Molecular Classification of Autofluorescence Excision Margins in Oral Potentially Malignant Disorders

Camile S Farah¹,², Farzaneh Kordbacheh², Keziah John², Nigel Bennett², Simon A Fox¹

¹Australian Centre for Oral Oncology Research & Education, UWA Dental School, University of Western Australia, Nedlands, WA, Australia.

²University of Queensland Centre for Clinical Research, The University of Queensland, Herston, QLD, Australia.

Corresponding Author:
Professor Camile Farah
UWA Dental School, University of Western Australia
17 Monash Ave, Nedlands WA, 6009, Australia
T: 08 9346 7636
E: camile.farah@uwa.edu.au

Running title: OPMD Molecular Autofluorescence Excision Margins

Keywords: oral potentially malignant disorders; autofluorescence; excision margins; RNA sequencing; gene expression profiling; Velscope

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/odi.12818

This article is protected by copyright. All rights reserved.
Abstract

Objective
To define molecular differences between autofluorescence and white light defined excision margins in oral potentially malignant disorders (OPMD) using transcriptome expression profiles.

Materials and Methods
Excisional biopsy specimens were taken from 11 patients at three different sites for each lesion: centre, white light margin, and autofluorescence margin. The lesions were diagnosed histopathologically as oral epithelial dysplasia, oral lichenoid dysplasia, oral lichen planus or other. Transcriptome analysis was performed by RNA sequencing, hierarchical clustering, differential expression and biological pathway analysis.

Results
For hierarchical clustering the samples broadly clustered according to histology rather than the margins with lichenoid samples clustering together. Differential expression analysis showed that independent of histology, there was greater molecular dysregulation between the lesion centre and autofluorescence margin compared to the lesion centre and white light margin. Furthermore, the autofluorescence and white light margins were molecularly distinct indicating the white light margins harboured abnormality.

Conclusion
Our results indicate that the molecular profile of OPMD changes with divergence away from the centre of the lesion, and that autofluorescence determined margins are superior to the white light margin in achieving a clear molecular margin when excising an OPMD.
Introduction

Early detection and treatment of oral potentially malignant disorders (OPMD) and oral squamous cell carcinoma (OSCC) is the key to reducing the burden of patient morbidity and mortality from oral cancer (Brocklehurst et al., 2013, Walsh et al., 2013). The current gold standard for clinical detection of OPMD is conventional oral examination (COE) employing visual and tactile examination under white light (Farah & McCullough, 2008). However, this approach has deficiencies since many subtle mucosal changes may be missed (Bhatia et al., 2014). Furthermore, there is histological evidence of dysplasia and microinvasive carcinoma found in mucosa assessed as clinically normal by COE (Thomson, 2002). This has led to new technologies being developed which highlight OPMDs aiding early diagnosis and management of otherwise occult mucosal lesions, whether benign or malignant, and providing some information as to the underlying molecular nature of the mucosal abnormality (Bhatia et al., 2013, Lingen et al., 2008, McCullough et al., 2010, Rethman et al., 2010).

Tissue autofluorescence (AF) is one technique which has been employed to enhance the detection and visualisation of OPMD and early OSCC (Bhatia et al., 2013). This method is based upon the principle that endogenous fluorophores can be excited to fluoresce under external illumination in the range of 400-460 nm (Richards-Kortum & Sevick-Muraca, 1996, Pavlova et al., 2008, Rethman et al., 2010, Poh et al., 2009, De Veld et al., 2005, Wu & Qu, 2006, Roblyer et al., 2009). Since each fluorophore is associated with specific excitation and emission wavelengths, changes in tissue architecture and concentrations of fluorophores results in altered absorption and scattering properties of the tissue (Bhatia et al., 2013). Typically, dysplastic and malignant tissues exhibit decreased autofluorescence (Pavlova et al., 2009, Pavlova et al., 2008, Poh et al., 2006). VELscope™ (LED Medical Diagnostics Inc., Barnaby Canada) is a hand-held device for the enhanced visualisation of oral mucosal abnormalities by autofluorescence. Under the VELscope™, normal oral mucosa appears pale green, while abnormal tissue appears darker in colour due to loss of autofluorescence (LAF) (Rethman et al., 2010).

