Original Article
Transcriptome profiles of moderate dysplasia in oral mucosa associated with malignant conversion

Jiali Zhang¹², Si Chen¹³, Lan Yin⁴, Xinming Chen¹²

¹The State Key Laboratory Breeding Base of Basic Science of Stomatology (Hubei_MOST) & Key Laboratory of Oral Biomedicine Ministry of Education, School & Hospital of Stomatology, Wuhan University, Wuhan 430079, China; ²Department of Oral Histopathology, School and Hospital of Stomatology, Wuhan University, Wuhan 430079, China; ³Department of Oral Implantology, School and Hospital of Stomatology, Wuhan University, Wuhan 430079, China; ⁴ABLife, Inc., Wuhan 430075, China

Received January 24, 2016; Accepted April 23, 2016; Epub June 1, 2016; Published June 15, 2016

Abstract: Background: Oral malignancies are among the top most deadly cancers in the world. Early diagnosis of oral premalignant and malignant lesions is essential for treatment decision-making and prognosis improvement. Methods: RNA sequencing was applied to diseased tissues and paired normal tissue from three patients with moderate epithelial dysplasia (MED). Hierarchical clustering analysis was performed to evaluate the correlation of gene expression patterns of different specimens. EdgeR software was used to identify differentially expressed genes between MED and paired normal tissues, gene ontology analysis was carried out on DAVID online platform. qRT-PCR was conducted to confirm the accuracy of our RNA-seq data. Results: A total of 21556 genes were detected in microdissected MED and paired normal tissues. Hierarchical clustering analysis revealed distinct gene expression patterns between the moderate dysplasia samples. 346 differentially expressed genes that may contribute to the pathogenesis of moderate dysplasia were identified. Among these genes, ISG15 showed different transcription profiles that close related with the alterations of gene expression patterns in moderate dysplasia samples and paired normal tissues. Conclusions: We speculated that, as the potential marker to signal the risk of malignant transformation, ISG15 together with its close related genes may involve in the malignant conversion from oral dysplasia to oral squamous cell carcinoma.

Keywords: Oral dysplasia, RNA sequencing, transcriptome, interferon-stimulated gene 15

Introduction

Oral cancer is the most common subtype of head and neck squamous cell carcinoma and a significant global cancer killer [1]. Oral cancer develops through a series of histopathological stages: through mild, moderate, and severe dysplasia to carcinoma in situ and then to squamous cell carcinoma, defined by the World Health Organization (WHO) classical oral system [2]. It is regarded that severe epithelial dysplasia and carcinoma in situ show high risk of malignant transformation, with a wide range of 7-50% [3], moderate epithelial dysplasias (MED) have a malignant transformation potential of 3-15% [4], whereas only a small proportion of mild epithelial dysplasia progress to oral squamous cell carcinoma (OSCC) (< 5%) [3]. Surgical excision, conservative or active surveillance and chemoprevention are the current treatment options for oral dysplasia. However, many patients have been over treated because it is difficult to predict accurately which dysplasia lesion will progress to OSCC. Due to wide intra- and inter- observer variability [5], the grading system of oral epithelial dysplasia (OED) has been shown to be subjective and not ideal. Consequently, novel approaches and validated molecular markers for defining progression likelihood for histopathologically similar OEDs are required.

Numerous microarray studies on oral malignancies have been performed to reveal the tumor-specific disruptive genes [6, 7]. Nevertheless, such hybridization-based approach typically
Transcriptome profiling of oral dysplasia

resides on existing knowledge about genome sequence, and has a limited dynamic range of detection owing to both high background levels and saturation of signals [8]. The recent advent of massively parallel sequencing has started to revolutionize biomedical studies, RNA sequencing (RNA-seq) becomes a superior approach that allows the entire transcriptome to be surveyed in a very high-throughput and quantitative manner [8]. By sequencing the whole transcriptome of the diseased tissue and a paired normal specimen, we can accurately measure the transcript expression levels of a panel of genes which can be used effectively to make informed clinical decisions.

RNA-seq of OSCC and matched normal tissues showed the benefit of this powerful tool in measurement of the low abundance transcripts [9]. However, the comprehensive transcriptome profiles of MED patients and its association with oral pathogenesis and carcinogenesis were seldom reported. Herein, we conducted a genome-wide study to investigate the gene expression pattern in MED by interrogating RNA-seq data from three MED samples and their matched normal samples. The significantly mis-regulated genes were identified and may mediate different pathways that contribute to the malignant conversion from oral epithelial dysplasia to OSCC.

