Five extracellular matrix-associated genes upregulated in oral tongue squamous cell carcinoma: An integrated bioinformatics analysis

PINGPING ZHONG1,4*, LIYA LIU5,6*, ALING SHEN5,6, ZHONGXIN CHEN7, XIAOYAN HU2,4, YICHAO CAI2,4, JIE LIN2,4, BANGYANG WANG2,4, JIESEN LI2,4, YOUQIN CHEN8 and JUN PENG5,6

1Department of Orthodontics, School and Hospital of Stomatology, Fujian Medical University; 2Key Laboratory of Stomatology, Fujian Province University; 3Fujian Biological Materials Engineering and Technology Center of Stomatology; 4Institute of Stomatology, Fujian Medical University, Fuzhou, Fujian 350002; 5Academy of Integrative Medicine; 6Fujian Key Laboratory of Integrative Medicine on Geriatrics, Fujian University of Traditional Chinese Medicine, Fuzhou, Fujian 350122; 7West China School of Stomatology Sichuan University, Chengdu, Sichuan 610041, P.R. China; 8Department of Pediatrics, Case Western Reserve University School of Medicine, Rainbow Babies and Children’s Hospital, Cleveland, OH 44106, USA

Received November 3, 2018; Accepted July 26, 2019

DOI: 10.3892/ol.2019.10982

Abstract. Despite advancements in treatment regimens, the mortality rate of patients with oral tongue squamous cell carcinoma (OTSCC) is high. In addition, the signaling pathways and oncoproteins involved in OTSCC progression remain largely unknown. Therefore, the aim of the present study was to identify specific prognostic marker for patients at a high risk of developing OTSCC. The present study used four original microarray datasets to identify the key candidate genes involved in OTSCC pathogenesis. Expression profiles of 93 OTSCC tissues and 76 normal tissues from GSE9844, GSE13601, GSE31056 and GSE75538 datasets were investigated. Differentially expressed genes (DEGs) were determined, and gene ontology enrichment and gene interactions were analyzed. The four GSE datasets reported five upregulated and six downregulated DEGs. Five upregulated genes (matrix metalloproteinase 1, 3, 10 and 12 and laminin subunit gamma 2) were localized in the extracellular region of cells and were associated with extracellular matrix disassembly. Furthermore, analysis for The Cancer Genome Atlas database revealed that the aforementioned five upregulated genes were also highly expressed in OTSCC and head and neck squamous cell carcinoma tissues. These results demonstrated that the five upregulated genes may be considered as potential prognostic biomarkers of OTSCC and may serve at understanding OTSCC progression. Upregulated DEGs may therefore represent valuable therapeutic targets to prevent or control OTSCC pathogenesis.

Introduction

Head and neck squamous cell carcinoma (HNSCC) is a common cancer worldwide and accounts for >600,000 new cases annually (1). Despite significant advancements in treatments, including reconstructive microvascular free tissue transfer, hyperfractionated radiotherapy and concurrent chemoradiation, the survival rate of patients with HNSCC has not sufficiently improved over the last 50 years, with overall survival of ~50%, resulting in increased mortality rates worldwide every year (2). HNSCC is a highly complex disease emerging from the oral cavity, tongue, pharynx or larynx (3). Each tumor harbors unique mutations, presents variable clinical outcomes and is associated with specific risk factors (4,5). For example, TP53 inactivation, either through somatic mutation or HPV infection, appears nearly universal in this malignancy (6). The present study particularly focused on one type of HNSCC, the oral tongue squamous cell carcinoma (OTSCC), because of its poor diagnosis, high incidence
rate, aggressive clinical behavior and poor outcome (7-10). A recent study reported a five-year survival rate of 63% for patients with OTSCC in The Netherlands (11). In 2017, nearly 16,400 new cases of tongue cancer were diagnosed and 2,400 tongue cancer-associated mortality cases were recorded in the United States (12). Several prognostic factors for OTSCC exist, including occult node positivity, tumor depth, lymphovascular invasion and perineural invasion (13). Nevertheless, robust and reliable molecular prognostic biomarkers need to be determined in order to identify patients with advanced stages of OTSCC.