VELscope™ has demonstrated benefit for assisting the detection of oral epithelial dysplasia (OED) and OSCC not visible by COE warranting tissue biopsy and aiding in demarcating margins (Kois & Truelove, 2006, Poh et al., 2006). However, since LAF may also occur in benign mucosal inflammation leading to false positives, clinicians have been advised to use the VELscope™ in conjunction with COE (Farah et al., 2012). Furthermore, many inflammatory conditions display diascopic fluorescence (DF), wherein tissues display normal fluorescence pattern with the application of pressure and this can assist in differentiating inflammatory from neoplastic lesions (Farah et al., 2012, Bhatia et al., 2014). We have previously investigated the molecular mechanisms underlying LAF and DF in OPMD using RNA sequencing to elucidate correlations between...
autofluorescence behaviour and differential transcriptional mechanisms across histological groups (Kordbacheh et al., 2016). It is recognised that the feature of autofluorescence that make it advantageous in the detection of OPMD can also make it useful for the determination of excision margins (Poh et al., 2006, Elvers et al., 2015, Ohnishi et al., 2016). However, there are few studies that provide molecular validation for this application.

The aim of this study was to use RNA sequencing to molecularly validate the accuracy of oral autofluorescence imaging compared to white light visualisation for determination of surgical excision margins of OPMD. We analysed the transcriptome of samples from the centre of the lesion (L), white light margin (WL), and autofluorescence (VELscope™) margin (AF), in a cohort of patients undergoing surgical resection of OPMD. By bioinformatic analysis of gene expression profiles we were able to determine the degree of molecular dysregulation across the margins and gain insight into differentially regulated biological pathways.

**Materials and methods**

**Patients and Samples**
A total of 36 fresh frozen oral cavity biopsies from 11 patients were collected prospectively for use in the study. Samples were stored in RNAlater® (Ambion, Life Technologies, Carlsbad, USA) at -80°C acquired from patients referred for investigation of suspicious oral mucosal lesions between 2010 and 2013. For each patient, biopsies were taken at centre of the lesion (L), at the white light margin (WL), and at the autofluorescence (VELscope™) determined margin (AF). 4 mm punch biopsies were taken at the margins, not extending more than 1-2 mm beyond the autofluorescence margin, ensuring the biopsies were sitting within the lesion itself (example shown in Figure 1). A portion of each biopsy was reserved for histopathological diagnosis which was undertaken separately by three oral pathologists with final diagnosis based upon consensus. Criteria used for oral lichen planus and oral lichenoid lesions were according to those by van der Meij and van der Waal (van der Meij & van der Waal, 2003), while oral lichenoid dysplasia was according to Krutchkoff and Eisenberg (Krutchkoff & Eisenberg, 1985).

The study was approved by The University of Queensland and the Royal Brisbane & Women’s Hospital Human Research Ethics Committees (project numbers UQ2007001478 and HREC/10/QRBW336). Informed consent was obtained for all patients in the study in accordance with the Declaration of Helsinki.
RNA isolation, library preparation and sequencing

Total RNA extraction and quality control was carried out for all samples as previously described by us (Kordbacheh et al., 2016). Briefly, frozen biopsies were ground in liquid nitrogen and RNA extracted using an AllPrep DNA/RNA/miRNA Universal kit (Qiagen, Netherlands) and quantity and quality assessed using NanoDrop spectrophotometer (Thermofisher Scientific, MA, USA), Qubit fluorometer (Thermofisher) and 2100 Bioanalyzer (Agilent Technologies, CA, USA). The RiboMinus™ Eukaryote System v2 kit (Life Technologies, USA) was used to deplete ribosomal RNA (rRNA). 50ng of rRNA depleted RNA input was used for library preparation with an Ion Total RNA-Seq Kit v2 (Thermofisher), barcoded with the Ion Xpress™ RNA-Seq Barcode 1-16 Kit (Thermofisher) according to the manufacturer’s instructions. Templating and chip loading was performed with 110pM of each library using the Ion PI IC kit (ThermoFisher) on an automated Ion Chef system (ThermoFisher) using Ion PIv3 sequencing chips. Loaded chips were sequenced on a Proton sequencer (ThermoFisher) with data collection and adaptor trimming using Torrent Suite v4.2 software.

