Int J Clin Exp Pathol 2016;9(6):6107-6116
www.ijcep.com /ISSN:1936-2625/IJCEP0024835

*Original Article*

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

Jiali Zhang[1,2], Si Chen[1,3], Lan Yin[4], Xinming Chen[1,2]

*[1]The 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; [2]Department of Oral Histopathology, School and Hospital of Stomatology, Wuhan University, Wuhan 430079, China; [3]Department of Oral Implantology, School and Hospital of Stomatology, Wuhan University, Wuhan 430079, China; [4]ABLife, 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

relies 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. Microdisected tissue was directly placed into TRIzol (Invitrogen, Carlsbad, CA, USA) in RNAase 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% agrose 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

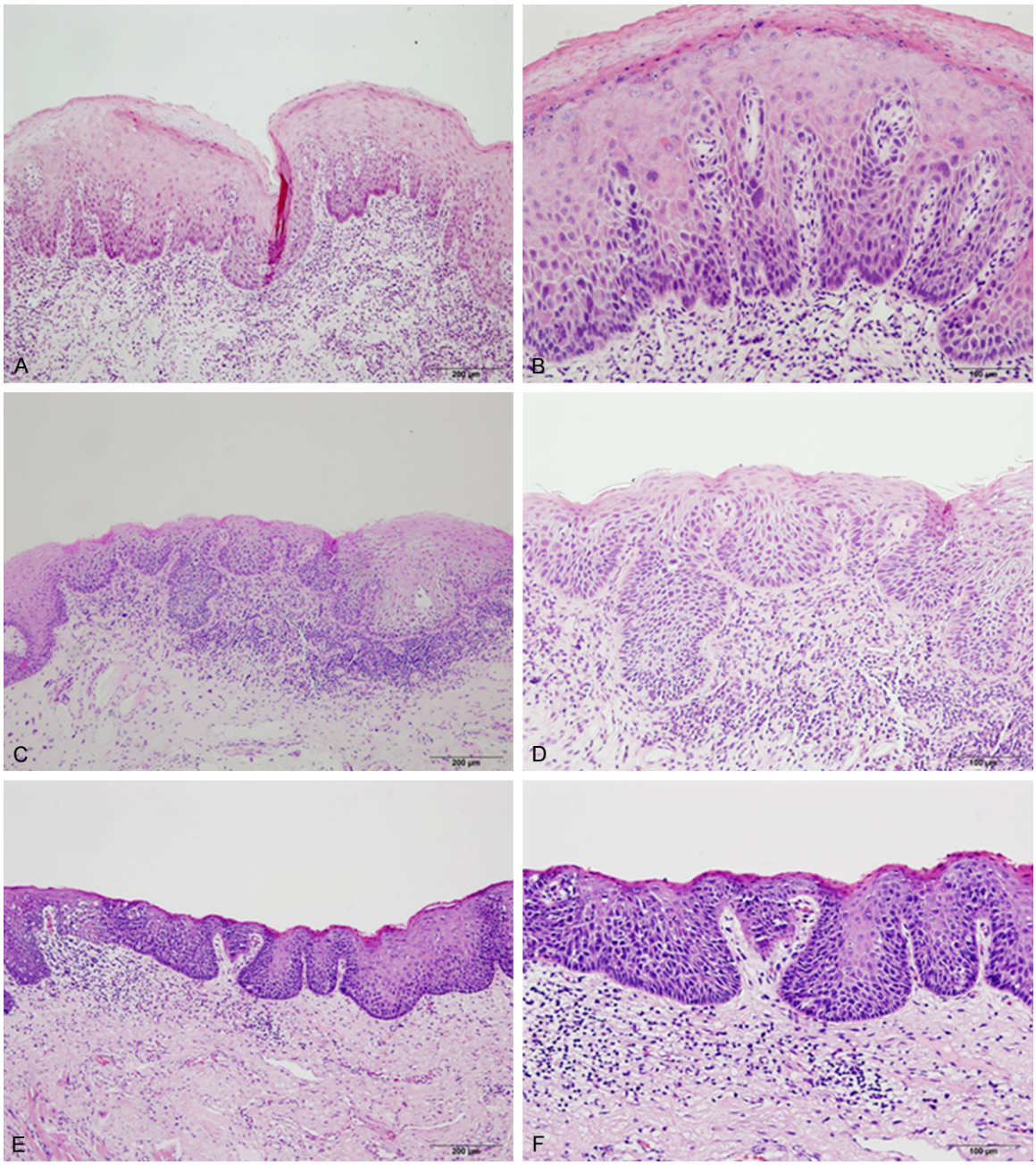


**Figure 1.** Histopathological examination of biopsy specimens from the three patients with MED. A-F. Photomicrographs of hematoxylin and eosin H&E. stained slides demonstrate a proliferation of atypical squamous cells extending into the middle third of the epithelium, basal cell hyperplasia, regional 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.

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-

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