Several thousands of tumor biomarkers have been discovered and are associated with the prognosis of various types of cancer (14). In particular, markers for OTSCC, including microtubule associated scaffold protein 1 and beta-parvin, have attracted much attention due to their crucial role in OTSCC pathogenesis (15). Furthermore, it was demonstrated that downregulation of these markers markedly decreases cancer cell survival (15-19).

Thanks to the rapid development and extensive application of microarrays for gene identification in various types of cancer (20-22), a high number of differentially expressed genes (DEGs) in OTSCC have been identified (13,19). However, the results between studies are inconsistent, which might be due to the variability of tissue samples used. In addition, no reliable biomarkers for OTSCC have been established. Subsequently, the combination of bioinformatics methods and expression profiling techniques may represent a novel approach to resolve these problems. The present study used four microarray datasets [GSE9844 (3), GSE13601 (23), GSE31056 (24), and GSE75538 (25)] from the National Center for Biotechnology Information (NCBI)-Gene Expression Omnibus (GEO) database and an mRNA sequencing (mRNA-seq) dataset from The Cancer Genome Atlas (TCGA). DEGs were filtered using the GEO2R tool according to conventional data processing standards, and gene ontology (GO) enrichment analysis was performed to screen for DEGs using The Database for Annotation, Visualization and Integrated Discovery (DAVID). Expression levels of DEGs in samples from TCGA were assessed using Gene Expression Profiling Interactive Analysis (GEPIA), and corrections were applied for DEGs in samples from The Cancer Genome Atlas database. GO enrichment analysis of DEGs. Usually, genes and their products are annotated to identify characteristic biological function and processes of a high-throughput genome or transcriptome (28). DAVID (https://david.ncifcrf.gov/) is a web-based bioinformatics resource for gene annotation and visualization with an integrated discovery function. It is therefore useful for determining gene biological attributes (29). P<0.01 and the Benjamini corrected P<0.01 (30) were set as cutoffs. Co-expressed DEGs that were downregulated or upregulated in the two sets of gene expression profiles were identified using Venn diagrams (http://bioinformatics.psb.ugent.be/webtools/Venn/).

Comparison of gene expression in patients with OTSCC. The mRNA-seq data of the five OTSCC genes of interest from HNSC samples were obtained from the TCGA database.
In addition, clinical data of OTSCC were downloaded using TCGA Assembler (Table I). mRNA-seq data of 147 OTSCC tissues and 15 adjacent normal tongue tissues were obtained on an Illumina HiSeq RNASeq platform (32). Since the TCGA dataset was developed as a community resource project, no additional approval was required for this study from Fujian Medical University. The present study complied with the TCGA publication guidelines and data access policies (33).

Analysis of gene expression patterns in all tumors from TCGA through GEPIA. The expression patterns of the five genes of interest in various types of cancers and normal tissues were determined using GEPIA (http://gepia.cancer-pku.cn) (34). The program was used to analyze RNA sequencing data of 33 types of cancer (including adrenocortical carcinoma, bladder urothelial carcinoma, breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), lymphoid neoplasm diffuse large B-cell lymphoma, esophageal adenocarcinoma (ESCA), glioblastoma multiforme, head and neck squamous cell carcinoma (HNSC), kidney chromophobe (KICH), kidney renal clear cell carcinoma, kidney renal papillary cell carcinoma, acute myeloid leukemia, brain lower grade glioma, liver hepatocellular carcinoma, lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), mesothelioma, ovarian serous cystadenocarcinoma, pancreatic adenocarcinoma, pheochromocytoma and paraganglioma, prostate adenocarcinoma, rectum adenocarcinoma (READ), sarcoma, skin cutaneous melanoma, stomach adenocarcinoma (STAD), testicular germ cell tumor, thyroid carcinoma, thymoma, uterine corpus endometrial carcinoma, uterine carcinosarcoma, uveal melanoma) and adjacent normal samples from TCGA following the standard processing pipeline. P-value <0.01 and |logFC|>2 were set as cutoffs. Box plots were generated to visualize the associations.

