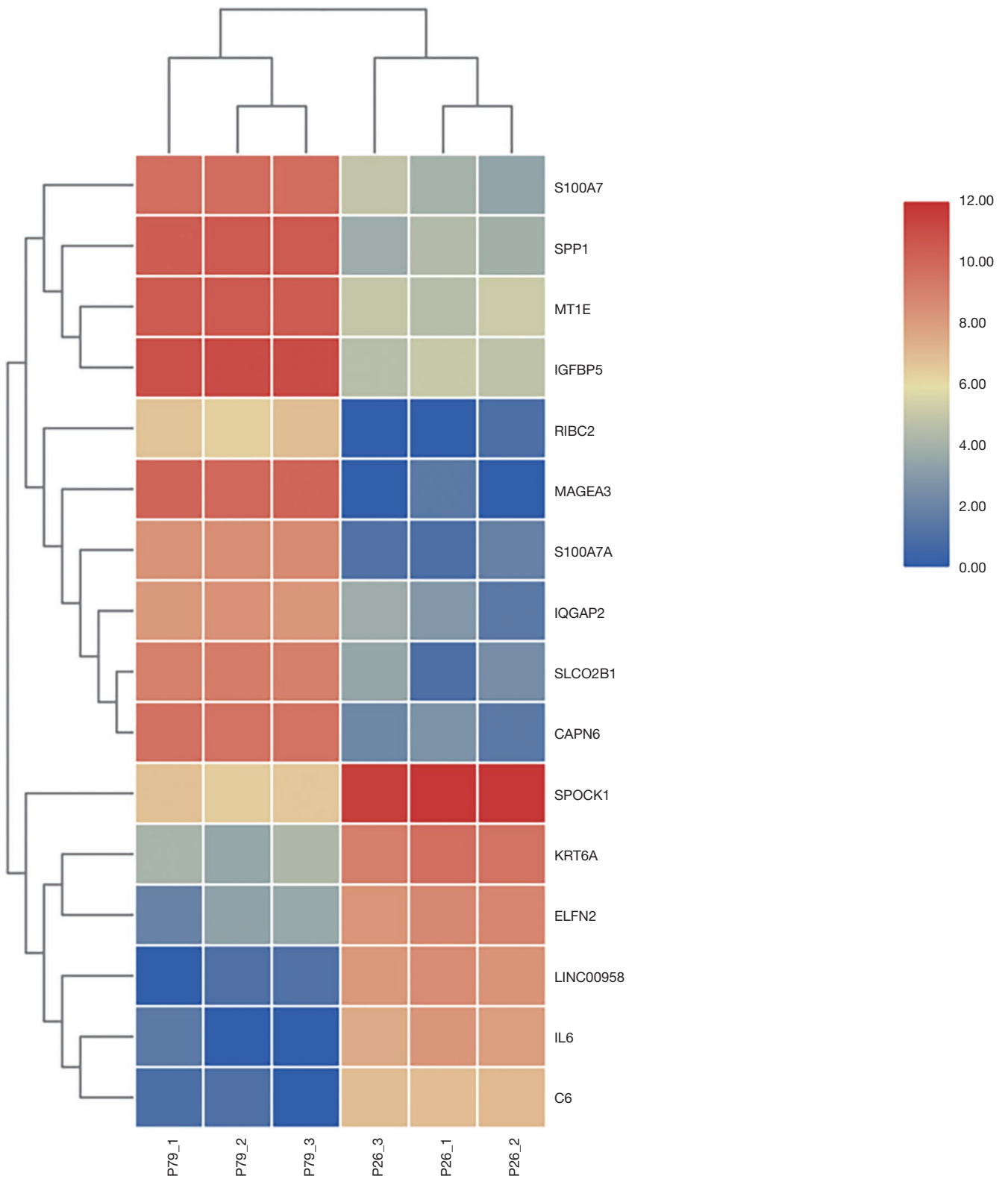
### Enrichment analysis

GO category and KEGG pathway were analyzed for relationships among all DEGs in human embryonic esophageal epithelial cells were analysed using GO and KEGG. A GO analysis is a comprehensive database describing gene function and can be divided into three parts: biological process (BP), molecular function (MF), and cellular component (CC). GO analysis results showed that changes in BP were enriched in ribosome biogenesis, ribonucleoprotein complex biogenesis, DNA replication, in utero embryonic development, ncRNA processing, response to endoplasmic reticulum stress, cell cycle phase transition, mitotic cell cycle phase transition, response to type I interferon, DNA-dependent DNA replication. Changes in MF mainly enriched in transcription factor binding, damaged DNA binding, structure-specific DNA binding, double-stranded DNA binding, ubiquitin protein ligase binding, ubiquitin-like protein ligase binding, protein serine/threonine kinase activity, magnesium ion

binding, ubiquitin-like protein transferase activity, helicase activity. Changes in CC mainly enriched in nucleolar part, mitochondrial inner membrane, organelle inner membrane, nuclear membrane, nuclear envelope, preribosome, mitochondrial matrix, centrosome, focal adhesion, transferase complex and transferring phosphorus-containing groups (*Figure 5*). KEGG is a comprehensive database that integrates genomic, chemical and system function information. As shown in *Figure 6*, the three most affected pathways are ribosome biogenesis in eukaryotes, protein processing in the endoplasmic reticulum and the cell cycle. *Table 4* shows the expressions of the genes involved in each pathway.

### Construction of the ceRNA network

CeRNA network was built on the basis of the differential expression lncRNA, miRNA, and mRNA. As shown in *Figure 7*, 5 lncRNA nodes, 12 miRNA nodes, 383 mRNA nodes and 1596 edges were in the network.



**Figure 2** Prominent differential gene heat map. Red indicates high gene expression, while Blue indicates low gene expression.