Sequencing data filtering, alignment and exploratory analysis

Prior to further analysis raw reads were quality checked with fastQC R tool and pre-processed to remove low quality sequence using trimmomatic R tool (Bolger et al., 2014). The processed reads were mapped to reference genome hg19 using Burrows-Wheeler Aligner (BWA-MEM) (Li & Durbin, 2009). Following the mapping step, the alignment files were analysed with RNAseQC for descriptive quality control (QC) of the mapped reads. Read count matrices of summarised data for each gene were generated using the R package Rsubread (Liao et al., 2014). Data was transformed by variance stabilizing transformation (VST) implemented in voom prior to downstream analysis. For exploratory analysis of the relationship between samples hierarchical clustering was performed using the Morpheus webserver (https://software.broadinstitute.org/morpheus/).

Differential Expression Analysis

Differential gene expression analysis was performed with the R/Bioconductor packages limma and DESeq2. Limma uses linear modelling (Smyth, 2004) while DEseq2 uses a negative binomial model (Love et al., 2014). Analysis with these packages was performed using standard parameters. Voom transformed data was used for limma while for DEseq2 input was read counts since DEseq2 incorporates variance stabilization (Love et al., 2014). Adjusted p-values (padj) for multiple testing, using Benjamini-Hochberg to estimate the false discovery rate (FDR), were calculated for final estimation of differential expression (DE) significance using a 0.1 cutoff. We also used discriminant analysis to identify genes whose differential expression stratifies between the different margins. The
discriminant method was sparse partial least square discriminant analysis (sPLS-DA) which is a supervised approach to classification implemented in mixOmics (Le Cao et al., 2009). sPLS-DA was run with two components and 100 genes to be selected per component to separate each group. A cross-validation method (re-running the sPLS-DA many times removing one sample at a time) identified genes that were selected in at least 70% of the runs in either of the components.

**Gene set enrichment analysis for cancer related genes, biological pathways and gene ontology terms**

Overrepresentation analysis (ORA) of differentially expressed genes for pathways and gene ontology terms was performed using the web based platform EnrichR (http://amp.pharm.mssm.edu/Enrichr) which permits interrogation of multiple databases (Kuleshov et al., 2016). Up and down regulated gene lists were evaluated for significant enrichment against the following gene set libraries: GO Biological Process, GO Cellular Component, GO Molecular Function (all from http://www.geneontology.org), the Wikipathways database (http://www.wikipathways.org/) and KEGG (http://www.kegg.jp/kegg) (all current as of August 2017). Enriched annotations/pathways were selected and ranked based upon the combined score which was calculated by the EnrichR platform following Z-score permutation background correction on the Fischer Exact Test p-value (Kuleshov et al., 2016). In order to identify cancer associated genes present in the differentially expressed genes the Oncosearch database of cancer genes was used (Lee et al., 2014).

**Results**

**Patient and sample details**

Details of the demographic and clinical features of the subjects enrolled in the study are presented in Table 1 and represent a typical range for patients referred to the clinic (Kordbacheh et al., 2016, Vu et al., 2015, Allen & Farah, 2015, Dost et al., 2014). The mean age at biopsy was 60.83 years (±9.33) and there were 7 males and 4 females. Of the 11 patients in the study one had biopsies taken on the same day from two different lesion sites (buccal mucosa and lateral tongue) so that the total number of sample sets was 12 (36 biopsies). Specimens were categorized based on pathology: dysplasia (OED; mild or moderate), oral lichenoid dysplasia (OLD; mild or moderate), oral lichen planus (OLP), or other (ulceration, chronic hyperplastic candidosis, hyperplasia/keratosis); as well as based on sample location: L, WL or AF. Tissues were classified as OLD if both lichenoid and dysplastic features were present histologically, while samples classified as OED did not contain a lichenoid component. For the purposes of this study, the grade of dysplasia was ignored for bioinformatic analysis. There were 7 samples from the lateral tongue, 4 from buccal mucosa, and one sample from the soft palate.

