Integrated bioinformatics analysis to identify key genes related to the prognosis of esophageal squamous cell carcinoma

Ying Yang, Zhiwei Sun, Youwu Shi, Jing Sun, Xiaodong Zhang

The VIP-II Gastrointestinal Cancer Division of Medical Department, Key Laboratory of Carcinogenesis & Translational Research (Ministry of Education), Peking University Cancer Hospital & Institute, Beijing, China

Contributions: (I) Conception and design: Y Yang, X Zhang; (II) Administrative support: X Zhang; (III) Provision of study materials or patients: Y Yang, Z Sun, Y Shi, J Sun; (IV) Collection and assembly of data: Y Yang, Y Shi, J Sun; (V) Data analysis and interpretation: Y Yang, Z Sun; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Xiaodong Zhang. The VIP-II Gastrointestinal Cancer Division of Medical Department, Key Laboratory of Carcinogenesis & Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing, China. Email: zhangxd0829@163.com.

Background: Esophageal squamous cell carcinoma (ESCC) is a serious threat to human health and life. The National Center for Biotechnology Information Gene Expression Omnibus (NCBI-GEO) database provides valuable information on genes related to the pathogenesis and prognosis of ESCC, which helps us to make in-depth understanding about the disease and improve its prognosis.

Methods: Four microarray profiles [GSE77861 (African Americans), GSE26886 (Germans), GSE17351 (Americans), and GSE45670 (Chinese)] from the NCBI-GEO including 49 ESCC tissues and 41 corresponding normal tissues were collected. Integrated bioinformatics methods, including protein-protein interaction (PPI) network analysis, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses, and Kaplan-Meier plotter were applied to determine the differentially expressed genes (DEGs) in ESCC together with their core functions and relationship with survival.

Results: A total of 220 upregulated and 112 downregulated genes were identified as DEGs in ESCC, of which, 40 upregulated genes were core function genes. The DEGs were mostly involved in DNA replication and cell cycle pathways. Survival analysis and Bonferroni adjustment showed kinesin family member 18A (KIF18A) and TTK protein kinase (TTK) to be related to prognosis in ESCC.

Conclusions: The findings of the present study verified the previously proposed association between TTK and patient survival in ESCC, and identified KIF18A as ESCC prognosis-related gene markers for the first time. The underlying mechanism needs to be further investigated using larger sample size studies and biological experiments in future.

Keywords: Bioinformatics analysis; prognosis; esophageal squamous cell carcinoma (ESCC)

Submitted Nov 07, 2020. Accepted for publication Feb 26, 2021.
doi: 10.21037/tcr-20-3220
View this article at: http://dx.doi.org/10.21037/tcr-20-3220

Introduction

Esophageal cancer presents a serious threat to human health and life. In 2018, there were an estimated 572,034 new cases of esophageal cancer and 508,585 deaths, which saw this disease rank 7th for incidence and 6th for mortality out of all malignant tumors (1). In East Asia, esophageal squamous cell carcinoma (ESCC) is the major pathological type of esophageal cancer (2). Despite advancements in surgery, radiotherapy techniques, and systematic treatments, the prognosis of ESCC is still poor, with the 5-year survival rate standing at around 15% (3). Together with primary prevention, the promotion of early cancer screening programs, and the development of innovative therapy
strategies, a comprehensive understanding of the genetic changes of ESCC would aid in improving the prognostic outlook for patients. Acquiring in-depth knowledge of the pathogenesis, progression, and prognosis of ESCC would help to deepen our perceptions of the disease, stratify patients by prognosis, and provide the foundation for precision medicine and targeted therapies. However, previous reports on ESCC genetic susceptibility genes have been inconclusive (4-8). More investigations on genetic alterations are needed to improve our understanding of ESCC and explore its underlying mechanisms; specifically, analysis of differentially expressed genes (DEGs) is an effective approach to identifying genes of interest in the disease.

Gene chip is a reliable and effective technique to quickly identify DEGs (9). DEGs identified using this technique can be deposited in public databanks for future reference. Access to such open data and integrated information can provide researchers with worthwhile hints for novel study ideas.

To identify genetic factors relating to the pathogenesis and prognosis of ESCC, and to explore the potential mechanisms, we conducted an integrated bioinformatics analysis based on the National Center for Biotechnology Information Gene Expression Omnibus (NCBI-GEO) microarray database. Unlike the previous studies (4,5), we selected microarray datasets based on the same gene chip to reduce heterogeneity generated from experimental tests. Also, we selected gene profiles with no restrictions on patients' race, which might make our conclusion more generalizable. We present the following article in accordance with the MDAR checklist (available at http://dx.doi.org/10.21037/tcr-20-3220).