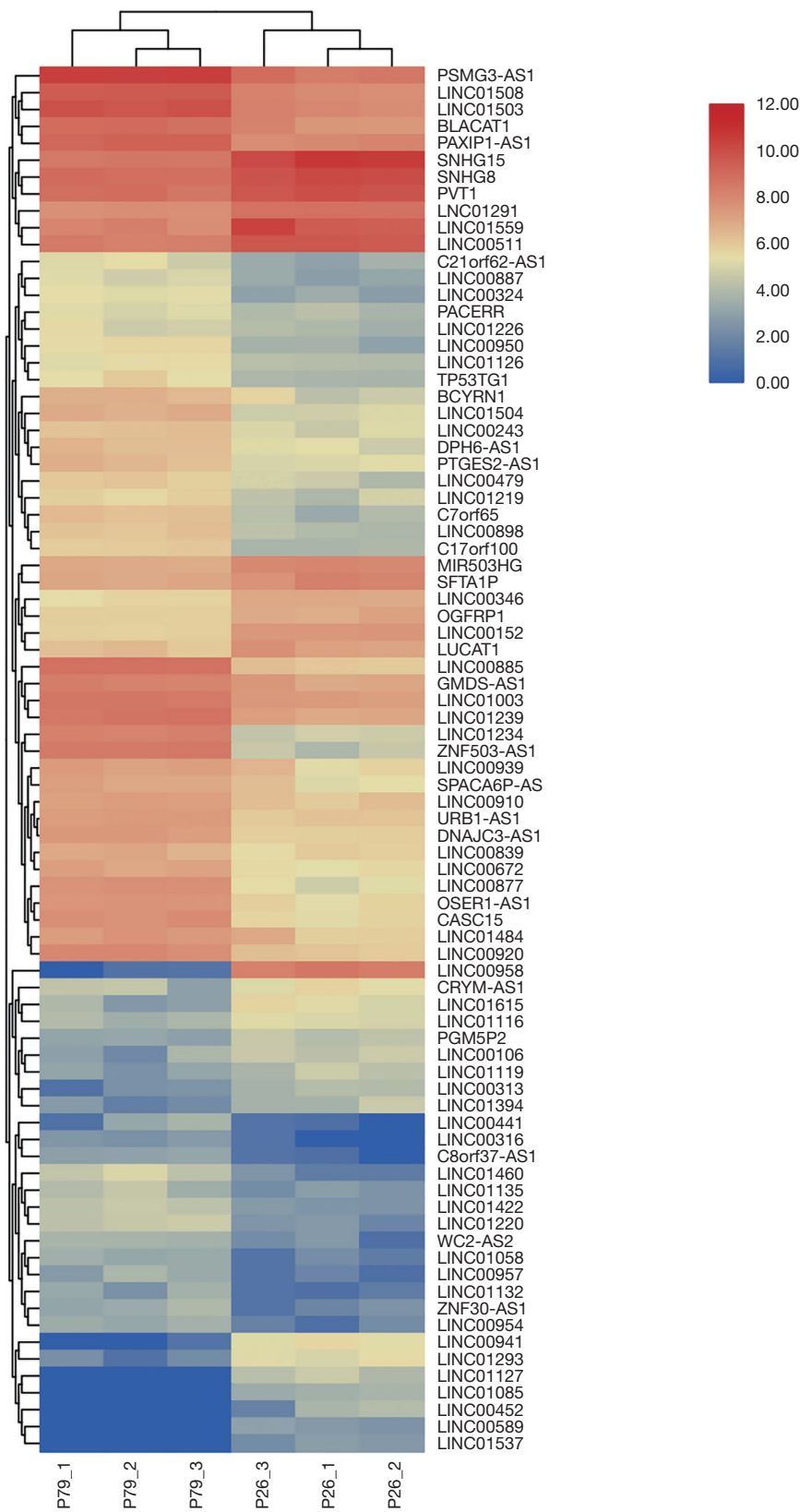


Figure 3 A heatmap of all lncRNAs in DEGs. Red indicates high gene expression, while Blue indicates low gene expression.

**Table 3** Differential lncRNAs between SHEE P26 and P79, as measured by RNA-Seq. Log2FC >2.0, Padj ≤0.05

Gene name	GeneType	Log2FC	Padj
LINC00958	LncRNA	-6.44381	2.90E-60
LINC00941	LncRNA	-3.81448	6.93E-15
LINC01127	LncRNA	-2.9287	6.05E-08
LINC01293	LncRNA	-2.83693	3.91E-10
LINC01085	LncRNA	-2.27758	5.04E-05
LINC00452	LncRNA	-2.0481	3.01E-04
CASC15	LncRNA	2.13179	3.86E-26
C7orf65	LncRNA	2.282585	5.18E-13
LINC01460	LncRNA	2.285909	3.32E-06
LINC00877	LncRNA	2.451322	1.46E-30
LINC00885	LncRNA	2.647319	1.01E-65
LINC01234	LncRNA	3.437637	3.48E-60
ZNF503-AS1	LncRNA	4.048393	1.79E-77

### PPI & functional analyses of mRNAs

The top 464 differentially expressed mRNAs were input into PPI network by Omicsbean (Figure 8A). Figure 8B shows that IL6 plays an important role in malignant transformation of human embryonic esophageal epithelial cells by HPV18 E6E7. It is the hub and connects multiple classic cancer pathways. Survival analysis was performed IL6 maybe serve as prognostic biomarkers (Figure 8C).

### Real-time QPCR verification

To verify the results of the RNA-Seq, we selected seven genes to corroborate the RNA-seq data using quantitative qRT-PCR. The 26th generation of SHEE cells (P26) was used as reference sample. The result shows qRT-PCR data have a strong correlation with the RNA-Seq data (Figure 9) and confirms that RNA-Seq results are reliable in this study.