This article is protected by copyright. All rights reserved.
Sequencing Analytics and Unsupervised clustering

We performed RNAseq on a total of 36 samples with an average read depth of 50.7 million reads. Pre-processed reads were mapped against hg19 reference with an average efficiency of 92.6% (range 87.4%-96.6%) with detection of around 30,000 transcripts overall per library. In order to explore transcriptome relationships across the entire data set we performed hierarchical clustering of the samples using the 1000 most variably expressed genes as determined by median absolute deviation (Figure 2). Broadly across the dataset there were two major clusters, however, samples did not cluster separately according to the margin sampled but tended rather to cluster by patient. As can be seen from the sample annotations generally the samples from subjects with lichenoid features (OLD and OLP) clustered together in the first major cluster (Figure 2) while other samples were found in a second major cluster. The exceptions to this were biopsies 8L and 8WL (OED) which were found in the ‘lichenoid’ cluster whereas the autofluorescence biopsy from lesion 11 was in the second major cluster.

Differential Expression Analysis

In order to investigate the molecular dysregulation across the margins, differential expression analysis was performed initially using voom transformed data and limma. Multiple comparisons were made between the different margins (L vs WL, L vs AF, WL vs AF) for the entire dataset combining all diagnostic groups and separately for each diagnostic group provided that at least 3 samples per group were present. There were no DE genes for any of the comparisons tested when using limma which met the cut-off of padj<0.1. We then analysed the dataset using DEseq2 with pairwise comparisons which is suitable for studies with small sample sizes (Seyednasrollah et al, 2015). Figure 3 illustrates a broad overview of this analysis across the entire dataset. It demonstrates the extent of molecular difference across the margins as well as the statistically significant DE genes. This clearly shows that there was greater molecular difference in the transcriptome between the L and AF margins both generally and in the numbers of significant DE genes compared with either L vs WL or WL vs AF (Figure 3, Table 2). Furthermore, DE analysis between WL and AF margins (Figure 3c) is supportive of molecular distinction between these margins. When individual diagnostic groups (e.g. OLD or OED) were analysed for differential expression only a few DE genes were identified across the margins. For example for the L vs AF comparison in the subset of OLD samples alone only four genes (AQP5, SPINK9, AZGP1, GOLGA7B) were differentially expressed (Padj <0.1). This was most likely due to the small sample numbers in each group.

This article is protected by copyright. All rights reserved.
Discriminant analysis was also applied to the dataset in order to further explore molecular differences between the margins. Supervised sPLS-DA was applied to the gene expression data and was able to discriminate DE genes between margin comparisons within the combined grouping of all samples (Table 2). This sPLS-DA analysis identified a larger number of discriminating genes than DEseq2 and were supportive in showing greater molecular difference between the L and AF margins than other comparisons (Table 2). We used venn analysis to examine genes commonly identified by sPLS-DA across the different comparisons (Figure 4) and broadly there were few common genes across the comparisons. Complete annotated gene lists for the individual comparisons can be found in Supplementary file 1 and these were subject to further gene ontology and pathway analysis as described below.

**Gene Ontology and Pathway Analysis**

In order to explore the biological implications of the DE genes identified from sPLS-DA analysis we used overrepresentation analysis of gene ontology and biological and disease associated pathways to identify those which were enriched in our datasets. The findings of this functional exploration of the molecular dysregulation across the margins was consistent with our other results with significant enrichment of biological pathways in the L vs AF comparison but no significantly enriched pathways identified in either the L vs WL or WL vs AF genesets. For the L vs AF comparison enriched pathways comprised those associated with inflammatory signalling (cytokine response) and cancer progression (Jak-STAT signalling) (Table 3). We next examined the cancer association of differentially expressed genes in the L vs AF comparison using the Oncosearch database of genes dysregulated in cancer (http://oncosearch.biopathway.org). Of the 89 genes we found that 27 were represented in the database as dysregulated in cancer and with a role in carcinogenesis. A number of these genes have reported association with head and neck SCC (HNSCC) or other SCC (Table 4) and are involved in biologically relevant pathways.