Methods

Microarray profiles

Six gene expression profiles were collected from that NCBI-GEO microarray database: GSE77861 (African Americans), GSE26886 (Germans), GSE17351 (Americans), and GSE100942, GSE45670, and GSE33810 (Chinese). Each of the 6 microarray profiles contained both ESCC and normal esophageal epitheliums tissues, and all expression data were built on GPL570 platform (HG-U133_Plus_2; Affymetrix Human Genome U133 Plus 2.0 Array).

Data processing of DEGs

GEO2R online tools were applied to identify DEGs between ESCC and normal esophageal epithelial tissues (10). The filter criteria included |logFC|>0 and adjusted P<0.05. Two Chinese datasets, GSE100942 and GSE33810, were excluded from further analysis, due to having an adjusted P value of ≥0.05. Finally, GSE77861, GSE26886, GSE17351, and GSE45670, including 49 cancer tissues and 41 normal tissues, were included in the genetic analysis. Common DEGs among the 4 microarray datasets were determined using a Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/). DEGs were considered to be upregulated if logFC >0, and downregulated if logFC <0.

PPI network analysis

The protein–protein interaction (PPI) network of the identified DEGs was constructed by estimating the probable interactions between the DEGs using the online instrument Search Tool for the Retrieval of Interacting Genes (STRING) in Cytoscape under the conditions of maximum number of interactions 0 and confidence score ≥0.4 (11,12). The Cytoscape Molecular Complex Detection app was applied to build modules of the PPI network, with the degree cutoff set at 2, maximum depth at 100, k-core at 2, and node score cutoff at 0.2.

Gene Ontology (GO) and pathway enrichment analysis

Based on high-throughput transcriptomic or genomic data, we used the bioinformatics tool Database for Annotation, Visualization and Integrated Discovery (DAVID) to conduct a GO analysis of hub DEGs to analyze their genetic features, including molecular functions (MFs), cellular components (CCs), and biological processes (BPs) (Bonferroni-corrected P<0.05/counts) (13,14). DAVID was also employed to visualize Kyoto Encyclopedia of Genes and Genomes (KEGG) biological pathways (P<0.05) (15,16).

Survival analysis

Associations between the identified hub genes and ESCC survival in public genome databases, including the European Genome-phenome Archive (EGA), The Cancer Genome Atlas (TCGA), and GEO (Affymetrix microarrays only), were evaluated using the online Kaplan-Meier plotter (17). The results were presented as hazard ratios (HRs) with 95% confidence intervals (CIs) and as the log-rank P value. Because of the large number of genes to be tested,
we applied both Bonferroni-corrected P<0.05/the number of genes and false discovery rate (FDR) <0.2 to adjust for multiple comparisons.

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). All data used in our study are available in public databases, including the NCBI-GEO microarray database, EGA, and TCGA. Individual consent for this analysis was waived.

Results

A total of 2,720 upregulated genes and 2,725 downregulated genes were obtained from GSE77861 (African Americans); 15,133 upregulated genes and 6,263 downregulated genes were obtained from GSE26886 (Germans); 567 upregulated genes and 301 downregulated genes were obtained from GSE17351 (Americans); and 6,428 upregulated genes and 5,572 downregulated genes were obtained from GSE45670 (Chinese). The DEGs identified from 49 cancer tissues and 41 normal matching tissues in the above 4 microarray profiles were compared. After the exclusion of unknown genes, 332 common DEGs, including 220 upregulated genes and 112 downregulated genes, were identified (Figure 1 and Table 1).

STRING was used to identify the nodes with the most connections as core function genes (Figure S1). As shown in Figure 2, there were 40 hub genes, all of which were upregulated.

The 40 core DEGs were subjected to GO and KEGG pathway enrichment analyses. In terms of BPs, the DEGs were significantly enriched in cell division, mitotic nuclear division, DNA replication and sister chromatid cohesion, chromosome segregation, DNA replication initiation, and G1/S transition of mitotic cell cycle. In regard to CCs, the DEGs were significantly enriched in the nucleoplasm, nucleus, chromosome, centromeric region, midbody, and spindle midzone. In terms of MFs, the DEGs were mostly involved in ATP binding, protein binding, ATP-dependent microtubule motor activity, plus-end-directed, and chromatin binding (Table 2).