| Sample | Patient #1 | | Patient #2 | | Patient #3 | |
|---|---|---|---|---|---|---|
| | Normal | MED | Normal | MED | Normal | MED |
| Total reads | 22 M | 26 M | 28 M | 20 M | 28 M | 32 M |
| Clean reads | 20 M, 91.37% | 23 M, 86.70% | 25 M, 90.96% | 18 M, 90.61% | 25 M, 87.68% | 28 M, 88.73% |
| Aligned to genome | 18 M, 83.40% | 20 M, 78.24% | 23 M, 84.47% | 17 M, 83.79% | 23 M, 81.34% | 27 M, 82.89% |
| Uniquely aligned to genome | 17 M, 79.07% | 19 M, 73.70% | 22 M, 80.40% | 16 M, 80.26% | 22 M, 77.38% | 25 M, 78.81% |
| Detected genes (skipping isoform) | 19447 | 19396 | 20125 | 19490 | 20085 | 20203 |

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 ($\log_2$ 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 $2^{-\Delta\Delta Ct}$ 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

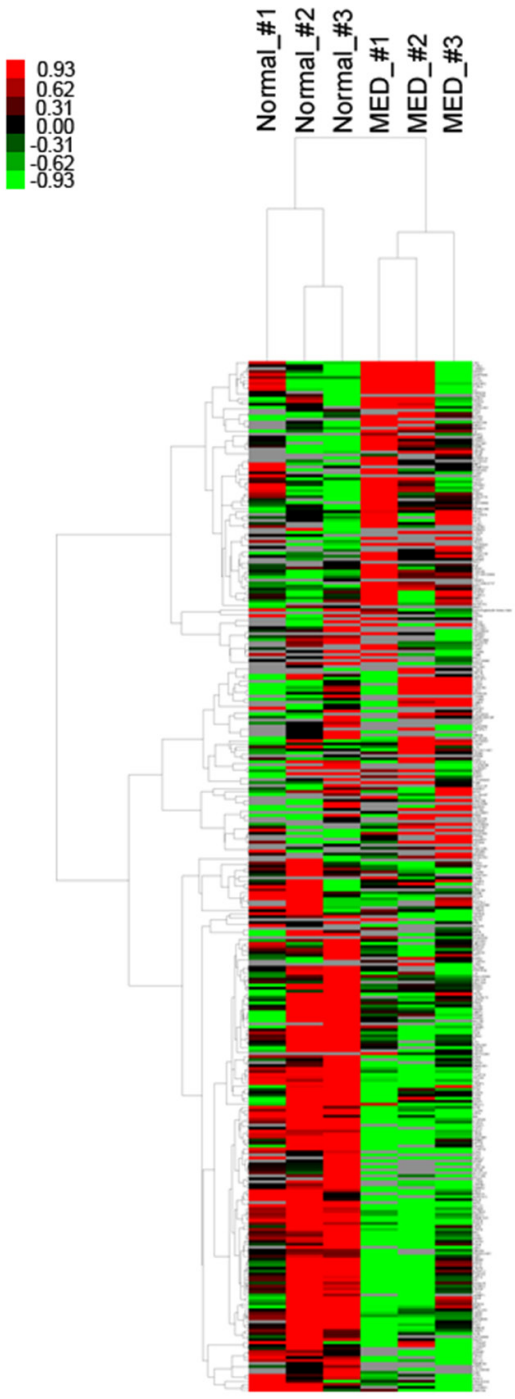


**Figure 2.** Hierarchical clustering analysis of differentially expressed genes. The heatmap illustrates the expression profiles of differentially expressed genes ($log_2$ RPKM, y-axis) in six tissue samples (x-axis). The red/green color code corresponds to the expression deviation from the average expression among all samples. Number is case number.

were detected in each sample (**Table 1**). On the whole, 21556 of the 28518 annotated genes (75.59%) were detected in these MED and matched-normal RNA samples.