Correlations among genes expression in HNSC samples from TCGA through cBioPortal. The correlations between expression patterns of the genes of interest in cancer tissues from TCGA-HNSC dataset were analyzed using Pearson and Spearman correlation coefficient via cBioPortal (http://www.cbioportal.org) (35-36).

Statistical analysis. Statistical analyses were performed using Graphpad Prism 5.0 (GraphPad Software, Inc.). Mann-Whitney U test was performed to compare gene expression between tumor and adjacent normal tissues. A two-tailed P-value <0.01 was considered statistically significant.

Results

Some biomarkers distinguished OTSCC tissues from normal tissues. Four GEO datasets (GSE9844, GSE13601, GSE31056, and GSE75538) were imported into the GEO2R analysis tool. Based on DEGs between OTSCC and normal tissues, five upregulated genes [matrix metalloproteinases 1, 3, 10 and 12 (MMP1, MMP3, MMP12, MMP10) and laminin subunit gamma 2 (LAMC2)] (Fig. 1A) and six downregulated genes (dermatopontin, cartilage intermediate layer protein, keratin 4, ATP binding cassette subfamily A member 8, alcohol dehydrogenase 1B and protein phosphatase 1 regulatory

Table II. Gene ontology analysis of upregulated genes associated with oral tongue squamous cell carcinoma.

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>Count</th>
<th>Genes</th>
<th>P-value</th>
<th>Benjamini corrected P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0022617-extracellular matrix disassembly</td>
<td>5</td>
<td>MMP10, LAMC2, MMP3, MMP12, MMP1</td>
<td>3.9x10^{-10}</td>
<td>7.7x10^{-9}</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0030574-collagen catabolic process</td>
<td>4</td>
<td>MMP10, MMP3, MMP12, MMP1</td>
<td>2.1x10^{-7}</td>
<td>2.1x10^{-6}</td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0004222-metalloendopeptidase activity</td>
<td>4</td>
<td>MMP10, MMP3, MMP12, MMP1</td>
<td>1.2x10^{-6}</td>
<td>9.3x10^{-6}</td>
</tr>
<tr>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0005578-proteinaceous extracellular matrix</td>
<td>4</td>
<td>MMP10, MMP3, MMP12, MMP1</td>
<td>1.2x10^{-5}</td>
<td>1.1x10^{-4}</td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0004252-serine-type endopeptidase activity</td>
<td>4</td>
<td>MMP10, MMP3, MMP12, MMP1</td>
<td>1.3x10^{-5}</td>
<td>5.4x10^{-5}</td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0004175-endopeptidase activity</td>
<td>3</td>
<td>MMP3, MMP12, MMP1</td>
<td>6.0x10^{-5}</td>
<td>1.6x10^{-4}</td>
</tr>
<tr>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0005576-extracellular region</td>
<td>5</td>
<td>MMP10, LAMC2, MMP3, MMP12, MMP1</td>
<td>6.1x10^{-5}</td>
<td>2.7x10^{-4}</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0006508-proteolysis</td>
<td>4</td>
<td>MMP10, MMP3, MMP12, MMP1</td>
<td>1.0x10^{-4}</td>
<td>6.8x10^{-4}</td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0005509-calcium ion binding</td>
<td>4</td>
<td>MMP10, MMP3, MMP12, MMP1</td>
<td>3.0x10^{-4}</td>
<td>5.9x10^{-4}</td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0008270-zinc ion binding</td>
<td>4</td>
<td>MMP10, MMP3, MMP12, MMP1</td>
<td>1.3x10^{-3}</td>
<td>2.0x10^{-3}</td>
</tr>
</tbody>
</table>
Figure 1. Venn diagram of (A) upregulated and (B) downregulated DEGs in four GEO datasets (GSE9844, GSE13601, GSE31056 and GSE75538). Each colored area represents the number of DEGs in one corresponding dataset. The cross-areas indicated altered DEGs. DEGs were identified using Mann-Whitney U test and statistically significant DEGs were defined for P<0.01 and (A) log FC>2 or (B) log FC<−2 as cutoffs. DEGs, differentially expressed genes.

