Integrated Transcriptomics Explored the Cancer-Promoting Genes CDKN3 in Esophageal Squamous Cell Cancer

Wanpeng Wang

Lianshui County People's Hospital

Kai Liao

Shihezi University

HaoChun Guo

Southeast University

Suqin Zhou

Lianshui County People's Hospital

Ran Yu

Lianshui County People's Hospital

Yanyan Liu

Lianshui County People's Hospital

Yan Pan

Lianshui County People's Hospital

Juan Pu (✉ lianshuipujuan@163.com)

Lianshui County People's Hospital <https://orcid.org/0000-0002-3278-2591>

Research article

Keywords: Bioinformatics, Integrated transcriptomics, Module mining, Esophageal squamous cell carcinoma (ESCC), Cyclin - dependent kinase inhibitor 3 (CDKN3)

Posted Date: December 15th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-127861/v1>

License: (cc) (i) This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background and objectives: The aims of the present study were to explore the critical genes that related to development of ESCC by integrated transcriptomics and investigate the clinical significance by experimental validation.

Methods: The datasets of protein-coding genes expression which involved in ESCC were downloaded from GEO database. The "RobustRankAggreg" package was used for data integration, and the different expression genes (DEGs) were identified based the cut-off criteria as follows: $\text{adjust p-value} < 0.05$, $|\text{Log}_2 \text{fold change (FC)}| \geq 1.5$; The protein expression of seed gene in 184 cases of primary ESCC were detected by immunohistochemistry; The relationship between the expression level of seed genes and clinical significance were analyze.

Results: A total of 244 DEGs were identified by comparing gene expression patterns between ESCC patients and the controls based on integrating dataset of GSE77861, GSE77861, GSE100942, GSE26886, GSE17351, GSE38129, GSE33426, GSE20347 and GSE23400; The CDKN3 were identified the seed gene of top cluster by use of PPI network and plug-in MCODE; The level of CDKN3 mRNA was significantly increased in ESCC patients compared to controls; The positive expression rate of CDKN3 protein in ESCC tissue samples was 32.0% and 61.4% in control, respectively, differences were statistically significant; There was significant correlation between the expression level of CDKN3 and lymph node metastasis or clinical staging of ESCC patients.

Conclusion: Integrated transcriptomics is an efficient approach to system biology. By this procedure, our study improved the understanding of the transcriptome status of ESCC.

Introduction

Esophageal squamous cell carcinoma (ESCC) is a dominant malignant tumor, which accounts for mostly 90% of esophageal carcinoma[1]. Previous studies indicated that a combination of pathological stages and genetic backgrounds cooperatively contributes the progress of ESCC but the concrete molecular mechanism is elusive[2, 3]. Currently, a number of sample data of cancer genomics are accessible on professional network and provides a huge of benefits for further bio-analysis of those cancers[4]. Each individual studies, however, is limited to multi-factors such as sample sizes, batch effects, experimental conditions or so on, and potentially lead to a significant difference of results among them. This problem implied that an effective *in silico* method to integrate those individual study could provide a more profound and valuable conclusion to screen the crucial genes of ESCC[5].

In this study, robust rank aggregation (RRA) method was performed to integrate ESCC data from different public platforms to obtain different expression genes (DEGs) that were used to construct protein-protein interaction (PPI) and screen the hub genes. Expression data of ESCC tissue from TCGA and immunohistochemistry analysis conducted in our hospital were performed to verify crucial genes. This study aims to further explore new bio-markers of ESCC.

Materials And Methods

Data Source

Gene expression profiles were obtained by a systematic retrieval on the GEO (<http://www.ncbi.nlm.nih.gov/geo/>) database with keywords. A total of 9 series (GSEs) with more than 3 cases of ESCC samples and matched normal controls, respectively, were downloaded for further study and their general information of each data sets were shown in Table 1.

Data Preprocessing and Integration of Differentially Expressed Genes

The raw data of GSEs were preprocessed using R package "Affy", including background corrections, normalization, missing data imputation and calculation of gene expression. The R package "limma"[6] was utilized to screen and compare the preprocessed data of ESCC samples with matched controls samples using Bayes test. Corrected P value and absolute values of Fold Chang ($|\text{Log}_2\text{FC}|$) from each data sets were obtained and formed matrix of 9 differential expression matrix. Besides, the R package "RobuStrankaggreg"[5, 7] was utilized to integrate the matrix based RRA method. Genes with $|\text{Log}_2\text{FC}| > 1.5$ and $P < 0.05$ were considered to be DEGs.