Materials and methods

Patients and tissue samples

Three patients with moderate epithelial dysplasia were recruited from Wuhan University Stomatological Hospital, Hubei, China from December, 2012 to August, 2013, they were named as #1, #2, and #3. Another three moderate dysplasia and two severe dysplasia specimens used for qRT-PCR validation of ISG15 expression were also obtained from Wuhan University Stomatological Hospital. All the patients were diagnosed and graded according to WHO criteria [10]. The diseased tissues and paired normal tissues were collected and embedded in optimal cutting temperature compound (OCT) (VWR Scientific, San Diego, CA). The frozen biopsy specimens were cryosectioned into thin slices of 4 μm, the first and the last 4 μm H&E-stained sections were used for orientation, and cells of the MEDs were manually microdissected from 2 to 4 adjacent sections. Microdissected tissue was directly placed into TRIzol (Invitrogen, Carlsbad, CA, USA) in RNase free eppendorf tube. All study protocols were reviewed and approved by the Ethics Committee of Wuhan University Stomatological Hospital. Informed consent was obtained from all patients in accordance with our institutional guidelines.

RNA isolation

Total RNA was extracted from each tissue sample using TRIzol kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer’s instructions. Subsequently, the RNA samples were treated with RQ1 DNase (Promega, Madison, WI, USA) to remove both double- and single-stranded DNA. The quality and quantity of the purified RNA were determined by measuring the absorbance at 260 nm/280 nm (A260/A280) on SmartSpec Plus Spectrophotometer (BioRad, Philadelphia, PA, USA). RNA integrity was further verified by 1.5% agarose gel electrophoresis and assessed by Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

cDNA library preparation

For each sample, 10 mg of total RNA was used for RNA-seq library preparation. Polyadenylated mRNAs were purified and concentrated with Magnetic Beads Oligo (dT) (Invitrogen, Carlsbad, CA, USA) before used for directional RNA-seq library preparation. Purified mRNAs were iron fragmented at 95°C followed by end repair and 5’ adaptor ligation. Then, reverse transcription was performed with RT primer harboring 3’ adaptor sequence and randomized hexamer. The cDNAs were purified and amplified and PCR products corresponding to 200-500 bps were purified, quantified and stored at -80°C until used for sequencing.

Whole transcriptome sequencing

For high-throughput sequencing, the prepared libraries were applied to Illumina HiSeq 2000 system 100 nt pair-end sequencing by ABlife, Inc. (Wuhan, China).

RNA-Seq raw data clean and alignment statistics

First, raw reads containing more than 2-N bases were discarded, then the rest reads were
processed by clipping adaptor and low quality bases were removed, too short reads (less than 16nt) were also dropped. FASTX-Tool kit (Version 0.0.13) was used to get the clean reads. Afterwards, clean reads were aligned to hg19 build of the human genome from University of California Santa Cruz (UCSC) genome browser database (http://genome.ucsc.edu) by Tophat2 [11]. Based on the genome location of the reads, aligned reads with more than one genome location were discarded due to their ambiguous location. Uniquely localized reads were used to calculate reads number and RPKM value (RPKM repre-

Figure 1. Histopathological examination of biopsy specimens from the three patients with MED. A-F. Photomicrographs of hematoxylin and eosin H&E. stained slides demonstrate a proliferation of atypical squamous cells extending into the middle third of the epithelium, basal cell hyperplasia, regional absence of basal polarity, loss of intercellular adhesion. A and B correspond to patient #1 (H&E), C and D correspond to patient #2 (H&E), E and F correspond to patient #3 (H&E). Bar in A, C, E = 200 μm, and bar in B, D, F = 100 μm.
Transcriptome profiling of oral dysplasia

Table 1. Alignment statistics for transcriptome reads from the six clinical samples