Figure 2. Expression levels of (A) MMP1, (B) LAMC2, (C) MMP3, (D) MMP10 and (E) MMP12 in oral tongue squamous cell carcinoma and normal tissues based on TCGA database. *P<0.01. LAMC2, laminin subunit gamma 2; MMP1, 3, 10 and 12, matrix metalloproteinase 1, 3, 10 and 12; N, normal tissue; T, cancer tissue.
subunit 3C) (Fig. 1B) were highlighted in all four datasets. Whether the five upregulated genes could be considered as potential biomarkers for distinguishing OTSCC tissues from normal tissues was therefore assessed.

**GO function enrichment analysis.** GO enrichment analysis was performed using DAVID tool. The five specific upregulated genes \( \text{MMP1, LAMC2, MMP3, MMP10 and MMP12} \) were uploaded into the DAVID software. GO results indicated that these genes were specifically involved in certain biological processes, including ‘extracellular matrix disassembly’, ‘collagen catabolism’ and ‘proteolysis’. Furthermore, with regards to the molecular function, the products of these five genes mainly comprised metalloendopeptidases, serine-type endopeptidases, calcium ion binders and zinc ion binders. Similarly, GO cell component analysis revealed that the proteins encoded by these genes were significantly enriched for functions associated with the proteinaceous extracellular matrix and extracellular region (\( P<0.01 \), Table II).

**Gene expression of \( \text{MMP1, LAMC2, MMP3, MMP10 and MMP12 in OTSCC samples from TCGA.} \)** TCGA database was used to analyze the expression levels of the five genes of interest in OTSCC and adjacent normal tissues. The results demonstrated that the expression levels of \( \text{MMP1, LAMC2, MMP3, MMP10 and MMP12} \) were significantly upregulated in OTSCC tissues compared with adjacent normal tissues (Fig. 2).
ESCA, HNSC, and READ (Fig. 3C). MMP10 expression was significantly upregulated in ESCA, HNSC and LUSC (Fig. 3D). In addition, MMP12 was upregulated in CESC, ESCA, HNSC, LUAD, LUSC and STAD, whereas it was downregulated in pancreatic adenocarcinoma (Fig. 3E). Overall, the expression level of MMP1, LAMC2, MMP3, MMP10 and MMP12 was significantly upregulated in HNSC tissues.

Correlations among genes expression in samples from the TCGA-HNSC database. To select the five genes of interest for further investigation, the correlations between gene expression patterns in samples from TCGA were analyzed using cBioPortal (35-36). The results demonstrated that MMP1 expression level was positively correlated with LAMC2 (Fig. 4A), MMP3 (Fig. 4B) and MMP10 (Fig. 4C) expression levels. In addition, LAMC2 expression level was positively correlated with MMP10 expression level (Fig. 4D).

Pearson correlation coefficients ranged from 0.33 to 0.69, and Spearman correlation coefficients ranged from 0.55 to 0.83. All together, these results suggested that MMP1, LAMC2, MMP3 and MMP10 expression levels were positively correlated in HNSC tissues.