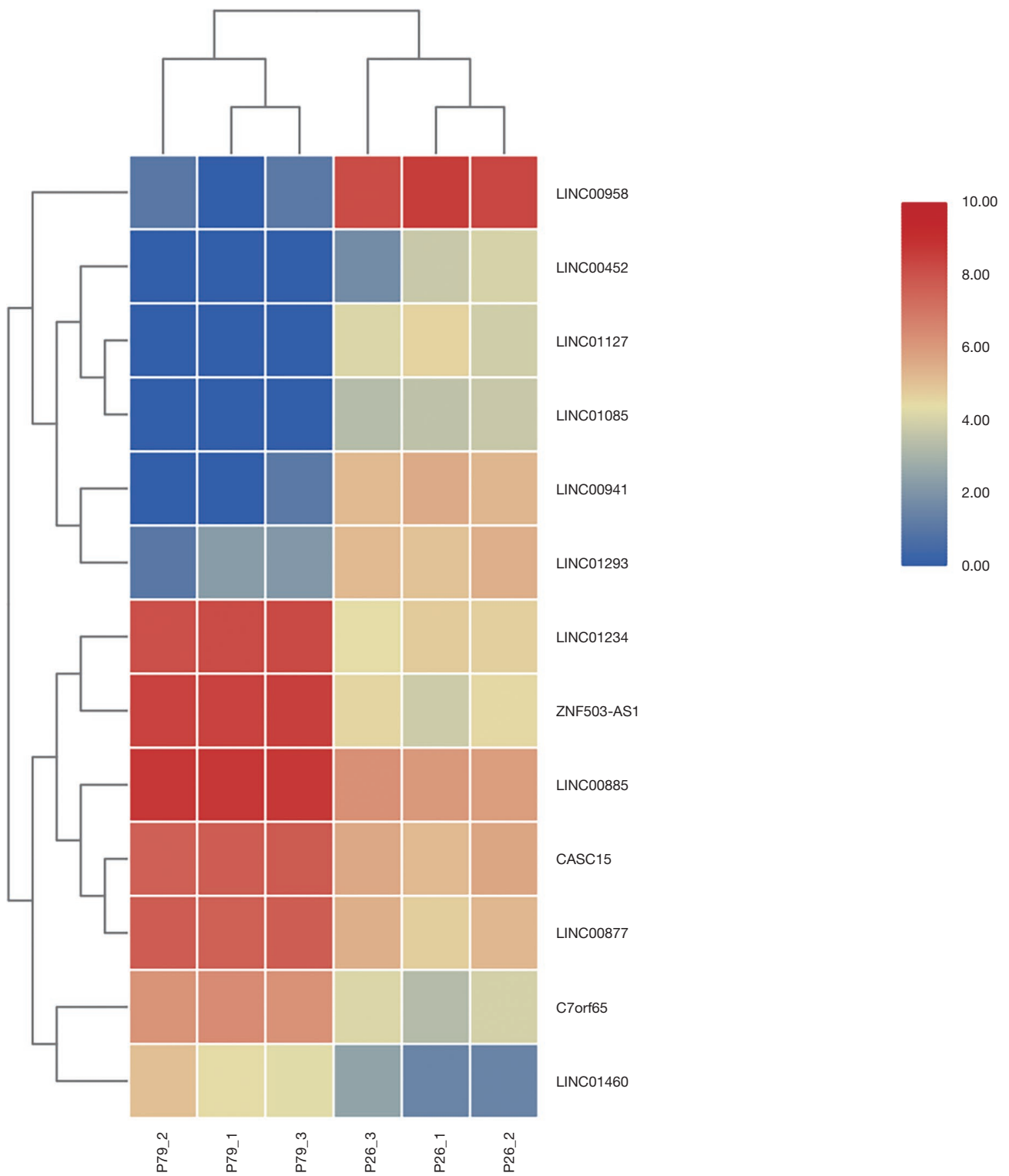
## Discussion

HPV is a small double-stranded DNA virus that can cause squamous epithelium proliferation in human skin mucosa and is closely related to the occurrence of numerous tumours. HPV is reported to be a necessary causative factor for cervical and anal-genital tumours (such as anal cancer, penile cancer and vulvar cancer) (16). However, the

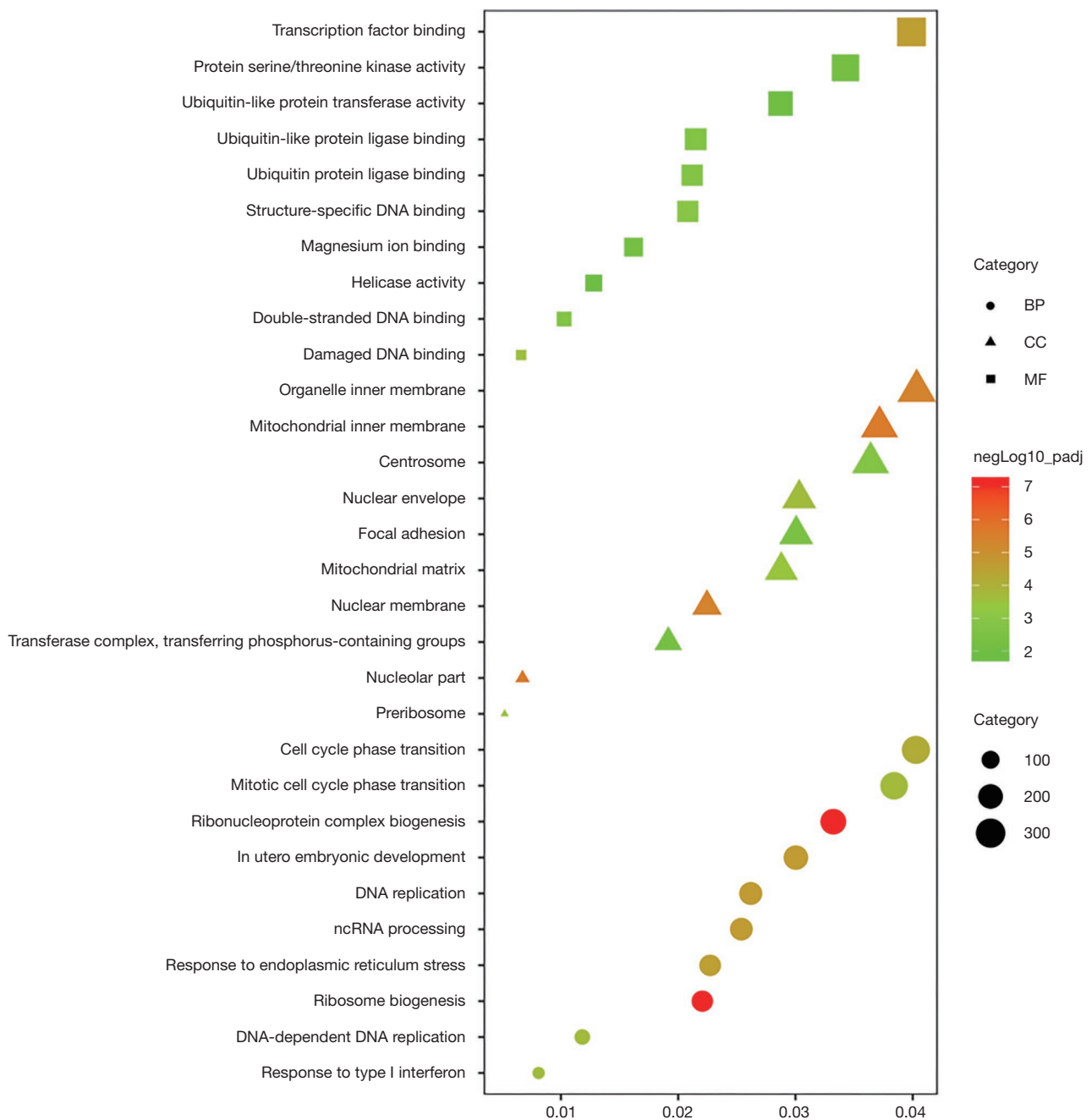
mechanism of HPV in cell transformation is still unclear. The SHEE cell line, which was established by this paper's authors, consists of immortalised epithelial cells from the embryonic oesophagus after being induced by the E6E7 genes of HPV 18. We previously demonstrated that carcinogenesis of esophageal epithelial cells induced by HPV was a multistage process that went through the initial, immortal, premalignant and malignant transformation stages (14,17-19). Therefore, this cell line may be useful in elucidating the carcinogenic mechanism of HPV. Considering the significant difference in differentiation between the 26th and 79th generations of the cells, we conducted a comparative transcriptomic study on these two generations of cells to analyse the global gene expression and the signal transduction pathways, hence exploring the molecular mechanism of malignant transformation of human embryonic esophageal epithelial cells by HPV18 E6E7. According to the results of the GO analysis (Figure 5), the most significant difference was ribosome biogenesis and ribonucleoprotein complex biogenesis. Ribosomes are complex organelles composed of different proteins and nucleic acids and are responsible for protein synthesis in living cells. Therefore, ribosome biogenesis is a very important biological process in cells. Dysregulated ribosome biogenesis plays an important role in the development and progression of most cancers. Upregulated ribosome biogenesis lead to downregulate the expression and activity of p53, thus promote tumor development. Coincidentally, p53 is also the target of E6 protein. At the same time, some studies have shown that the inhibition of ribosome biogenesis may provide a viable treatment for cancer treatment (20-23). The ribonucleoprotein complex is formed by the interaction of RNA and RNA-binding proteins in the cells and participates in almost all cellular life functions. The ribonucleoprotein complex, such as ribonucleic acid-binding motif protein 8A (RBM8A), has been shown to be involved in transcriptional translation, gene expression, cell cycle regulation and the apoptosis regulation of RNA, and its increased expression may play a carcinogenic role (24,25). Therefore, the imbalance in the biogenesis of the ribonucleoprotein complex plays an important role in cell malignant transformation.