**Discussion**

In this study we performed RNA sequencing on 36 samples from 11 subjects in order to investigate the transcriptome profile of lesion margins as determined by conventional white light illumination compared to autofluorescence. We investigated a clinically typical distribution of lesion types across this subject group. Our results and analysis provides molecular evidence for the superiority of autofluorescence in the delineation of margins for the excision of OPMD. The data also provides insight into biological pathways broadly dysregulated across the margins. Furthermore, the results
attest to the power of genome wide transcriptome analysis in even small study sizes particularly when paired lesion/margin are used.

This study was necessarily limited in some respects compared for example with our similar study of OSCC margins (Farah et al., 2016) since the level of molecular abnormality is much greater in tumour tissue than in pre-malignant lesions. Furthermore, physically OPMD lesions are relatively small, particularly compared to larger OSCC resections, and therefore the margins are spatially close and may overlap. This is evidenced by the comparatively low number of significant differentially expressed genes found in the present study. For ethical reasons, since we were obtaining multiple samples from each subject at the time of diagnostic biopsy, it was not practicable to restrict the study to a single diagnostic grouping. Despite this diversity of lesions, both clinically and histopathologically, we were able to successfully show molecular discrimination across the three biopsy locations.

Unsupervised hierarchical clustering of the samples in this study showed a tendency of cases to cluster according to histology rather than margin site. This is consistent with a reduced level of gene dysregulation in OPMD compared with cancer (eg OSCC) where samples do tend to cluster according to margin site (Farah et al., 2016). The apparent clustering of samples (including the margins) from patients with lichenoid lesions (OLP and OLD) generally away from the other samples is not unexpected considering the inflammatory nature of these lesions. Further investigation of this in a larger study group is required to allow exploration of these differences at the level of specific molecules and biological pathways, and to provide overdue molecular clarification on the differences and similarities between OLP, OLD and OED.

Application of both DEseq2 and sPLS-DA was successful in identifying discriminating genes between the margins, independent of histology, and demonstrating that molecular differences were greatest for the L vs AF comparison. For DE genes between the centre of the lesion and the autofluorescence margin the results indicate that irrespective of the pathology there were dysregulated pathways identified (Table 3) which are associated with inflammatory and oncogenic signalling. Amongst these genes are those with a previously identified association with tumour progression specifically in HNSCC (Table 4). Notably these include the matrix metalloproteinase MMP1 (Ye et al., 2008, Kidacki et al., 2017) and the chemokine ligand CCL18 (Jiang et al., 2016) both molecules implicated in tumour invasion in oral cancer.
Although the greatest molecular distinction was found between the lesion and the autofluorescence margin we also established that there were significant molecular differences between the lesion and the white light margin. Furthermore, when comparing the two margins, WL and AF, there was molecular dysregulation irrespective of pathology. This is significant because it indicates that the AF marginal tissue is significantly different to the lesion itself and clinicians should give serious consideration to using autofluorescence determined margins instead of WL for excision of OPMD amenable to surgery. This approach would ensure more of the lesion is excised leaving less moleculely and histopathologically abnormal tissue thereby reducing the likelihood of recurrent disease, and facilitating surveillance.

Conclusions
Our results indicate that the molecular profile of OPMD changes with divergence away from the centre of the lesion, and that autofluorescence determined margins are superior to the white light margin in achieving a clear molecular margin when excising an OPMD. Our study also demonstrates the power of gene expression profiling of the transcriptome even when applied to small study sizes.

Acknowledgements
Funding for this study was received from the Queensland Government Smart Futures Co-Investment Fund and Cancer Australia awarded to CSF. KJ would also like to thank the Australian Dental Research Foundation for funding support. The authors thank Jennifer Wu from UQ Centre for Clinical Research for assistance with RNA extraction; and Roxane Legaie from QFAB for assistance with bioinformatics and biostatistical data analysis. The authors also thank Omar Kujan and Norman Firth for assistance with pathological diagnosis.

Conflict of Interest
The authors declare no personal or financial interests in LED Medical Diagnostics Inc. or VELscope™ which may bias this manuscript.