<table>
<thead>
<tr>
<th>Sample</th>
<th>Patient #1</th>
<th>Patient #2</th>
<th>Patient #3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
<td>MED</td>
<td>Normal</td>
</tr>
<tr>
<td>Total reads</td>
<td>22 M</td>
<td>26 M</td>
<td>28 M</td>
</tr>
<tr>
<td>Clean reads</td>
<td>20 M, 91.37%</td>
<td>23 M, 86.70%</td>
<td>25 M, 90.96%</td>
</tr>
<tr>
<td>Aligned to genome</td>
<td>18 M, 83.40%</td>
<td>20 M, 78.24%</td>
<td>23 M, 84.47%</td>
</tr>
<tr>
<td>Uniquely aligned to genome</td>
<td>17 M, 79.07%</td>
<td>19 M, 73.70%</td>
<td>22 M, 80.40%</td>
</tr>
<tr>
<td>Detected genes (skipping isoform)</td>
<td>19447</td>
<td>19396</td>
<td>20125</td>
</tr>
</tbody>
</table>

1 Read counts in the middle section are expressed in millions (left) or as a percentage of the total reads processed (right) for each sample.

...sents reads per kilobase of transcript per million mapped reads) for each gene, according to their genome location. Other statistical results, such as gene coverage and depth, reads distribution around start codon and stop codon, were also obtained.

**Defining the differentially expressed genes (DEGs) between two samples**

Each patient is considered as a biological replicate, and the diseased tissues and paired normal tissues taken from the three patients are independent of each other. DEGs between the diseased tissue samples and paired normal tissue samples were analyzed by using edgeR [12], one of R packages. For each gene, the P-value and FDR were obtained based on the model of negative binomial distribution. The fold changes were also estimated within the edgeR statistical package. 0.01 p-value and 2 fold change were set as the threshold to define DEGs.

**Hierarchical clustering analysis**

Hierarchical clustering was performed to calculate the cluster of gene set by Cluster 3.0 software. Heatmaps showing expression profiles (log2 RPKM) were generated using Java TreeView.

**Functional enrichment analysis**

Lists of DEGs were submitted to DAVID [13, 14] web server (http://david.abcc.ncifcrf.gov/) for enrichment analysis. Enrichment clusters were sorted by the enrichment score on descending order. Categories of one cluster were sorted by p-value on descending order. Fold Enrichment, Bonferroni and Benjamini corrected p-value and FDR were also presented for each category of each cluster.

**qRT-PCR**

To validate the accuracy of our RNA-seq profiling results, 6 randomly selected candidate genes were evaluated by qRT-PCR. Total RNA was extracted from MED and normal tissue samples with TRIzol kit (Invitrogen, Carlsbad, CA, USA), and the total RNA was reverse transcribed by using M-MLV Reverse Transcriptase according to the manufacturer's protocol (Promega, Madison, WI, USA). The primer sequences were listed in Table S1. GAPDH was used as an internal control. Real-time monitoring of PCR was performed with ABI3700 (Applied Biosystems, Grand Island, NY, USA) and TransStart Top Green qPCR SuperMix (TransGen Biotech, Beijing, China). Reaction were performed at 94°C for 30 s, and then cycled at 94°C for 5 s, 58°C for 30 s, 72°C for 31 s for 40 cycles. Each assay was performed in triplicate. Data analysis was performed using the 2ΔΔCt method described previously [15].

**Results**

All three patients were histologically diagnosed as MED (Figure 1) by at least two oral pathologists according to WHO criteria [10] and underwent surgical excision. Due to the relative low incidence of oral dysplasia in central China, only a small number of MED biopsy samples from three patients were obtained for RNA-seq analysis.

**Measurement of gene expression in diseased and normal tissue by RNA-seq**

We sequenced total RNA extracted from diseased and paired normal tissues from three patients with MED, yielding 20-32 million 100 bp long sequence reads per sample. An overview of the results is displayed in Table 1. The majority of reads from each sample (78%-85%) were successfully aligned; those that did not align are likely to be polyclonal, low quality or have origins outside the reference human genome. Between 16 and 25 million reads per sample were uniquely aligned to the human genome and over 19,000 expressed genes...
Hierarchical clustering analysis of gene expression levels in each of MED and paired-normal samples

Hierarchical clustering analysis was performed to investigate the expression patterns of genes during MED pathogenesis. As shown in Figure S1, MED_#3 was assigned to the adjacent node of the corresponding normal tissue (Normal_#3), whereas MED_#1 and MED_#2 were separated from their corresponding normal tissues (Normal_#1 and Normal_#2). This finding suggested that the gene expression pattern of MED_#3 was similar with Normal_#3, while the gene expression patterns of MED_#1 and MED_#2 were distinct from Normal_#1 and Normal_#2, respectively.