Our analysis of DEGs in the current study demonstrated that the most significant genes involved MAGEA3 which is the most significant up-regulated gene and LINC00958 which is the most significant down-regulated gene. MAGEA3, as a member of the melanoma antigen gene family, has been reported as having a significant inhibitory





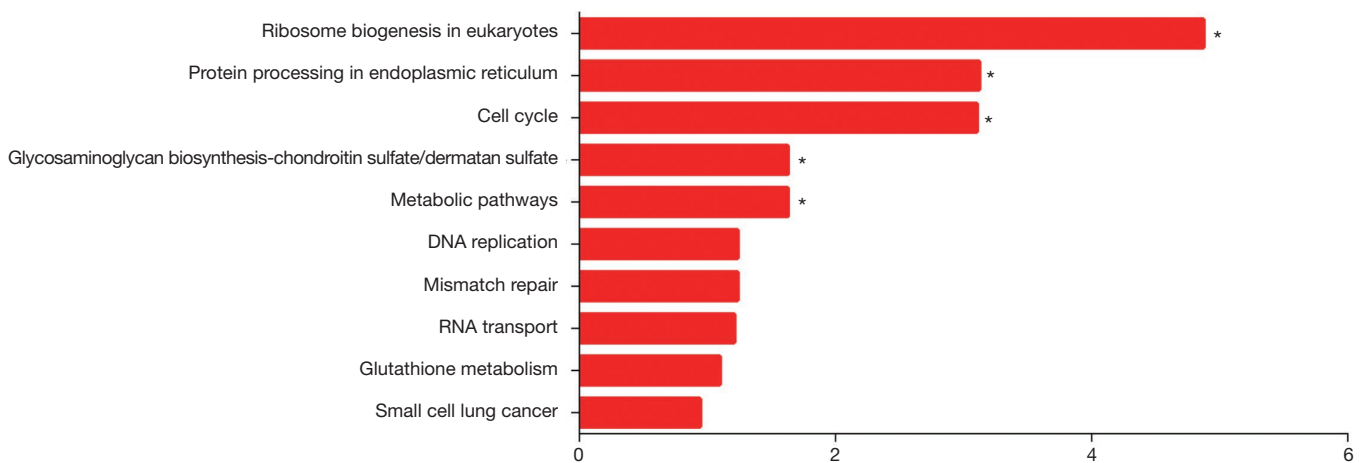
**Figure 4** A heatmap of differentially expressed lncRNAs with an absolute value >5 of log2FC and the padj <0.05.



**Figure 5** Gene ontology (GO) enrichment analysis of DEGs between SHEE P26 and P29.

effect on apoptosis. MAGEA3 will enhance the activity of specific E3 ubiquitin-synthase and AMP-activated protein kinase  $\alpha 1$  (AMPK $\alpha 1$ ) to reduce AMP-activated protein kinase (AMPK) and plays a role in promoting the growth of cancer cells and inhibiting apoptosis (26). Abikhair *et al.* reported that MAGEA3 is only expressed

in cancer and testis, so MAGEA3 has the potential to be a diagnostic marker (27). LINC00958 is a significant downregulated long-chain noncoding RNA in our analysis result. Few studies have been done on LINC00958, but it has been reported that cell viability is observed by silencing LINC00958 in bladder cancer. The migration