Protein-protein Interaction (PPI) Network Construction and ModuleMining

DEGs were further analyzed by STRING (<https://string-db.org/>) to predicts PPI network and a confidence score of 0.4 was set as the threshold value. The PPI network then was visualized using Cytoscape (V3.5.1), with the Molecular Complex Detection (MCODE) plug-in that performed the module analysis. Parameters setting: a degree cut-off >5 , k-core >5 and the rest are default settings.

The verification of mRNA level of hub genes

The mRNA level of hub genes was tested via ESCC data from TCGA. Briefly, expression gene data of ESCC samples and collaterally clinic information were downloaded (<http://xena.ucsc.edu/welcome-to-ucsc-xena/>). The data set was based on IlluminaHiSeq_RNASeqV2 high-throughput RNA sequencing platform, and the expression values were all relative values normalized by computer programming language. The hub genes transcriptase sequencing data of 81 ESCC patients with clinical data and 11 control tissues were extracted for subsequent analysis.

Collected Cases

There were 184 eligible ESCC patients selected from Lianshui County People's Hospital between January 2013 and December 2015 were included in this study. Inclusion criteria: 1) patients with ESCC were pathologically diagnosed by our pathology department. 2) patients weren't undertaken radiotherapy before sampling. 3) there was no history of recent infection or hematologic disease among included patients. Among the 184 ESCC patients, 157 were male and 27 were female with age ranged from 36 to

86 years old. The study protocol was approved by the ethical review committee of Lianshui County People's Hospital. 50 adjacent tissues were collected at the same time as control group.

Immunohistochemistry staining

Paraffin-embedded sections (4 μm) of ESCC and matched normal tissues, saved in our pathology department, were used for CDKN3 immunostaining (Abcam Group, Inc;). After dewaxing, washing and incubating with the primary antibodies (1:200) and secondary antibody in turn, the slides were coloured with DAB and then counterstained with hematoxylin and dehydrated and mounted. Two experienced pathologists were independently evaluated the immunostaining slides by recording the staining intensity of tumor cells and the rate of percentage of positive cells. Concrete criteria were previous article[8].

Statistical analysis

The SPSS 22.0 was used for statistical analysis and the Graphpad Prime 5 was used for drawing statistical pictures. Normal distribution data were indicated as the standard deviation of sample means and their groups were compared using t test. Skewness distribution data were indicated as inter quartile range and their groups were compared using Mann-Whitney test. Enumeration data were represented by frequency or percentage (%) and were tested by χ^2 test. The *P* value of less than 0.05 was considered statistically significant.

Results

DEGs Screening

After performing integrated analysis of 9 series of gene expression profiles from the GEO database, a total of 244 DEGs were found, of which 93 were upregulated and 151 were downregulated ($P < 0.05$ and $|\log_2(\text{FC})| > 1.5$). The top 10 upregulated and downregulated DEGs are shown in Figure 1.

PPI Network Construction and ModuleMining

To explore the biological functions of these 244 DEGs, a PPI network that included 194 nodes and 864 edges via STRING (Figure 2A). Then, modules with core significance were obtained via modules mining and analysis using MCODE app from cytoscape software. Results show that the module with the highest score (23.304) contain 24 nodes and 268 edges (Figure 2B), of which the seed gene is cyclin dependent kinase inhibitor 3 (CDKN3)

The verification of mRNA level of CDKN3 in ESCC

Results of TCGA analysis showed that the relative expression level of CDKN3 is 3.291 (IQR: 2.833 ~ 3.659) and that of 11 control groups is 1.184 (IQR: 0.734 ~ 1.72) (Figure. 3A) with statistically significance ($U = 18.00$, $P < 0.001$). Analysis of receiver operating characteristic curve (ROC) showed that

area under the curve (AUC) is 0.980 (Figure. 3B) with a 2.149 of cut off value. The sensitivity and specificity were 90.91% (95%CI: 58.72% ~ 99.77%) and 92.59% (95%CI: 84.57% ~ 97.23%), respectively.

Immunohistochemical analysis for CDKN3 protein

Immunohistochemical analysis was used to detect CDKN3 expression in 184 ESCC tissue and 50 matched normal tissues. We found that the rate of positive expression of CDKN3 protein in ESCC tissues (61.4%, 113/184) were higher than that in matched normal tissues (16.49%, 34/ 50) with statistically significance ($\chi^2 = 13.75$, $p < 0.001$) (Figure. 4A-4D).