*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,

**Table 2.** Identified 18 candidate genes expressed no less than 1 RPKM in all six tissue samples

| Gene symbol | UCSC ID | logFC | logCPM | *p*-value | #1 | | #2 | | #3 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Normal | MED | Normal | MED | Normal | MED |
| Six upregulated genes | | | | | | | | | | |
| ISG15 | uc001acj.4 | 1.7844919 | 6.6576219 | 0.0090016 | 98.97 | 460.92 | 74.49 | 114.8 | 43.37 | 55.46 |
| ART3 | uc003hji.3 | 2.4420135 | 6.958955 | 0.000475 | 41.73 | 369.78 | 45.35 | 97.77 | 32.02 | 79.86 |
| BCL2A1 | uc002bfc.4 | 1.9063964 | 4.7878854 | 0.0056437 | 30.76 | 104.74 | 5.02 | 17.91 | 4.51 | 7.5 |
| AF086184 | uc001pen.1 | 1.7827139 | 6.0446015 | 0.0090791 | 21.13 | 10.3 | 56.83 | 265.5 | 61 | 191.01 |
| WTAPP1 | uc001phh.1 | 1.7978784 | 5.0401065 | 0.0086838 | 10.41 | 31.96 | 1.66 | 11.65 | 5.36 | 10.23 |
| CTHRC1 | uc003ylk.4 | 1.8798286 | 4.3419185 | 0.0063349 | 9.85 | 32.87 | 3.95 | 4.09 | 7.79 | 35.4 |
| Twelve downregulated genes | | | | | | | | | | |
| KRT4 | uc031qhk.1 | -4.666401 | 10.539026 | 1.42E-09 | 743.2 | 4.09 | 3166.1 | 133.04 | 470.21 | 30.59 |
| CRNN | uc001ezx.2 | -2.44704 | 9.6463353 | 0.0004596 | 698.35 | 29.54 | 659.58 | 143.97 | 938.52 | 244 |
| MAL | uc002stx.2 | -1.902885 | 6.4894901 | 0.0055258 | 126.13 | 15.96 | 199.52 | 54.49 | 76.68 | 33.13 |
| KRT19 | uc002hxd.4 | -2.873009 | 6.1388636 | 5.32E-05 | 38.76 | 7.74 | 182.15 | 2.81 | 56.73 | 25.24 |
| CLCA4 | uc009wcs.3 | -2.702051 | 6.334215 | 0.0001304 | 25.84 | 1.56 | 63.8 | 12.94 | 53.66 | 6.72 |
| ADH1B | uc003hus.4 | -2.963738 | 5.8686567 | 3.37E-05 | 24.56 | 5.46 | 64.96 | 3.05 | 36.97 | 6.11 |
| PRH1-PRR4 | uc001qzb.4 | -1.769372 | 6.3903275 | 0.0095922 | 17.3 | 32.77 | 31.4 | 19.66 | 226.66 | 16.33 |
| GREM2 | uc001hys.3 | -1.78758 | 3.5427541 | 0.0092801 | 11.97 | 4.06 | 38.15 | 3.08 | 24.57 | 13.32 |
| LTF | uc003cpq.3 | -3.146563 | 6.4928228 | 1.23E-05 | 7.47 | 4.82 | 20.57 | 14.16 | 184.76 | 2.57 |
| AZGP1 | uc003ush.3 | -3.115075 | 5.5408003 | 1.50E-05 | 4.69 | 2.9 | 22.78 | 10.04 | 194.94 | 11.02 |
| PRR4 | uc001qyz.4 | -7.460766 | 7.1376237 | 2.83E-17 | 1.65 | 3.39 | 79.97 | 2.58 | 1602.83 | 2.45 |
| LCN1 | uc004cfz.2 | -8.598177 | 8.6345969 | 1.22E-20 | 1.1 | 2.65 | 2.16 | 2.41 | 3434.1 | 2.68 |

DEGs were filtered for a threshold of 1 RPKM in all six samples.UCSC means UCSC Genome Browser website (genome.ucsc.edu).