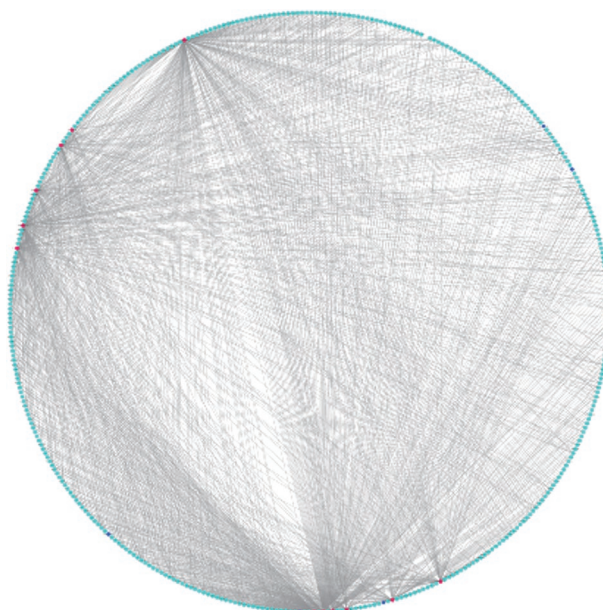


**Figure 6** The first 10 biological functions identified by KEGG analysis. \* for significant differences.

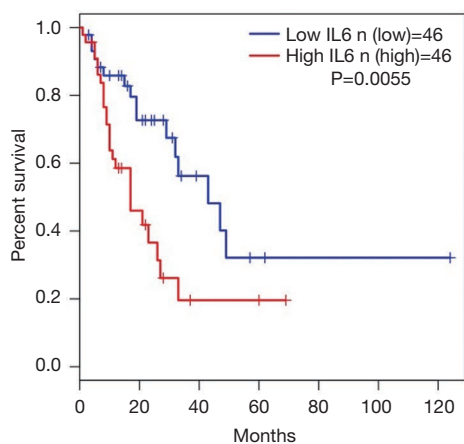
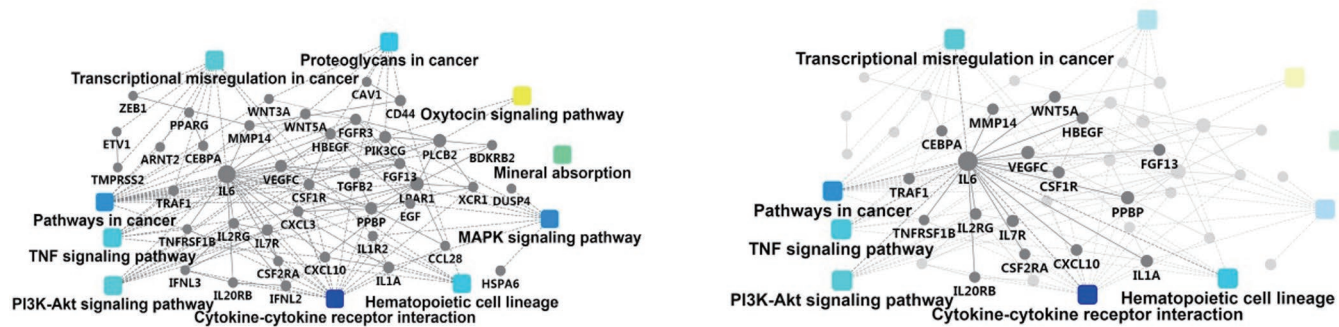
**Table 4** Gene expression in significant anomalous pathway in KEGG analysis

KEGG pathway	Count	P value	Genes
Ribosome biogenesis in eukaryotes↓	61	6.01E-08	<i>REXO2</i> ↑, <i>XRN2</i> ↑, <i>CSNK2A2</i> ↑, <i>NOL6</i> ↓, <i>NOP56</i> ↓, <i>DKC1</i> ↓, <i>WDR43</i> ↓, <i>GNL3</i> ↓, <i>TCOF1</i> ↓, <i>RCL1</i> ↓, <i>RPP25</i> ↓, <i>GTPBP4</i> ↓, <i>UTP6</i> ↓, <i>SNU13</i> ↓, <i>TBL3</i> ↓, <i>WDR36</i> ↓, <i>UTP15</i> ↓, <i>WDR3</i> ↓, <i>UTP14A</i> ↓, <i>RRP7A</i> ↓, <i>IMP4</i> ↓, <i>NXF1</i> ↓, <i>POP1</i> ↓, <i>RIOK1</i> ↓, <i>GNL2</i> ↓, <i>NAT10</i> ↓, <i>RAN</i> ↓, <i>NOP58</i> ↓, <i>CIRH1A</i> ↓, <i>GAR1</i> ↓, <i>MDN1</i> ↓, <i>UTP18</i> ↓, <i>HEATR1</i> ↓, <i>NXT1</i> ↓, <i>DROSHA</i> ↓, <i>BMS1</i> ↓, <i>NHP2</i> ↓, <i>REXO1</i> ↓, <i>NVL</i> ↓, <i>EIF6</i> ↓, <i>XPO1</i> ↓, <i>RBM28</i> ↓, <i>POP5</i> ↓, <i>RPP40</i> ↓, <i>FBL</i> ↓, <i>LSG1</i> ↓, <i>POP7</i> ↓, <i>RPP25L</i> ↓, <i>POP4</i> ↓, <i>RIOK2</i> ↓, <i>IMP3</i> ↓, <i>EFTUD1</i> ↓, <i>WDR75</i> ↓, <i>SPATA5</i> ↓, <i>TAF9</i> ↓, <i>GNL3L</i> ↓, <i>MPHOSPH10</i> ↓, <i>RPP30</i> ↓, <i>SBDS</i> ↓, <i>EMG1</i> ↓, <i>NMD3</i> ↓
Protein processing in endoplasmic reticulum↑	114	6.82E-06	<i>PDIA4</i> ↑, <i>MBTPS1</i> ↑, <i>CALR</i> ↑, <i>HSPA5</i> ↑, <i>HSP90B1</i> ↑, <i>PDIA3</i> ↑, <i>HERPUD1</i> ↑, <i>UBE2G1</i> ↑, <i>SEC24D</i> ↑, <i>LMAN1</i> ↑, <i>AMFR</i> ↑, <i>MARCH6</i> ↑, <i>CAPN1</i> ↑, <i>HYOU1</i> ↑, <i>XBP1</i> ↑, <i>RNF5</i> ↑, <i>MAPK10</i> ↑, <i>ERO1B</i> ↑, <i>EIF2AK1</i> ↑, <i>ERP29</i> ↑, <i>UGGT1</i> ↑, <i>ATXN3</i> ↑, <i>ATF4</i> ↑, <i>DDOST</i> ↑, <i>OS9</i> ↑, <i>MAN1A1</i> ↑, <i>NFE2L2</i> ↑, <i>P4HB</i> ↑, <i>PDIA6</i> ↑, <i>EDEM2</i> ↑, <i>SSR2</i> ↑, <i>UBE4B</i> ↑, <i>UBE2D1</i> ↑, <i>UBXN6</i> ↑, <i>FBXO2</i> ↑, <i>UBE2E1</i> ↑, <i>GANAB</i> ↑, <i>PRKCSH</i> ↑, <i>MAN1B1</i> ↑, <i>SEL1L</i> ↑, <i>WFS1</i> ↑, <i>DNAJC3</i> ↑, <i>DAD1</i> ↑, <i>YOD1</i> ↑, <i>ATF6B</i> ↑, <i>BCAP31</i> ↑, <i>TRAM1</i> ↑, <i>CAPN2</i> ↑, <i>SEC23A</i> ↑, <i>STT3B</i> ↑, <i>TUSC3</i> ↑, <i>SEC31A</i> ↑, <i>UBE2J1</i> ↑, <i>CUL1</i> ↑, <i>RPN2</i> ↑, <i>SEC24A</i> ↑, <i>EIF2AK3</i> ↑, <i>UFD1L</i> ↑, <i>SKP1</i> ↑, <i>ERLEC1</i> ↑, <i>SEC61G</i> ↑, <i>BAG1</i> ↓, <i>DNAJA1</i> ↓, <i>HSPA1A</i> ↓, <i>HSPH1</i> ↓, <i>HSP90AA1</i> ↓, <i>UBE2G2</i> ↓, <i>SEC23B</i> ↓, <i>HSPA8</i> ↓, <i>HSP90AB1</i> ↓, <i>PLAA</i> ↓, <i>BAG2</i> ↓, <i>VCP</i> ↓, <i>HSPBP1</i> ↓, <i>DNAJC1</i> ↓, <i>HSPA1B</i> ↓, <i>CKAP4</i> ↓, <i>EIF2S1</i> ↓, <i>MAP3K5</i> ↓, <i>UBE2J2</i> ↓, <i>DNAJB2</i> ↓, <i>RAD23A</i> ↓, <i>BAX</i> ↓, <i>UGGT2</i> ↓, <i>PPP1R15A</i> ↓, <i>DDIT3</i> ↓, <i>EIF2AK2</i> ↓, <i>UBE2D4</i> ↓, <i>SSR3</i> ↓, <i>DNAJB1</i> ↓, <i>SSR4</i> ↓, <i>STUB1</i> ↓, <i>UBE2D2</i> ↓, <i>HSPA6</i> ↓, <i>ERN1</i> ↓, <i>EDEM1</i> ↓, <i>TRAF2</i> ↓, <i>DNAJB12</i> ↓, <i>UBE2E2</i> ↓, <i>MBTPS2</i> ↓, <i>SIL1</i> ↓, <i>RPN1</i> ↓, <i>NGLY1</i> ↓, <i>RAD23B</i> ↓, <i>HSPA4L</i> ↓, <i>NSFL1C</i> ↓, <i>SEC31B</i> ↓, <i>VIMP</i> ↓, <i>BCL2</i> ↓, <i>PREB</i> ↓, <i>SSR1</i> ↓, <i>NPLOC4</i> ↓, <i>RBX1</i> ↓, <i>SEC61A2</i> ↓
Cell cycle↑	88	1.06E-05	<i>RBL2</i> ↑, <i>TFDP2</i> ↑, <i>TGFB2</i> ↑, <i>CDKN1B</i> ↑, <i>ABL1</i> ↑, <i>GADD45B</i> ↑, <i>CCNB2</i> ↑, <i>SMAD4</i> ↑, <i>CDC25C</i> ↑, <i>SKP2</i> ↑, <i>CDKN1C</i> ↑, <i>HDAC2</i> ↑, <i>BUB1</i> ↑, <i>CDK2</i> ↑, <i>CDC14A</i> ↑, <i>ATM</i> ↑, <i>ORC4</i> ↑, <i>CDKN2C</i> ↑, <i>PTTG1</i> ↑, <i>WEE2</i> ↑, <i>CDC14B</i> ↑, <i>STAG2</i> ↑, <i>CDC26</i> ↑, <i>GADD45G</i> ↑, <i>YWHAQ</i> ↑, <i>CUL1</i> ↑, <i>ORC2</i> ↑, <i>ANAPC2</i> ↑, <i>SKP1</i> ↑, <i>MYC</i> ↓, <i>CDC16</i> ↓, <i>CDC25B</i> ↓, <i>TFDP1</i> ↓, <i>ANAPC7</i> ↓, <i>CDC6</i> ↓, <i>CCND1</i> ↓, <i>MCM4</i> ↓, <i>CDKN1A</i> ↓, <i>GADD45A</i> ↓, <i>MCM3</i> ↓, <i>STAG1</i> ↓, <i>ANAPC5</i> ↓, <i>PKMYT1</i> ↓, <i>ORC1</i> ↓, <i>CDC25A</i> ↓, <i>MCM2</i> ↓, <i>PRKDC</i> ↓, <i>SFN</i> ↓, <i>ATR</i> ↓, <i>SMC1A</i> ↓, <i>E2F4</i> ↓, <i>ORC5</i> ↓, <i>PCNA</i> ↓, <i>CDK6</i> ↓, <i>E2F1</i> ↓, <i>CDC45</i> ↓, <i>MCM5</i> ↓, <i>CDC7</i> ↓, <i>YWHAH</i> ↓, <i>ORC6</i> ↓, <i>RBL1</i> ↓, <i>ORC3</i> ↓, <i>GSK3B</i> ↓, <i>CDK4</i> ↓, <i>CCNA2</i> ↓, <i>MAD1L1</i> ↓, <i>CDK1</i> ↓, <i>CCNE1</i> ↓, <i>MAD2L2</i> ↓, <i>E2F2</i> ↓, <i>ANAPC11</i> ↓, <i>CCNE2</i> ↓, <i>CCNH</i> ↓, <i>E2F5</i> ↓, <i>RBX1</i> ↓, <i>CDC23</i> ↓, <i>MAD2L1</i> ↓, <i>CREBBP</i> ↓, <i>YWHAZ</i> ↓, <i>BUB3</i> ↓, <i>ANAPC4</i> ↓, <i>CHEK2</i> ↓, <i>CDC20</i> ↓, <i>MCM7</i> ↓, <i>E2F3</i> ↓, <i>PLK1</i> ↓, <i>TTK</i> ↓, <i>TP53</i> ↓