Correlation between between CDKN3 and ESCC patients

Correlation between the protein expression of CDKN3 and clinicopathological features of ESCC patients are shown in Table 2. Briefly, there is no statistic correlation on age ($\chi^2 = 0.788$, $p = 0.375$), gender ($\chi^2 = 0.788$, $p = 0.375$), tumor location ($\chi^2 = 0.017$, $p = 0.898$), differentiation grades ($\chi^2 = 0.328$, $p = 0.567$), T stage ($\chi^2 = 0.025$, $p = 0.874$), M stage ($\chi^2 = 1.479$, $p = 0.224$) but a significantly statistic correlation on N stage ($\chi^2 = 10.352$, $p = 0.001$) and clinical stage ($\chi^2 = 6.158$, $p = 0.013$).

Discussion

Current dominant *in silico* methods of integrated transcriptomics include: 1) to analysis each expression profile and make an intersection between each DEGs. 2) to remove batch effects via 'combat' function of sva package. The former methods is supposed to be limited in batch effects according to our previous experience in other study[9]. However, the latter method cannot be conducted in cross-platform analysis due to its deep reliance on similar experiment backgrounds[10]. In this study, we performed RRA to integrate transcriptomics because this method is not only avoid the interference of cross-platform, but also enlarge the simple size. Our results indicated that there were 244 DEGs were screened via this method. Besides, many genes among DEGs such as MMP1[11], MAGEA6[12] and MAL[13] were closely associated with the progress of ESCC, which also implied the reliability of RRA.

The pathological mechanism of ESCC is complicated and involved a number of pathways and genes, which cause a deep restriction on traditional biological study. In this study, the PPI were constructed by DEGs to explore the crucial module of gene-gene interaction. The modules with the highest importance consist of 24 gene, of which genes such as FOXM1[14] or DTL[15] were considered as crucial genes in ESCC. The Cyclin-dependent protein kinase (CDK), a central gene in module, encodes a cell cycle regulatory protein which is associated with multi-tumors[16]. Our results indicated that compared with control group, the mRNA level of CDKN3 is significantly higher. Besides, our immunohistochemical study indicated that there is an abnormal expression of CDKN3 protein in ESCC patients, which preliminarily confirmed its association with the progress of ESCC. Meanwhile, recent studies suggested that CDKN3 is partly responsible for the proliferation, invasion and metastasis of esophageal squamous cell carcinoma via AKT sign pathway dependently or independentl, which also is a corroborative evidence for the reliability of our study[17, 18].

Conclusions

In conclusion, our method is to explore the pathogenesis of ESCC and its candidate bio-markers of diagnose and prognosis at the molecule level. This study is also of instructive value for other cancer studies.

Declarations

Acknowledgements

Not applicable

Authors' contributions

PJ and PY are responsible for the study design. WW, LK and GC performed the experiments and draft the manuscript. ZS, YR and LY participated in the data analysis and interpretation. All authors have read and approved the final manuscript.

Funding

This study was sponsored by the Natural Science Research Project of Huai'an City (HAB201949) and the Medical Scientific Research Project of Health and Family Planning Commission of Jiangsu Province (Z2020022, Z2018026, Z2019045).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The present study was approved by the Medical Ethics Committee of Lianshui County People's Hospital and all patients provided written informed consent

Consent for publication

Not applicable.

Competing interests

The authors have no conflict of interest to declare.