Discussion

The American Joint Committee on Cancer has updated the Tumor-Node-Metastasis system for cancer staging system according to diagnostic class, treatment choices and prognosis for OTSCC (37). However, with the development of novel modern therapeutic approaches for cancer, genetic analysis is an outstanding tool for early diagnosis of cancer that can prolong patient survival (38). Previous clinical studies on OTSCC development and progression reported increased incidence and mortality rates due to the lack of knowledge on
OTSCC (39,40). The present study identified 11 DEGs (five upregulated and six downregulated genes) from four mRNA expression profile datasets via GEO. These DEGs were significantly associated with OTSCC pathogenesis. The present study therefore focused on the five upregulated DEGs since they may be considered as potential reliable biomarkers. These five target genes were categorized into three classes according to their associated biological functions (molecular functions, biological processes and cellular components) thanks to the GO enrichment using multiple approaches. The results demonstrated that all genes were involved in extracellular matrix disassembly and were localized in the extracellular region during cancer development. Furthermore, expression levels of these genes were positively correlated with each other and upregulated in OTSCC and HNSC tissues compared with adjacent normal tissues, as confirmed using DAVID.

Expression levels of four MMP family members, including MMP1, MMP3, MMP10 and MMP12, were upregulated in HNSC tissues compared with adjacent normal tissues. These genes serve important functions in numerous physiological and pathological processes involved in tumor progression and can promote tumor-induced angiogenesis and extracellular matrix disassembly, enhancing therefore tumor invasion and metastasis (41,42). MMP1 is known to be significantly associated with poor clinical outcomes and is therefore a robust prognostic factor for various types of cancer, including colorectal cancer and breast cancer (43-45). Furthermore, MMP3 has been proposed as a vital tumor oncogene in numerous cancers, including pancreatic, pulmonary, and mammary carcinoma (46). In addition, MMP10 has been demonstrated to be a potential clinical marker for cancer stem-like cells/cancer-initiating cells in epithelial ovarian cancer (EOC) and may serve as a therapeutic target in chemotherapy-resistant EOCs (47). MMP12 has been reported to be highly expressed in nasopharyngeal carcinoma both in vitro and in vivo, and is therefore a powerful tumor marker for cancer metastasis (48). Furthermore, LAMC2, which encodes the major component of laminin-5, was one of the upregulated genes in the present study. Highly expression levels of LAMC2 have been previously reported in numerous invasive tumors (49,50), including carcinomas of the lung (51), the colorectum (52,53) and the pancreas (54,55). In addition, it has been reported that LAMC2 overexpression in esophageal cancer is associated with poor survival (56,57). Subsequently, MMP1, MMP3, MMP10, MMP12 and LAMC2 may serve as novel promising prognostic factors in advanced metastasis of OTSCC. The protein expression of MMP1, MMP3, MMP10, MMP12 and LAMC2 should be further confirmed in OTSCC tissues. Furthermore, the correlation between MMP1, MMP3, MMP10, MMP12 and LAMC2 protein expression and the survival and clinic characteristics of patients with OTSCC should be further assessed. Tissue samples from patients with OTSCC are currently being collected for this purpose, and will be used for immunohistochemistry analysis to detect MMPs protein expression. Furthermore, the clinical significance, biological function and underlying mechanism of MMPs on the pathogenesis and progression of OTSCC will be further investigated.

In conclusion, the present study identified five upregulated genes that may be considered as robust and reliable biomarkers in OTSCC prognosis. The findings from this study may help discovering the role of oncogenes in cancer progression and determining the underlying pathways in order to develop novel therapeutic strategies for OTSCC.

Acknowledgements

Not applicable.

Funding

The present study was funded by the National Natural Science Foundation of China (grant no. 81673721) and the Natural Science Foundation of Fujian Province (grant no. 2017J0007).

Availability of data and materials

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

Authors' contributions

PZ, LL, YC and JP conceived and designed the experiments. PZ, AS, ZC and XH performed GEO and GEO2R analyses. LL, XH, YC and JL conducted gene ontology analyses using DAVID. BW, JL, YC and JP analyzed TCGA database using GEPIA. LL, AS, ZC and YC analyzed expressions of genes in OTSCC from the TCGA database. XH, JL, BW and JL assessed the correlations among gene expressions using cBioPortal. PZ, LL, YC and JP drafted the manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interest

The authors declare that they have no competing interests.

References


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.