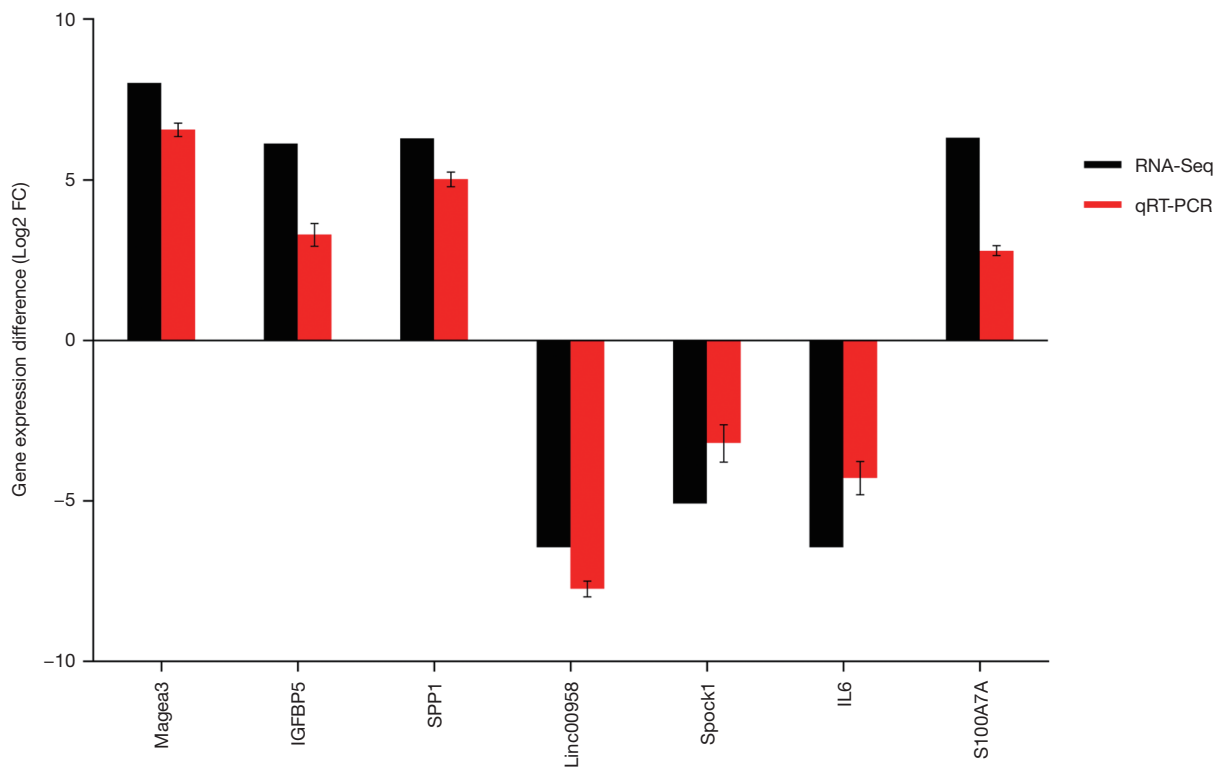
The arrows ↑ and ↓ indicate up-regulation or down-regulation.