Identification of differentially expressed genes in MED

To identify genes associated with MED, we examined the gene expression profiles in normal and dysplastic tissues. A total of 132 genes showed significant upregulation ($P$-value < 0.01 and FC > 2) in MEDs compared with normal tissues (Table S2), and 214 genes differentially downregulated ($P$-value < 0.01 and FC < 0.5) between MEDs and normal tissues (Table S3). Apparently, the number of genes differentially downregulated between MED and normal tissues was almost 2-fold greater than the amount of upregulated genes. Further, the expression levels of differentially expressed genes were hierarchically clustered, and the normal tissues and MEDs form tight clusters (Figure 2). MEDs of patient #1 and #2 were classified into the same cluster, whereas MED of patient #3 formed a cluster with a distinct pattern of gene expression. This finding suggested that patient #1 and #2 have more similar gene expression features than patient #3 during the conversion from normal tissues to moderate dysplasia.

Next, an additional 1 RPKM filter was added to ensure confidence values of gene expression. As shown in Table 2, six up-regulated genes and twelve down-regulated genes showed an expression no less than 1 RPKM in all six tissue samples, these genes are likely to be closely related to the pathogenesis of oral dysplasia.

To more systematically assess the biological functions of commonly mis-regulated genes,
DAVID online platform was employed for enrichment analysis (Tables S4 and S5). A selection of the results is displayed in Table S6. The process of signal prominently amongst both the up and down regulated gene sets. Whereas, the upregulated genes tend to function in epidermis development, tumor antigen, cytokine-cytokine receptor interaction, and the downregulated genes more often function in glycosylation, mucosa protection, cell-cell signaling, ion transport, membrane-bounded vesicle, hexose metabolic process, cell-cell adhesion and regulation of secretion.

Validation of RNA-seq data by quantitative RT-PCR

In order to verify the expression pattern of those genes involved in the pathogenesis of moderate dysplasia, six mis-regulated genes were selected for further qRT-PCR analysis. Consistently, with the RNA-seq data, mRNA expression level of ISG15 was significantly increased between the normal tissues and MEDs in patient #1 and #2, and MMP3 expression level in patient #1 was sharply increased in MED compared to the normal tissue, the expression level of MMP3 in patient #2 and #3 were not detectable. On the contrary, TUSC5 and KRT4 expression levels were declined during the conversion from normal tissues to MEDs in all three patients (Figure 3). Taken together, although the exact fold changes of the selected genes in three patients varied between RNA-seq data and qRT-PCR analysis, most selected genes (13/18) showed a similar expression trends found from the two different approaches. These findings indicated that the data obtained from RNA-seq analysis was reliable. Further, we focused on ISG15 among these identified misregulated genes, because previous microarray studies have indicated that the expression of this gene is enhanced according to the degree of malignant transformation [7, 16].

To evaluate the mRNA expression of ISG15, qRT-PCR assay was performed on three moderate dysplasia and two severe dysplasia samples. As shown in Figure 4, 1.39 to 9.35 fold upregulation of ISG15 was detected in moderate dysplasia samples, and much higher mRNA expression (59.5 fold and 22.94 fold) was found in severe dysplasia samples. Our data further proved that the enhancement of ISG15 expression is correlated with the progression of epithelial dysplasia to oral cancer.
Discussion

By sequencing the diseased and paired normal transcriptomes of three individuals with MED, we have characterized the changes in gene expression associated with oral dysplasia development.