Author Contributions
CF conceived and initiated this study. CF coordinated the experiments, the clinical sample collection and provided clinical information. FK, NB and KJ performed sample curation, RNA sequencing and preliminary data analysis. FK and KJ prepared the initial draft of the manuscript with significant input.
from CF. SF performed the bioinformatic analysis, clinical correlation and prepared the final draft of the manuscript with significant input from CF. All authors read and approved the final manuscript.

References


Kidacki M, Lehman HL, Green MV, Warrick JI and Stairs DB (2017). p120-Catenin Downregulation and PIK3CA Mutations Cooperate to Induce Invasion through MMP1 in HNSCC. *Mol Cancer Res.*

This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


Figure Legends

**Figure 1.** Example of an OPMD from patient RL_0376 on the right lateral tongue viewed under white light (a) and autofluorescence (b). (c) shows the regions where three punch biopsies were taken: at the centre of the lesion (L), white margin (WL), and autofluorescence margin (AF). The red outline in (c) shows the extension of the excision corresponding to that seen in (d). (d) shows the excision margin for this lesion based on fluorescence and the removal of the 3 punch biopsy samples and removal of the anterior portion of the lesion.

**Figure 2.** Hierarchical clustering of all samples. Samples are annotated according to diagnosis and relative expression of selected genes is shown by truncated heatmap.
**Figure 3.** Molecular diversity is greater across autofluorescence margins than white light margins. MA plots of differential expression analysis showing log fold change between margins relative to normalised mean expression level. (a) L vs AF for all samples, (b) L vs WL for all samples and (c) WL vs AF for all samples. Green dots are genes with greater than 1.5 fold expression difference (log2 = 0.58). Red dots are significant differentially expression genes, as determined by DESeq analysis (FDR < 0.1).

**Figure 4.** Venn analysis of discriminatory genes identified as differentially expressed by sPLS-DA. Illustrated are the number of genes unique and common between the different comparisons.
<table>
<thead>
<tr>
<th>Biopsy ID</th>
<th>Patient ID</th>
<th>Sex</th>
<th>Age at Biopsy (years)</th>
<th>Anatomical Site</th>
<th>Clinical Presentation</th>
<th>Histopathological Diagnosis</th>
<th>Definitive Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RL_0126</td>
<td>M</td>
<td>57</td>
<td>Tongue (lateral)</td>
<td>Erosive OLP</td>
<td>Moderate OLD</td>
<td>Moderate OLD</td>
</tr>
<tr>
<td>2*</td>
<td>RL_0212</td>
<td>M</td>
<td>71</td>
<td>Buccal mucosa</td>
<td>Multiple Non-Homogenous Leukoplakia</td>
<td>Mild OLD</td>
<td>Mild OLD</td>
</tr>
<tr>
<td>3*</td>
<td>RL_0212</td>
<td>M</td>
<td>71</td>
<td>Tongue (lateral)</td>
<td>Multiple Non-Homogenous Leukoplakia</td>
<td>Mild OLD</td>
<td>Mild OLD</td>
</tr>
<tr>
<td>4</td>
<td>RL_0280</td>
<td>F</td>
<td>76</td>
<td>Tongue (lateral)</td>
<td>Chronic Ulcer</td>
<td>Ulcer with keratosis</td>
<td>Chronic ulcer with keratosis</td>
</tr>
<tr>
<td>5</td>
<td>RL_0376</td>
<td>F</td>
<td>65</td>
<td>Tongue (lateral)</td>
<td>Plaque type OLP</td>
<td>OLP</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>RL_0406</td>
<td>M</td>
<td>46</td>
<td>Tongue (lateral)</td>
<td>Non-Homogeneous Leukoplakia</td>
<td>Mild OED</td>
<td>Mild OED</td>
</tr>
<tr>
<td>7</td>
<td>RL_0408</td>
<td>M</td>
<td>62</td>
<td>Buccal mucosa</td>
<td>Non-Homogeneous Leukoplakia</td>
<td>Hyperkeratosis/Hyperplasia</td>
<td>Hyperkeratosis/Hyperplasia</td>
</tr>
<tr>
<td>8</td>
<td>RL_0413</td>
<td>F</td>
<td>52</td>
<td>Buccal mucosa</td>
<td>Non-Homogeneous Leukoplakia</td>
<td>Mild OED</td>
<td>Mild OED</td>
</tr>
<tr>
<td>9</td>
<td>RL_0423</td>
<td>M</td>
<td>48</td>
<td>Soft palate</td>
<td>Homogeneous Leukoplakia</td>
<td>Mild OED</td>
<td>Mild OED</td>
</tr>
<tr>
<td>10</td>
<td>RL_0424</td>
<td>M</td>
<td>57</td>
<td>Buccal mucosa</td>
<td>Non-Homogeneous Leukoplakia</td>
<td>Chronic Hyperplastic Candidosis</td>
<td>Chronic Hyperplastic Candidosis</td>
</tr>
<tr>
<td>11</td>
<td>RL_0436</td>
<td>M</td>
<td>64</td>
<td>Tongue (lateral)</td>
<td>Homogeneous Leukoplakia</td>
<td>Mild OED</td>
<td>Mild OED</td>
</tr>
<tr>
<td>12</td>
<td>RL_0444</td>
<td>F</td>
<td>61</td>
<td>Tongue (lateral)</td>
<td>Reactive Keratosis</td>
<td>Hyperplasia/Keratosis</td>
<td>Hyperplasia/Keratosis</td>
</tr>
</tbody>
</table>