**Figure 7** The lncRNA-miRNA-mRNA Competing endogenous RNA network. light blue nodes represent mRNAs, dark blue nodes represent lncRNAs and red nodes represent miRNAs.



**Figure 8** PPI network and Kaplan–Meier curve analysis. (A) PPI Network of differentially expressed mRNA; (B) the PPI Network of IL6 as a hub nodes; (C) Kaplan-Meier survival curves of IL6 in esophageal cancer.



**Figure 9** Correlation between the results of qRT-PCR and RNA-Seq. Black bars represent RNA-Seq measurements, and red bars represent qRT-PCR experimental results, calculated using the  $2^{-\Delta\Delta C_t}$  method and normalized to GAPDH.

ability was reduced, and LINC00958 was shown as being combined with the initiation of translation and RNA post-transcriptionally modified proteins and their involvement in regulation. The same results were obtained in a study of LINC00958 in glioma (28-30).

In this study, we used bioinformatics to analyze the expression and function of aberrantly expressed lncRNA, miRNA, and mRNA and establish ceRNA network and PPI network to predict the regulation mechanism of DEGs. Through ceRNA network analysis, we predicted an interesting regulatory mechanism, TP53TG1- miR-6835-IGF2. TP53TG1 is a recently identified lncRNA, study have shown that TP53TG1 may play an important role of tumor suppressor gene or oncogene in different tumors (31). As we know, E6 protein can inactivate tumor suppressor p53 by inducing its degradation (32). The association between TP53TG1 and p53 has been confirmed by CHIP (33,34). Insulin-like growth factor 2 (IGF2) is a mitogenic peptide hormone expressed by liver and many other tissues. IGF2 is highly expressed in many cancers and promotes the development of cancer. IGF is associated with multiple signaling pathways, such as MAPK signaling pathway, Ras

signaling pathway and PI3K-AKT signaling pathway (35-37). These signaling pathways are closely related to the development of cancer. IGF2 has been proved to promote proliferation and migration of esophageal cancer. Meanwhile, IGF2 is also an excellent esophageal cancer therapeutic targets and biomarkers (38,39). Therefore, we boldly speculate TP53TG1 may contribute to the malignant transformation of SHEEs by acting as a competing endogenous RNA to competitively bind to miR-6835 and regulate IGF2 expression.

IL6 is the most critical node in the PPI network. IL-6 is one of the most characteristic proinflammatory cytokines that activate JAK-STAT signaling pathway, MAPK signaling pathway, PI3K-Akt signaling pathway and other pathways. In the PI3K-Akt signaling pathway, IL-6 can promote the proliferation of cancer cells and inhibit apoptosis *in vitro* and *in vivo*. In the JAK-STAT pathway, IL6 functions by regulating gp130, STAT3 is considered to be the main signal transduction factor downstream of the gp130 signal transduction, acting as an oncogene and key factor in the link between inflammation and cancer (40-42). Simultaneously, Kaplan- Meier survival curves has shown IL6 maybe serve as prognostic biomarkers and therapy target in HPV-induced esophageal cancer.

## Conclusions

Using RNA-Seq, we identified DEGs between the 26th and 79th generations of SHEEs and found that the significantly DEGs were involved in the development of cancer. Our results indicate that the activations of multiple biological processes, cellular components and molecular function are involved in the dynamic progressive process of carcinogenesis of esophageal epithelial cells induced by HPV. Furthermore, we predicted TP53TG1-miR-6835-IGF2 and IL6 play an important role in HPV-induced malignant transformation of human embryonic esophageal epithelial cells.

## Acknowledgments

*Funding:* This work is supported by the National Natural Science Foundation of China (No.21677006).

## Footnote

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/tcr.2020.02.23>). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Institutional ethical approval and informed consent were waived.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

1. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. *CA Cancer J Clin* 2015;65:87-108.
2. Stoner GD, Gupta A. Etiology and chemoprevention of esophageal squamous cell carcinoma. *Carcinogenesis* 2001;22:1737-46.
3. Poljak M, Sterbenc A, Lunar MM. Prevention of human papillomavirus (HPV)-related tumors in people living with human immunodeficiency virus (HIV). *Expert Rev Anti Infect Ther* 2017;15:987-99.
4. Zhou ZX, Li D, Guan SS, et al. Immunotherapeutic Effects of Dendritic Cells Pulsed with a Codon-optimized HPV 16 E6 and E7 Fusion Gene in Vivo and in Vitro. *Asian Pac J Cancer Prev* 2015;16:3843-7.
5. Zhou ZX, Zhao C, Li QQ, et al. A novel mutant of human papillomavirus type 18 E6E7 fusion gene and its transforming activity. *Asian Pac J Cancer Prev* 2014;15:7395-9.
6. Zhang K, Li JT, Li SY, et al. Integration of human papillomavirus 18 DNA in esophageal carcinoma 109 cells. *World J Gastroenterol* 2011;17:4242-6.
7. Syrjanen K, Parkkinen S, Mantyjarvi R, et al. Human papillomavirus (HPV) type as an important determinant of the natural history of HPV infections in uterine cervix. *Eur J Epidemiol* 1985;1:180-7.
8. Cao B, Tian X, Li Y, et al. LMP7/TAP2 gene polymorphisms and HPV infection in esophageal carcinoma patients from a high incidence area in China. *Carcinogenesis* 2005;26:1280-4.
9. Han C, Qiao G, Hubbert NL, et al. Serologic association between human papillomavirus type 16 infection and esophageal cancer in Shaanxi Province, China. *J Natl Cancer Inst* 1996;88:1467-71.
10. Farhadi M, Tahmasebi Z, Merat S, et al. Human papillomavirus in squamous cell carcinoma of esophagus in a high-risk population. *World J Gastroenterol* 2005;11:1200-3.
11. Weston AC, Prolla JC. Association between esophageal squamous cell carcinoma and human papillomavirus detected by Hybrid Capture II assay. *Dis Esophagus* 2003;16:224-8.
12. Acevedo-Nuño E, Gonzalez-Ojeda A, Vazquez-Camacho G, et al. Human papillomavirus DNA and protein in tissue samples of oesophageal cancer, Barrett's oesophagus and oesophagitis. *Anticancer Res* 2004;24:1319-23.
13. Monsonog J. [Cellular and molecular pathogenesis of cancer of the cervix]. *Contracept Fertil Sex* 1995;23:731-40.
14. Shen ZY, Xu LY, Chen MH, et al. Progressive transformation of immortalized esophageal epithelial cells. *World J Gastroenterol* 2002;8:976-81.
15. Shen Z, Cen S, Shen J, et al. Study of immortalization and malignant transformation of human embryonic esophageal epithelial cells induced by HPV18 E6E7. *J Cancer Res Clin Oncol* 2000;126:589-94.
16. Herrero R, Castellsague X, Pawlita M, et al. Human papillomavirus and oral cancer: the International Agency for Research on Cancer multicenter study. *J Natl Cancer*