The KEGG pathway enrichment analysis revealed that the DEGs were mostly involved in the DNA replication and cell cycle pathways. The implicated genes included replication factor C subunit 4 (RFC4), minichromosome maintenance complex component 2 (MCM2), MCM5, MCM6, cell division cycle 6 (CDC6), and TTK protein kinase (TTK) (Figure 3A,B; Table 3).

Finally, we conducted a survival analysis of 39 identified core DEGs, as PICALM interacting mitotic regulator (E4M64) was excluded due to a lack of survival data. With P<0.05 indicating statistical significance, 15 genes were found to be statistically significantly related to survival and are shown in Figure 4. The survival curves of the other 24 genes are presented in Figure S2.

The GO analysis of the 15 significant genes revealed that the most enriched BP, CC, and MF were cell division, nucleoplasm, and ATP binding, respectively (Table 4). KEGG pathway enrichment analysis again showed enrichment mainly in the cell cycle pathway (Figure 3B and Table 5).
Table 1 Common differentially expressed genes (DEGs) identified from 4 microarray datasets of esophageal squamous cell carcinoma tissues compared with normal esophageal tissues (n=332)

| DEGs          | Gene name                                                                 |
|---------------|---------------------------------------------------------------------------|
| Upregulated   | KRT17, IPO9, CDC25B, NUDT1, FOXM1, PLXNA1, DDX18, KIAA0930, KIF4A, MEX3D, MCM5, TMEM132A, MIR1292, CD276 |
| Downregulated | C4orf3, CPEB3, ALAD, PAQR8, UFL1, TFAP2B, RSU1, CHMP2B, CUL4B, KALRN, CALCOCO2, FBXL5, RBP7 |

For multiple comparisons, we applied a Bonferroni correction of 0.05/39 = 0.001282 as Bonferroni-adjusted statistical significance. We also applied FDR < 0.20 as a second reference. Eventually, 2 genes were identified as being significantly related to survival: kinesin family member 18A (KIF18A) (HR = 0.24, 95% CI: 0.10–0.59, log-rank P = 0.00072, FDR = 0.10) and TTK (HR = 0.20, 95% CI: 0.07–0.55, log-rank P = 0.00057, FDR = 0.10). Patients with a high expression of these genes had a better prognosis than those with a low expression.

Discussion

By performing a gene enrichment bioinformatics analysis of ESCC which did not factor in race, we identified 332 common DEGs from 4 microarray datasets (GSE77861, GSE26886, GSE17351, and GSE45670). Through module analysis, we identified 40 upregulated genes as core function genes, and the most involved pathways were the DNA replication and cell cycle pathways, which supported our previous knowledge of carcinogenesis. In the prognostic analysis, we identified 15 genes associated with survival. For these survival-related genes, the most enriched BP, CC, and MF were cell division, nucleoplasm, and ATP binding, respectively, and the most enriched pathway was the cell cycle pathway. Furthermore, for multiple comparisons, we applied the Bonferroni correction and FDR, and identified 2 genes, KIF18A (HR = 0.24, 95% CI: 0.10–0.59, log-rank P = 0.00072, FDR = 0.10) and TTK (HR = 0.20, 95% CI: 0.07–0.55, log-rank P = 0.00057, FDR = 0.10), as prognosis-related genes.

Although our findings present associations between KIF18A/TTK and ESCC development and prognosis, the exact mechanisms still are unclear, especially as research on these aspects of ESCC is lacking. A number of in vitro studies have explored the functional role of TTK in cancer, and have discovered the major pathways involved to be the TTK–AKT–mechanistic target of rapamycin kinase (mTOR) pathway in hepatocellular carcinoma and TTK/AKT interaction with B-Raf/mitogen-activated protein kinase.
kinase 1 (MAPK1) signals in melanoma (18,19). KIF18A, KIF18B, and KIF19 are kinesin-8 family members. KIF18A together with dyneins, that are as the key component of canonical plus-end stepping in the mitotic spindle and the depolymerization of microtubules, is primarily involved in mitotic chromosome alignment, maintenance of chromosomal stability, and cell division. It also acts as the major driving force for cell migration (20-26). KIF18A dysregulation may give rise to genomic instability (27), eventually resulting in neoplasm development and metastatic progression (28).

Earlier studies have shown that the overexpression of human KIF18A is related to a poor prognosis of colorectal cancer (20), invasive breast cancer (29), and primary hepatocellular carcinoma (30). However, studies on KIF18A in ESCC are scarce. To the best of our knowledge, the present study is the first to reveal the relationship of KIF18A with the prognosis of ESCC. Taken together, previously published findings (20,29,30) and the results of the present study imply that KIF18A is a possible prognostic predictor; however, the molecular mechanisms need further elucidation.