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 misregualted 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.

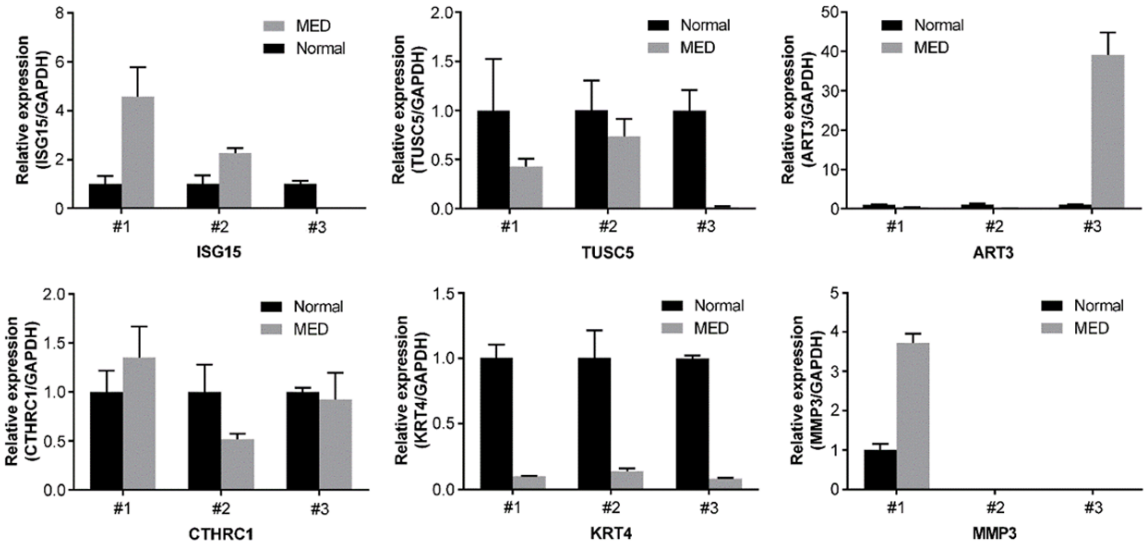


**Figure 3.** qRT-PCR analysis of the mRNA expression of six selected genes in normal tissues and MEDs. Mean values ± standard deviation of three replicated measurements.

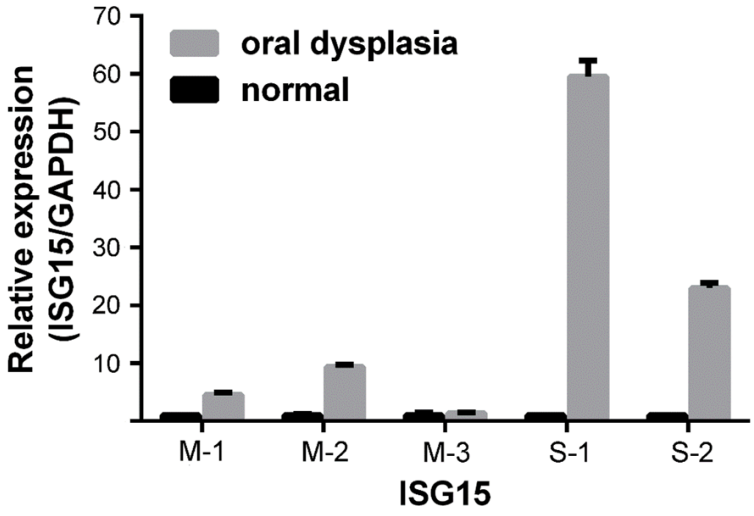


**Figure 4.** qRT-PCR quantification of ISG15 in three moderate dysplasia and two severe dysplasia specimens. M- means moderate dysplasia, S- means severe dysplasia.

## 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 misexpressed 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

Hospital of Stomatology, Wuhan University, Wuhan 430079, China. Tel: +86-27-87686229; Fax: +86-27-87647443; E-mail: xmchen3011@126.com

## References


[1] Parkin DM, Bray F, Ferlay J and Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005; 55: 74-108.