Hierarchical clustering analysis was performed to correlate the expression patterns of genes between MEDs and normal samples (Figure S1). MED tissue from patient #3 was assigned to the adjacent node of the paired normal tissue (Normal_#3), while MED tissue from patient #2 was assigned to the adjacent node of the normal tissue of patient #1 (Normal_#1), which was well correlated with the individual difference. Furthermore, a recent study including both OSCC and oral dysplasia samples presented that normal tissues and the majority of oral dysplasia samples had similar gene expression patterns and classified into the same cluster, whereas cancer tissues formed a cluster with a distinct pattern of gene expression [7], demonstrating gene expression patterns in dysplasia tissues were possibly less consistent than those in tumor tissues. Additionally, clustering of the differentially expressed genes revealed a closer gene expression features between the MEDs of patient #1 and #2 when compared with the MED of patient #3 (Figure 2). Taking together, in present study the same type of dysplasia tissues of the three patients exhibiting quite different gene expression patterns maybe associated with the different malignant potential of these specimens.
Among the genes differentially expressed in MEDs versus paired normal tissues, the most significantly regulated genes were found associated with multiple malignant bioprocesses. Proinflammatory cytokines promote systemic inflammation and have long been associated with many kinds of diseases [17]. Interleukin 6 (IL6), interleukin 11 (IL11), and interleukin 19 (IL19) all has been implicated to signal transduction through both Janus kinase 1 and 2 [18] whose activation has been widely demonstrated to be relevant various types of cancer, including OSCC [19]. A microarray-based transcriptional profiling also revealed that inflammatory cascades were the most perturbed pathway in oral dysplastic tissue [20]. Angiogenic stimulation related MMP3 was secreted by carcinoma-associated fibroblasts and mediated the cleavage of the extracellular domain of the adhesive protein E-cadherin on the surface of mammary epithelial cells [21]. Inhibin beta A (INHBA) is a subunit of both activin and inhibin protein complexes [22], both of which belong to transforming growth factor beta (TGF-β) protein superfamily [23]. Activin promotes cell proliferation and differentiation, wherein inhibin exerts tumor-suppressor activity [24]. As described before [9], the increased expression levels of INHBA in current MED tissues more likely reflected its biological roles as activin complexes. Melanoma-associated antigen-A (MAGEA) constitute a sub-family of Cancer-Testis Antigens, its expression is restricted to germ cells and various tumors, including OSCC [25]. A recent large sample size study revealed that, between progressive and non-progressive oral leukoplakia, MAGEAs expression was only detected in progressive oral leukoplakia, and showed no correlation with non-progressive leukoplakia or the grade of dysplasia [26]. At present study, only patient #3 had an evidently higher expression of MAGEAs in MED compared with the normal tissue (Table S2), indicating patient #3 may manifest a progressive oral leukoplakia.

Interferon-stimulated gene 15 (ISG15) is strongly upregulated by interferons (IFNs), pathogen infections, and cellular stresses [27], numerous studies indicated a potential role of free ISG15 as a cytokine or chemokine [28] in both intracellular and extracellular compartments. The increased expression of ISG15 has been associated with pancreatic adenocarcinoma, endometrial cancer, bladder cancer and OSCC [16, 29-31]. A recent investigation in ISG15 expression at both mRNA and protein levels in oral dysplasia and invasive OSCC tissues demonstrated the ISG15 expression level was enhanced according to the degree of malignant transformation [7], indicating ISG15 may serve as a potential marker to discriminate between oral dysplasias with and without malignant potential. In current study, ISG15 also showed an upregulated expression pattern during the conversion from normal tissues to moderate dysplasia samples. Moreover, among these three patients, patient #1 has the greatest degree of ISG15 upregulation between normal tissues and MED (4.66-fold), patient #2 showed a 1.54-fold upregulation in ISG15 and ranked the second, and patient #3 has relatively similar expression levels between normal and MED samples (Table S2). RNA-seq data suggested that patient #1 had a higher risk of progression to OSCC transformation than the other two patients. Our qRT-PCR assay performed on another three moderate dysplasia and two severe dysplasia specimens provided extra evidence for this conclusion (Figure 4). Hierarchical clustering analysis unveiled that the transcription profile of ISG15 was close related with the gene expression patterns in normal and MED specimens (Figure 2). The differential expressed genes in the MEDs of patient #1 and #2 showed similar gene expression features, while differential expressed genes in patient #3 MED have a distinct gene expression pattern. This findings suggested that ISG15, along with closely related mis-expressed genes, may involve in the malignant transformation of oral dysplasia to OSCC.

Accession numbers

The accession number for the RNA-seq data reported in this paper is GEO: GSE72627.

Acknowledgements

This project was funded by National Natural Science Foundation of China (Grant No. 81170971 and 81172571).

Disclosure of conflict of interest

None.

Address correspondence to: Xinming Chen, Department of Oral Histopathology, School and
Transcriptome profiling of oral dysplasia

References


Transcriptome profiling of oral dysplasia


Figure S1. Hierarchical clustering analysis of gene expression levels in each of the six samples. Shades of blue indicate lowered expression, relative to the mean across samples, whereas shades of pink indicate higher expression relative to the mean.