In terms of treatment, the mechanism and molecules relevant to mitosis are well-established targets for microtubule-binding chemotherapeutic agents, such as taxol and vinca alkaloids (31), which is consistent with the fact that regimens containing paclitaxel are the preferred chemotherapeutic treatment for ESCC in Chinese clinical practice (32). Additionally, mitotic kinesins have also been considered as potentially valuable targets in the development of anticancer drugs (33).

TTK is a cancer/testis antigen (CTA) and an HLA-A2402-restricted epitope peptide (34). As well as in the testes, early-evolving embryos, the thymus, and placentas, the expression of CTA can be detected in various malignant tumors, and it is rarely detected in benign tissues (35,36). TTK participates in cell proliferation and migration through the phosphorylation of multiple amino acids, such as serine, threonine, and tyrosine hydroxyamino acids, and functions to maintain genomic integrity through regulating the spindle assembly checkpoint (35,37-39).

Previous studies have revealed the associations between TTK overexpression and the growth of pancreatic ductal adenocarcinoma (40), hepatocellular carcinoma (18), breast cancer (41), and melanoma (19). A previously published study found that TTK was a key component in ESCC.
## Table 2 Gene Ontology analysis of 40 core function differentially expressed genes in esophageal squamous cell carcinoma

| Category | Term | Count | % | Genes | Bonferroni P value | FDR |
|----------|------|-------|---|-------|--------------------|-----|
| **GOTERM_BP_DIRECT** | GO:0051301 - cell division | 18 | 45 | KIF14, CDC6, CKS1B, NUF2, KNTC1, TPX2, KIF18B, CENPF, AURKA, UBE2C, MCM5, NCAPH, FAM64A, SPAG5, CDC2, CDC5, HELLS, CDC3 | 1.59E-16 | 7.75E-16 |
| | GO:0007067 - mitotic nuclear division | 13 | 32.5 | CENPN, CDC6, FAM64A, KNTC1, NUF2, TPX2, CDC2, CENPF, AURKA, AURKB, CDC5, HELLS, CDC3 | 6.10E-11 | 2.98E-10 |
| | GO:0006260 - DNA replication | 9 | 22.5 | CDC6, RFC4, DTL, MCM2, CHAF1A, MCM10, MCM5, MCM6, DSCC1 | 5.71E-07 | 2.79E-06 |
| | GO:0007062 - sister chromatid cohesion | 7 | 17.5 | CENPN, KIF18A, KNTC1, NUF2, CENPF, AURKB, CDC5 | 3.42E-05 | 1.67E-04 |
| | GO:0007059 - chromosome segregation | 6 | 15 | CENPN, SPAG5, NUF2, CDC2, CENPF, TOP2A | 1.31E-04 | 6.38E-04 |
| | GO:0006270 - DNA replication initiation | 5 | 12.5 | CDC6, MCM2, MCM10, MCM5, MCM6 | 2.29E-04 | 0.00112 |
| | GO:0000082 - G1/S transition of mitotic cell cycle | 6 | 15 | CDC6, MCM2, MCM10, MCM5, MCM6 | 9.84E-04 | 0.004811 |
| **GOTERM_CC_DIRECT** | GO:0005654 - nucleoplasm | 26 | 65 | CKS1B, KIF4A, FOXM1, AURKA, AURKB, MCM10, MYBL2, GTSE1, CDC5, TOP2A, CDC2, HELLS, TRIP13, KIF14, CDC6, RAD51AP1, DTL, TPX2, CENPF, ATAD2, MCM2, UBE2C, MCM5, MCM6, RFC4, SPAG5, UBE2T, DSCC1 | 5.52E-10 | 6.61E-09 |
| | GO:0005634 - nucleus | 29 | 72.5 | FOXM1, KNTC1, AURKA, AURKB, MCM10, NCAPF, TOP2A, CDC5, HELLS, TRIP13, KIF14, CDC6, RAD51AP1, DTL, KIF18A, TPX2, NUF2, KIF18B, ATAD2, CENPF, MCM2, ECT2, MCM5, MCM6, UHRF1, FAM64A, SPAG5, CHAF1A, UBE2T | 6.32E-06 | 7.57E-05 |
| | GO:0007775 - chromosome, centromeric region | 6 | 15 | CENPN, NUF2, CENPF, CDC5, HELLS, DSCC1 | 1.21E-05 | 1.45E-04 |
| | GO:0030496 - midbody | 7 | 17.5 | KIF14, KIF4A, SPAG5, CENPF, AURKA, AURKB, ECT2 | 2.74E-05 | 3.29E-04 |
| | GO:0051233 - spindle midzone | 4 | 10 | KIF14, CDC6, AURKA, AURKB | 7.80E-04 | 0.009348 |
| **GOTERM_MF_DIRECT** | GO:0005524 - ATP binding | 19 | 47.5 | KIF14, CDC6, KIF4A, KIF18A, TPX2, KIF18B, TTK, ATAD2, AURKA, AURKB, MCM2, UBE2C, MCM5, MCM6, RFC4, TOP2A, UBE2T, HELLS, TRIP13 | 2.80E-08 | 3.72E-07 |
| | GO:0005515 - protein binding | 36 | 90 | CKS1B, KIF4A, FOXM1, KNTC1, TTK, AURKA, AURKB, MYBL2, MCM10, GTSE1, NCAPF, TOP2A, CDC5, HELLS, CDC6, TRIP13, KIF14, CDC6, RAD51AP1, DTL, TPX2, KIF18A, NUF2, CENPF, KIF18B, MCM2, UBE2C, ECT2, MCM5, MCM6, UHRF1, FAM64A, RFC4, SPAG5, CHAF1A, DSCC1 | 1.49E-06 | 1.98E-05 |
| | GO:0008574 - ATP-dependent microtubule motor activity, plus-end-directed | 4 | 10 | KIF14, KIF4A, KIF18A, KIF18B | 5.16E-04 | 0.00686 |
| | GO:0003682 - chromatin binding | 8 | 20 | CENPF, ATAD2, CHAF1A, CDC5, TOP2A, UBE2T, MCM5, HELLS | 0.001531 | 0.020359 |