[2] Balasundaram I, Payne K, Al-Hadad I, Alibhai M, Thomas S and Bhandari R. Is there any benefit in surgery for potentially malignant disorders of the oral cavity? J Oral Pathol Med 2014; 43: 239-244.

[3] Bouquot JE, Speight PM and Farthing PM. Epithelial dysplasia of the oral mucosa-Diagnostic problems and prognostic features. Curr Diagn Pathol 2006; 12: 11-21.

[4] Napier SS and Speight PM. Natural history of potentially malignant oral lesions and conditions: an overview of the literature. J Oral Pathol Med 2008; 37: 1-10.

[5] Brothwell DJ, Lewis DW, Bradley G, Leong I, Jordan RC, Mock D, Leake JL. Observer agreement in the grading of oral epithelial dysplasia. Community Dent Oral Epidemiol 2003; 31: 300-305.

[6] Yu YH, Kuo HK, Chang KW and Rutherford S. The evolving transcriptome of head and neck squamous cell carcinoma: a systematic review. PLoS One 2008; 3: e3215.

[7] Sumino J, Uzawa N, Okada N, Miyaguchi K, Mogushi K, Takahashi KI, Sato H, Michikawa C, Nakata Y and Tanaka H. Gene expression changes in initiation and progression of oral squamous cell carcinomas revealed by laser microdissection and oligonucleotide microarray analysis. Int J Cancer 2013; 132: 540-548.

[8] Wang Z, Gerstein M and Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009; 10: 57-63.

[9] Tuch BB, Laborde RR, Xu X, Gu J, Chung CB, Monighetti CK, Stanley SJ, Olsen KD, Kasperbauer JL and Moore EJ. Tumor transcriptome sequencing reveals allelic expression imbalances associated with copy number alterations. PLoS One 2010; 5: e9317.

[10] Barnes L. Pathology and genetics of head and neck tumours. IARC, 2005.

[11] Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R and Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 2013; 14: R36.

[12] Robinson MD, McCarthy DJ and Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010; 26: 139-140.

[13] Huang DW, Sherman BT and Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids Res 2009; 37: 1-13.

[14] Huang DW, Sherman BT and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 2008; 4: 44-57.

[15] Livak KJ and Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 2001; 25: 402-408.

[16] Ye H, Yu T, Temam S, Ziober BL, Wang J, Schwartz JL, Mao L, Wong DT and Zhou X. Transcriptomic dissection of tongue squamous cell carcinoma. BMC Genomics 2008; 9: 69.

[17] Genctoy G, Altun B, Alper Kiykim A, Arici M, Erdem Y, Çağlar M, Yasavul Ü, Turgan Ç and Çağlar Ş. TNF Alpha-308 Genotype and Renin-Angiotensin System in Hemodialysis Patients: An Effect on Inflammatory Cytokine Levels? Artif Organs 2005; 29: 174-178.

[18] Quintás-Cardama A, Kantarjian H, Cortes J and Verstovsek S. Janus kinase inhibitors for the treatment of myeloproliferative neoplasias and beyond. Nat Rev Drug Discov 2011; 10: 127-140.

[19] Rossa C Jr, Sommer G, Spolidorio LC, Rosenzweig SA, Watson DK, Kirkwood KL. Loss of expression and function of SOCS3 is an early event in HNSCC: altered subcellular localization as a possible mechanism involved in proliferation, migration and invasion. PLoS One 2012; 7: e45197.

[20] Banerjee AG, Bhattacharyya I and Vishwanatha JK. Identification of genes and molecular pathways involved in the progression of premalignant oral epithelia. Mol Cancer Ther 2005; 4: 865-875.

[21] Watnick RS. The role of the tumor microenvironment in regulating angiogenesis. Cold Spring Harb Perspect Med 2012; 2: a006676.

[22] Ling N, Ying SY, Ueno N, Shimasaki S, Esch F, Hotta M and Guillemin R. Pituitary FSH is released by a heterodimer of the beta-subunits from the two forms of inhibin. Nature 1986; 321: 779-82.

[23] Kingsley DM. The TGF-beta superfamily: new members, new receptors, and new genetic tests of function in different organisms. Genes Dev 1994; 8: 133-146.