References

1. Siegel R, Ma J, Zou Z, et al. Cancer statistics, 2014. CA Cancer J Clin. 2014;64:9–29.

2. Tokuda E, Itoh T, Hasegawa J, et al. Phosphatidylinositol 4-phosphate in the Golgi apparatus regulates cell-cell adhesion and invasive cell migration in human breast cancer. *Cancer Res.* 2014;74:3054–66.
3. Lyu S, Lu J, Chen W, et al. High expression of eIF4A2 is associated with a poor prognosis in esophageal squamous cell carcinoma. *Oncol Lett.* 2020;20:177.
4. Xu J, Shu Y, Xu T, et al. Microarray expression profiling and bioinformatics analysis of circular RNA expression in lung squamous cell carcinoma. *Am J Transl Res.* 2018;10:771–83.
5. Kolde R, Laur S, Adler P, et al. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics.* 2012;28:573–80.
6. Kerr MK. Linear models for microarray data analysis: hidden similarities and differences. *J Comput Biol.* 2003;10:891–901.
7. Zhang H, Du Y, Wang Z, et al. Integrated Analysis of Oncogenic Networks in Colorectal Cancer Identifies GUCA2A as a Molecular Marker. *Biochem Res Int.* 2019;2019:6469420.
8. Ge H, Lu Y, Chen Y, et al. ERCC1 expression and tumor regression predict survival in esophageal squamous cell carcinoma patients receiving combined trimodality therapy. *Pathol Res Pract.* 2014;210:656–61.
9. Zuo Z, Shen JX, Pan Y, et al. Weighted Gene Correlation Network Analysis (WGCNA) Detected Loss of MAGI2 Promotes Chronic Kidney Disease (CKD) by Podocyte Damage. *Cell Physiol Biochem.* 2018;51:244–61.
10. Wang W, Fu S, Lin X, et al. miR-92b-3p Functions As A Key Gene In Esophageal Squamous Cell Cancer As Determined By Co-Expression Analysis. *Onco Targets Ther.* 2019;12:8339–53.
11. Liu M, Hu Y, Zhang MF, et al. MMP1 promotes tumor growth and metastasis in esophageal squamous cell carcinoma. *Cancer Lett.* 2016;377:97–104.
12. Hao J, Li S, Li J, et al. Investigation into the expression levels of MAGEA6 in esophageal squamous cell carcinoma and esophageal adenocarcinoma tissues. *Exp Ther Med.* 2019;18:1816–22.
13. Jin Z, Wang L, Zhang Y, et al. MAL hypermethylation is a tissue-specific event that correlates with MAL mRNA expression in esophageal carcinoma. *Sci Rep.* 2013;3:2838.
14. Xiao Z, Jia Y, Jiang W, et al. FOXM1: A potential indicator to predict lymphatic metastatic recurrence in stage IIA esophageal squamous cell carcinoma. *Thorac Cancer.* 2018;9:997–1004.
15. Dietzsch E, Parker MI. Infrequent somatic deletion of the 5' region of the COL1A2 gene in oesophageal squamous cell cancer patients. *Clin Chem Lab Med.* 2002;40:941–5.
16. Chang SL, Chen TJ, Lee YE, et al. CDKN3 expression is an independent prognostic factor and associated with advanced tumor stage in nasopharyngeal carcinoma. *Int J Med Sci.* 2018;15:992–8.
17. Yu H, Yao J, Du M, et al. CDKN3 promotes cell proliferation, invasion and migration by activating the AKT signaling pathway in esophageal squamous cell carcinoma. *Oncol Lett.* 2020;19:542–8.
18. Liu J, Min L, Zhu S, et al. Cyclin-Dependent Kinase Inhibitor 3 Promoted Cell Proliferation by Driving Cell Cycle from G1 to S Phase in Esophageal Squamous Cell Carcinoma. *J Cancer.* 2019;10:1915–22.

Tables

Table.1 The datasets of ESCC protein-coding genes expression included the present study

GSE	77861	100942	26886	17351	38129	33426	29001	20347	23400
Platform	GPL570				GPL571				GPL96/97
Number of probe (n)	54675				22277				44928
ESCC (n)	7	5	9	5	30	59	21	17	53
Con (n)	7	5	19	5	30	12	24	17	53

Table.2 The relationship between the expression level of CDKN3 and clinical significance

Items	N	CDKN3 protein		χ^2	P
		Positive	Negative		
Gender					
Male	157	94	63	1.071	0.301
Female	27	19	8		
Age					
≤60 years	91	52	39	0.788	0.375
>60 years	93	56	32		
Location					
neck/upper thoracic	56	34	22	0.017	0.898
mid/lower thoracic	128	79	49		
Grade					
G1/G2	125	75	50	0.328	0.567
G3	59	38	21		
T					
T1/T2	79	48	31	0.025	0.874
T3/T4	105	65	40		
N					
N0	79	38	41	10.352	0.001
N1	105	75	30		
M					
M0	175	107	69	1.479	0.224
M1	9	8	2		
Stage					
IIA	106	57	49	6.158	0.013
IIB	78	56	22		

Figures

Figure 1

The different expression genes (DEGs) identified by Integrated transcriptomics

Figure 2

Construction the protein-protein interaction (PPI) network. A: Construction the PPI network of the different expression genes (DEGs); B: The module with with the highest score identified by using MCODE.

Figure 3

The CDKN3 mRNA expression of primary ESCC and Control tissue. A: Expression levels of CDKN3 mRNA in ESCC patients and controls. B: Receiver operating characteristic (ROC) curves for CDKN3 mRNA in discriminating ESCC patients with controls.

Figure 4

Protein expression of CDKN3. A: The negative expression of CDKN3 protein in normal tissue samples. B: The positive expression of CDKN3 protein in normal tissue samples. C: The negative expression of CDKN3 protein in ESCC samples. D: The positive expression of CDKN3 protein in ESCC samples.