FDR, false discovery rate.
progression and predicted a poor prognosis, although the molecular mechanisms were unclear (4).

TTK inhibitors suppress the activity of monopolar spindle 1 kinase, specifically deactivate the spindle assembly checkpoint, which results in chromosome separation dysfunction, heteroploidy, and ultimately, cell death. Thus, TTK inhibition has arisen as a promising therapeutic strategy for triple-negative breast cancer and malignant
1.0
0.8
0.6
0.4
0.2
0.0

Expression

low
high

Number at risk
low
43
35
12
5
3
1
1
0

high
43
35
12
5
3
1
1
0

HR = 0.43 (0.19-0.98)
logrank P = 0.038

A
ATAD2

2.0
1.0
0.5
0.0

Expression

low
high

Number at risk
low
42
32
13
3
2
0
0
0

high
42
32
13
3
2
0
0
0

HR = 0.32 (0.13-0.82)
logrank P = 0.013

B
AURKA

1.0
0.8
0.6
0.4
0.2
0.0

Expression

low
high

Number at risk
low
27
20
6
1
1
0
0
0

high
54
44
15
7
4
2
1
0

HR = 0.29 (0.13-0.67)
logrank P = 0.0022

C
CDC6

1.0
0.8
0.6
0.4
0.2
0.0

Expression

low
high

Number at risk
low
19
14
3
2
2
0
0
0

high
82
53
18
6
3
2
1
0

HR = 0.39 (0.16-0.92)
logrank P = 0.027

D
CDCA3

1.0
0.8
0.6
0.4
0.2
0.0

Expression

low
high

Number at risk
low
19
13
3
1
1
0
0
0

high
82
51
18
7
4
2
1
0

HR = 0.44 (0.19-1.01)
logrank P = 0.047

E
CKS1B

1.0
0.8
0.6
0.4
0.2
0.0

Expression

low
high

Number at risk
low
21
14
5
2
1
0
0
0

high
80
50
16
6
4
2
1
0

HR = 0.43 (0.19-0.99)
logrank P = 0.04

F
GTSE1

1.0
0.8
0.6
0.4
0.2
0.0

Expression

low
high

Number at risk
low
31
22
7
2
1
0
0
0

high
50
42
14
6
4
2
1
0

HR = 0.24 (0.1-0.59)
logrank P = 0.00172

G
KIF1A

1.0
0.8
0.6
0.4
0.2
0.0

Expression

low
high

Number at risk
low
19
12
1
0
0
0
0
0

high
82
52
20
8
5
2
1
0

HR = 0.38 (0.14-1.03)
logrank P = 0.049

H
KIF1B

1.0
0.8
0.6
0.4
0.2
0.0

Expression

low
high

Number at risk
low
60
45
14
6
4
2
1
0

high
21
19
7
2
1
0
0
0

HR = 0.2 (0.05-0.84)
logrank P = 0.015

I
MCM6
Table 4 Gene Ontology analysis of 15 differentially expressed genes related to esophageal squamous cell carcinoma survival