- Inst 2003;95:1772-83.
17. Shen ZY, Xu LY, Li C, et al. A comparative study of telomerase activity and malignant phenotype in multistage carcinogenesis of esophageal epithelial cells induced by human papillomavirus. *Int J Mol Med* 2001;8:633-9.
  18. Shen ZY, Xu LY, Li EM, et al. The multistage process of carcinogenesis in human esophageal epithelial cells induced by human papillomavirus. *Oncol Rep* 2004;11:647-54.
  19. Shen Z, Shen J, Zeng Y. Biological characteristics of human fetal esophageal epithelial cell line immortalized by the E6 and E7 gene of HPV type 18. *Zhonghua Shi Yan He Lin Chuang Bing Du Xue Za Zhi* 1999;13:209-12.
  20. Derenzini M, Montanaro L, Trere D. Ribosome biogenesis and cancer. *Acta Histochem* 2017;119:190-7.
  21. Pelletier J, Thomas G, Volarevic S. Ribosome biogenesis in cancer: new players and therapeutic avenues. *Nat Rev Cancer* 2018;18:51-63.
  22. Penzo M, Montanaro L, Trere D, et al. The Ribosome Biogenesis-Cancer Connection. *Cells* 2019;8. doi: 10.3390/cells8010055.
  23. Bianco C, Mohr I. Ribosome biogenesis restricts innate immune responses to virus infection and DNA. *Elife* 2019;8. doi: 10.7554/eLife.49551.
  24. Ishigaki Y, Nakamura Y, Tatsuno T, et al. Depletion of RNA-binding protein RBM8A (Y14) causes cell cycle deficiency and apoptosis in human cells. *Exp Biol Med (Maywood)* 2013;238:889-97.
  25. Liang R, Lin Y, Ye JZ, et al. High expression of RBM8A predicts poor patient prognosis and promotes tumor progression in hepatocellular carcinoma. *Oncol Rep* 2017;37:2167-76.
  26. Wu F, Liu F, Dong L, et al. miR-1273g silences MAGEA3/6 to inhibit human colorectal cancer cell growth via activation of AMPK signaling. *Cancer Lett* 2018;435:1-9.
  27. Abikhair M, Roudiani N, Mitsui H, et al. MAGEA3 Expression in Cutaneous Squamous Cell Carcinoma Is Associated with Advanced Tumor Stage and Poor Prognosis. *J Invest Dermatol* 2017;137:775-8.
  28. Guo E, Liang C, He X, et al. Long Noncoding RNA LINC00958 Accelerates Gliomagenesis Through Regulating miR-203/CDK2. *DNA Cell Biol* 2018;37:465-72.
  29. Seitz AK, Christensen LL, Christensen E, et al. Profiling of long non-coding RNAs identifies LINC00958 and LINC01296 as candidate oncogenes in bladder cancer. *Sci Rep* 2017;7:395.
  30. He W, Zhong G, Jiang N, et al. Long noncoding RNA BLACAT2 promotes bladder cancer-associated lymphangiogenesis and lymphatic metastasis. *J Clin Invest* 2018;128:861-75.
  31. Zhang Y, Yang H, Du Y, et al. Long noncoding RNA TP53TG1 promotes pancreatic ductal adenocarcinoma development by acting as a molecular sponge of microRNA-96. *Cancer Sci* 2019;110:2760-72.
  32. Li X, Coffino P. High-risk human papillomavirus E6 protein has two distinct binding sites within p53, of which only one determines degradation. *J Virol* 1996;70:4509-16.
  33. Diaz-Lagares A, Crujeiras AB, Lopez-Serra P, et al. Epigenetic inactivation of the p53-induced long noncoding RNA TP53 target 1 in human cancer. *Proc Natl Acad Sci U S A* 2016;113:E7535-44.
  34. Nikulenkov F, Spinnler C, Li H, et al. Insights into p53 transcriptional function via genome-wide chromatin occupancy and gene expression analysis. *Cell Death Differ* 2012;19:1992-2002.
  35. Pereira SS, Monteiro MP, Costa MM, et al. IGF2 role in adrenocortical carcinoma biology. *Endocrine* 2019;66:326-37.
  36. Cacheux W, Lievre A, Richon S, et al. Interaction between IGF2-PI3K axis and cancer-associated-fibroblasts promotes anal squamous carcinogenesis. *Int J Cancer* 2019;145:1852-9.
  37. Ji Y, Wang Z, Li Z, et al. Silencing IGF-II impairs C-myc and N-ras expressions of SMMC-7721 cells via suppressing FAK/PI3K/Akt signaling pathway. *Cytokine* 2017;90:44-53.
  38. Gao T, He B, Pan Y, et al. Long non-coding RNA 91H contributes to the occurrence and progression of esophageal squamous cell carcinoma by inhibiting IGF2 expression. *Mol Carcinog* 2015;54:359-67.
  39. Zheng DN, Zhang CJ, Sun GP. Long non-coding RNA MNX1-AS1 promotes migration and invasion of esophageal squamous cell carcinoma by upregulating IGF2. *Eur Rev Med Pharmacol Sci* 2019;23:6179-85.
  40. Landskron G, De la Fuente M, Thuwajit P, et al. Chronic inflammation and cytokines in the tumor microenvironment. *J Immunol Res* 2014;2014:149185.
  41. Taniguchi K, Karin M. IL-6 and related cytokines as the critical lynchpins between inflammation and cancer. *Semin Immunol* 2014;26:54-74.
  42. Nguyen DP, Li J, Tewari AK. Inflammation and prostate cancer: the role of interleukin 6 (IL-6). *BJU Int* 2014;113:986-92.

**Cite this article as:** Tang D, Wang B, Khodahemmati S, Li J, Zhou Z, Gao J, Sheng W, Zeng Y. A transcriptomic analysis of malignant transformation of human embryonic esophageal epithelial cells by HPV18 E6E7. *Transl Cancer Res* 2020;9(3):1818-1832. doi: 10.21037/tcr.2020.02.23