[24] Lopez P, Vidal F, Rassoulzadegan M and Cuzin F. A role of inhibin as a tumor suppressor in Sertoli cells: down-regulation upon aging and repression by a viral oncogene. Oncogene 1999; 18: 7303-7309.

[25] Müller-Richter UD, Dowejko A, Peters S, Rauthe S, Reuther T, Gattenlöhner S, Reichert TE, Driemel O and Kübler AC. MAGE-A antigens in

patients with primary oral squamous cell carcinoma. Clin Oral Investig 2010; 14: 291-296.
[26] Ries J, Agaimy A, Vairaktaris E, Kwon Y, Neukam FW, Strassburg LH and Nkenke E. Evaluation of MAGE-A expression and grade of dysplasia for predicting malignant progression of oral leukoplakia. Int J Oncol 2012; 41: 1085-1093.
[27] Zhang D and Zhang DE. Interferon-stimulated gene 15 and the protein ISGylation system. J Interferon Cytokine Res 2011; 31: 119-130.
[28] Owhashi M, Taoka Y, Ishii K, Nakazawa S, Uemura H and Kambara H. Identification of a ubiquitin family protein as a novel neutrophil chemotactic factor. Biochem Biophys Res Commun 2003; 309: 533-539.
[29] Chi LM, Lee CW, Chang KP, Hao SP, Lee HM, Liang Y, Hsueh C, Yu CJ, Lee IN and Chang YJ. Enhanced interferon signaling pathway in oral cancer revealed by quantitative proteome analysis of microdissected specimens using 160/180 labeling and integrated two-dimensional LC-ESI-MALDI tandem MS. Mol Cell Proteomics 2009; 8: 1453-1474.
[30] Andersen JB, Aaboe M, Borden E, Goloubeva O, Hassel B and Ørntoft TF. Stage-associated overexpression of the ubiquitin-like protein, ISG15, in bladder cancer. Br J Cancer 2006; 94: 1465-1471.
[31] Desai SD, Haas AL, Wood LM, Tsai Y-C, Pestka S, Rubin EH, Saleem A, Nur-E-Kamal A and Liu LF. Elevated expression of ISG15 in tumor cells interferes with the ubiquitin/26S proteasome pathway. Cancer Res 2006; 66: 921-928.

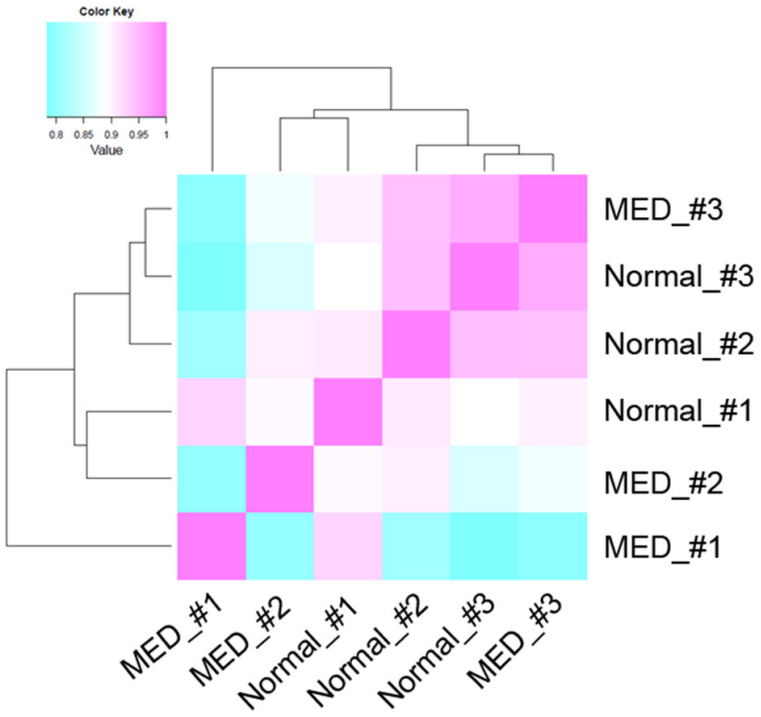


**Figure S1.** Hierarchical clustering analysis of gene expression levels in each of the six samples. Shades of blue indicate lowered expression, relative to the mean across samples, whereas shades of pink indicate higher expression relative to the mean.