| Category            | Term                          | Count | %     | Genes                                                                 | Bonferroni P value | FDR    |
|---------------------|-------------------------------|-------|-------|----------------------------------------------------------------------|-------------------|--------|
| GOTERM_BP            | GO:0051301–cell division      | 8     | 53.33333 | CKS1B, CDC6, NCAPH, SPAG5, TPX2, KIF18B, AURKA, CDCA3               | 5.17E-07          | 5.47E-06 |
| DIRECT              |                               |       |       |                                                                      |                   |        |
| GOTERM_CC            | GO:0005654–nucleoplasm        | 10    | 66.66667 | CKS1B, CDC6, RAD51AP1, SPAG5, TPX2, ATAD2, AURKA, MCM10, GTSE1, MCM6 | 0.002063          | 0.041061 |
| DIRECT              |                               |       |       |                                                                      |                   |        |
| GOTERM_MF            | GO:0005524–ATP binding        | 8     | 53.33333 | CDC6, KIF18A, TPX2, KIF18B, ATAD2, TTK, AURKA, MCM6                 | 0.003221          | 0.075191 |
| DIRECT              |                               |       |       |                                                                      |                   |        |

FDR, false discovery rate.

mesothelioma (37,42-46). Furthermore, the testis is an immunoprivileged organ, and CTAs are ideal candidates as tumor-related antigens (47). TTK has emerged as a valuable target for tumor immunotherapy, including cancer vaccines, chimeric antigen receptor T-cell immunotherapy, and immune checkpoint inhibitors, due to its cancerspecific expression and strong in vivo immunogenicity, together with its critical roles in cell proliferation and...
migration, and spindle assembly checkpoint regulation (42,47,48). In several pharmaceutical clinical trials on ESCC, TTK expression was reported as both a marker for the treatment response and a potential cancer vaccine target (49,50).

Our study has several limitations. First, a previous published study conducted using the laiurger microarray dataset (GSE38129 and GSE20347) recognized that the TPX2 microtubule nucleation factor, cyclin dependent kinase 1, and centrosomal protein 55 were associated with relapse-free survival (51). Instead, we selected another microarray dataset built on GPL570 platform, including diverse races, and chose overall survival as our study endpoint rather than relapse-free survival, which resulted in our conclusion being inconsistent with those of earlier studies. However, the discrepancy between our research and previous research revealed that the bioinformatics analysis results were unstable and emphasized the significance of molecular mechanism studies. Second, the survival data on ESCC were limited, and the small sample size reduced the reliability of the results; therefore, larger sample sizes are required in future research. Third, for both KIF18A or TTK, molecular mechanism studies were limited and none of them focused on ESCC. It should be recognized that gene enrichment analysis can only provide clues, and our findings should be further verified by biological experiments. Finally, bioinformatics analysis is a growing field. New analyses and gene expression microarray studies are in development, which might affect our conclusions. The challenge for researchers performing bioinformatics analyses is to integrate up-to-date analytical methods and datasets accordingly.

In conclusion, our analysis was the first to identify KIF18A related to ESCC prognosis and verified the previous association between TTK and ESCC survival. KIF18A is a key component of mitotic activity, and TTK is a good candidate for tumor immunotherapy. The results of the present study were in accordance with those of studies on the clinical use of paclitaxel and also shed light on directions for future anticancer immunotherapy research and development. However, our conclusions need to be further validated in studies with larger sample sizes and biological experiments.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the MDAR checklist. Available at http://dx.doi.org/10.21037/tcr-20-3220

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tcr-20-3220). 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). All data used in our study were available in public database, and individual consent for this analysis was 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/.

Table 5 Kyoto Encyclopedia of Genes and Genomes pathway analysis of 15 differentially expressed genes related to esophageal squamous cell carcinoma

| Term                        | Count | %  | P value    | Genes       | FDR          |
|-----------------------------|-------|----|------------|-------------|--------------|
| hsa04110: cell cycle        | 3     | 20 | 0.003111   | CDC6, TTK, MCM6 | 1.480007    |

FDR, false discovery rate.
References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.

2. Abnet CC, Arnold M, Wei WQ. Epidemiology of Esophageal Squamous Cell Carcinoma. Gastroenterology 2018;154:360-73.

3. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66:115-32.

4. He W, Chen L, Yuan K, et al. Gene set enrichment analysis and meta-analysis to identify six key genes regulating and controlling the prognosis of esophageal squamous cell carcinoma. J Thorac Dis 2018;10:5714-26.

5. Chen FF, Zhang SR, Peng H, et al. Integrative genomics analysis of hub genes and their relationship with prognosis and signaling pathways in esophageal squamous cell carcinoma. Mol Med Rep 2019;20:3649-60.

6. Song G, Xu J, He L, et al. Systematic profiling identifies PDLIM2 as a novel prognostic predictor for oesophageal squamous cell carcinoma (ESCC). J Cell Mol Med 2019;23:5751-61.

7. Yang Q, Lin W, Liu Z, et al. RAP80 is an independent prognosis biomarker for the outcome of patients with esophageal squamous cell carcinoma. Cell Death Dis 2018;9:146.

8. Sang M, Gu L, Liu F, et al. Prognostic Significance of MAGE-A11 in Esophageal Squamous Cell Carcinoma and Identification of Related Genes Based on DNA Microarray. Arch Med Res 2016;47:151-61.

9. Vogelstein B, Papadopoulos N, Velculescu VE, et al. Cancer genome landscapes. Science 2013;339:1546-58.

10. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics (Oxford, England) 2007;23:1846-7.

11. Damian S, Andrea F, Stefan W, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43:D447-52.

12. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research 2003;13:2498-504.

13. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44-57.

14. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25-9.

15. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27-30.

16. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009;37:1-13.

17. Szász AM, Lánzczky A, Nagy Á, et al. Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. Oncotarget 2016;7:49322-33.

18. Liu X, Liao W, Yuan Q, et al. TTK activates Akt and promotes proliferation and migration of hepatocellular carcinoma cells. Cancer Lett 2013;337:125-32.

19. Nagahara M, Nishida N, Iwatsuki M, et al. Kinesin 18A expression: clinical relevance to colorectal cancer progression. Int J Cancer 2011;129:2543-52.

20. Mayr MI, Hümmer S, Bormann J, et al. The human kinesin Kif18A is a motile microtubule depolymerase essential for chromosome congression. Curr Biol 2007;17:488-98.

21. Stumpff J, von Dassow G, Wagenbach M, et al. The kinesin-8 motor Kif18A suppresses kinetochore movements to control mitotic chromosome alignment. Dev Cell 2008;14:252-62.

22. Weaver LN, Ems-McClung SC, Stout JR, et al. Kif18A uses a microtubule binding site in the tail for plus-end localization and spindle length regulation. Curr Biol 2011;21:1500-6.

23. Gardner MK, Odde DJ, Bloom K. Kinesin-8 molecular motors: putting the brakes on chromosome oscillations. Trends Cell Biol 2008;18:307-10.

24. Kim H, Fonseca C, Stumpff J. A unique kinesin-8 surface loop provides specificity for chromosome alignment. Mol Biol Cell 2014;25:3319-29.

25. Locke J, Joseph AP, Peña A, et al. Structural basis of human kinesin-8 function and inhibition. Proc Natl Acad Sci U S A 2017;114:E9359-E48.

26. Cleveland DW, Mao Y, Sullivan KF. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. Cell 2003;112:407-21.

27. Hartwell LH, Kastan MB. Cell cycle control and cancer. Science 1994;266:222-6.

28. Kasahara M, Nagahara M, Nakagawa T, et al. Clinicopathological relevance of kinesin family member
18A expression in invasive breast cancer. Oncol Lett 2016;12:1909-14.
30. Liao W, Huang G, Liao Y, et al. High KIF18A expression correlates with unfavorable prognosis in primary hepatocellular carcinoma. Oncotarget 2014;5:10271-9.
31. Jordan MA, Wilson L. Microtubules as a target for anticancer drugs. Nat Rev Cancer 2004;4:253-65.
32. Kim JY, Do YR, Park KU, et al. A multi-center phase II study of docetaxel plus cisplatin as first-line therapy in patients with metastatic squamous cell esophageal cancer. Cancer Chemother Pharmacol 2010;66:31-6.
33. Rath O, Kozielski E. Kinesins and cancer. Nat Rev Cancer 2012;12:527-39.
34. Mizukami Y, Kono K, Daigo Y, et al. Detection of novel cancer-testis antigen-specific T-cell responses in TIL, regional lymph nodes, and PBL in patients with esophageal squamous cell carcinoma. Cancer Sci 2008;99:1448-54.
35. Mills GB, Schmandt R, McGill M, et al. Expression of TTK, a novel human protein kinase, is associated with cell proliferation. J Biol Chem 1992;267:16000-6.
36. Whitehurst AW. Cause and consequence of cancer/testis antigen activation in cancer. Annu Rev Pharmacol Toxicol 2014;54:251-72.
37. Thu KL, Silvester J, Elliott MJ, et al. Disruption of the anaphase-promoting complex confers resistance to TTK inhibitors in triple-negative breast cancer. Proc Natl Acad Sci U S A 2018;115:E1570-7.
38. Tao Y, Chai D, Ma L, et al. Identification of distinct gene expression profiles between esophageal squamous cell carcinoma and adjacent normal epithelial tissues. Tohoku J Exp Med 2012;226:301-11.
39. King JL, Zhang B, Li Y, et al. TTK promotes mesenchymal signaling via multiple mechanisms in triple negative breast cancer. Oncogenesis 2018;7:69.
40. Slee RB, Grimes BR, Bansal R, et al. Selective inhibition of the spindle assembly checkpoint kinase TTK enhances the efficacy of docetaxel in a triple-negative breast cancer model. Ann Oncol 2015;26:2180-92.
41. Görffy B, Bottai G, Lehmann-Che J, et al. TP53 mutation-correlated genes predict the risk of tumor relapse and identify MPS1 as a potential therapeutic kinase in TP53-mutated breast cancers. Mol Oncol 2016;10:808-19.
42. Mason JM, Wei X, Fletcher GC, et al. Functional characterization of CFI-402257, a potent and selective Mps1/TTK kinase inhibitor, for the treatment of cancer. Proc Natl Acad Sci U S A 2017;114:3127-32.
43. Riggs JR, Nagy M, Elsner J, et al. The Discovery of a Dual TTK Protein Kinase/CDC2-Like Kinase (CLK2) Inhibitor for the Treatment of Triple Negative Breast Cancer Initiated from a Phenotypic Screen. J Med Chem 2017;60:8989-9002.
44. Maia AR, de Man J, Boon U, et al. Inhibition of the spindle assembly checkpoint kinase TTK enhances the efficacy of docetaxel in a triple-negative breast cancer model. Ann Oncol 2015;26:2180-92.
45. Zhu D, Xu S, Deyanat-Yazdi G, et al. Synthetic Lethal Strategy Identifies a Potent and Selective TTK and CLK1/2 Inhibitor for Treatment of Triple-Negative Breast Cancer with a Compromised G(1)-S Checkpoint. Mol Cancer Ther 2018;17:1727-38.
46. Szymczek A, Carbone M, Pastorino S, et al. Inhibition of the spindle assembly checkpoint kinase Mps-1 as a novel therapeutic strategy in malignant mesothelioma. Oncogene 2017;36:6501-7.
47. Gordeeva O. Cancer-testis antigens: Unique cancer stem cell biomarkers and targets for cancer therapy. Semin Cancer Biol 2018;53:75-89.
48. Suda T, Tsunoda T, Daigo Y, et al. Identification of human leukocyte antigen-A24-restricted epitope peptides derived from gene products upregulated in lung and esophageal cancers as novel targets for immunotherapy. Cancer Sci 2007;98:1803-8.
49. Iinuma H, Fukushima R, Inaba T, et al. Phase I clinical study of multiple epitope peptide vaccine combined with chemoradiation therapy in esophageal cancer patients. J Transl Med 2014;12:84.
50. Kono K, Iinuma H, Akutsu Y, et al. Multicenter, phase II clinical trial of cancer vaccination for advanced esophageal cancer with three peptides derived from novel cancer-testis antigens. J Transl Med 2012;10:141.
51. Dong Z, Zhang H, Zhan T, et al. Integrated analysis of differentially expressed genes in esophageal squamous cell carcinoma using bioinformatics. Neoplasma 2018;65:523-31.

(English Language Editors: R. Scott and J. Reynolds)

Cite this article as: Yang Y, Sun Z, Shi Y, Sun J, Zhang X. Integrated bioinformatics analysis to identify key genes related to the prognosis of esophageal squamous cell carcinoma. Transl Cancer Res 2021;10(4):1679-1691. doi: 10.21037/tcr-20-3220
Figure S1 All differentially expressed genes in module analysis via Cytoscape. Each blue square represents a gene, and the connected lines represent interactions between genes.
Survival curves of 39 hub genes. PICALM-interacting mitotic regulator was excluded due to a lack of gene expression and survival information. Gene name is shown at the top of each figure part. Black line represents low expression of the identified gene, and red line represents high expression of the identified gene. Corresponding life table, hazard ratio, and log-rank P value are shown in each figure part.