*Biopsies from same patient at different sites; OLP: Oral Lichen Planus; OED: Oral Epithelial Dysplasia; OLD: Oral Lichenoid Dysplasia
Table 2 Comparison of differentially expressed genes between margins

<table>
<thead>
<tr>
<th>DE Analysis</th>
<th>L vs AF</th>
<th>L vs WL</th>
<th>WL vs AF</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEseq2</td>
<td>23</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>sPLS-DA</td>
<td>89</td>
<td>57</td>
<td>57</td>
</tr>
</tbody>
</table>

L=Lesion; WL=White Light; AF= Autofluorescence

Table 3 Top 5 Enriched Biological Pathways in L vs AF

<table>
<thead>
<tr>
<th>Pathway Name</th>
<th>Database ID</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytokines and Inflammatory Response</td>
<td>WP530^1</td>
<td>10.44</td>
</tr>
<tr>
<td>Jak-STAT signaling pathway</td>
<td>hsa04630^2</td>
<td>9.93</td>
</tr>
<tr>
<td>PPAR signaling pathway</td>
<td>hsa03320^2</td>
<td>9.72</td>
</tr>
<tr>
<td>EPO Receptor Signaling</td>
<td>WP581^1</td>
<td>9.70</td>
</tr>
<tr>
<td>Cytokine-cytokine receptor interaction</td>
<td>hsa04060^2</td>
<td>9.18</td>
</tr>
</tbody>
</table>

^1Wikipathways or ^2Kegg pathway databases ranked by EnrichR score.
### Table 4 Selected cancer associated differentially expressed genes identified by sPLS-DA in L vs AF

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Description</th>
<th>Cancer association</th>
<th>Reference</th>
<th>Kegg Pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMP1</td>
<td>matrix metalloproteinase 1</td>
<td>head and neck squamous cell carcinoma</td>
<td>(Franchi et al., 2002)</td>
<td>PPAR signaling pathway</td>
</tr>
<tr>
<td>CSF2</td>
<td>colony stimulating factor 2</td>
<td>squamous cell carcinoma</td>
<td>(Colasante et al., 1995)</td>
<td>Cytokine-cytokine receptor interaction</td>
</tr>
<tr>
<td>NTSR1</td>
<td>neurotensin receptor 1</td>
<td>head and neck squamous cell carcinoma</td>
<td>(Alshouk et al., 2011)</td>
<td>Calcium signaling pathway</td>
</tr>
<tr>
<td>CDKN3</td>
<td>cyclin dependent kinase inhibitor 3</td>
<td>head and neck squamous cell carcinoma</td>
<td>(Wang et al., 2015)</td>
<td>-</td>
</tr>
<tr>
<td>CCL18</td>
<td>C-C motif chemokine ligand 18</td>
<td>squamous cell carcinoma</td>
<td>(Jiang et al., 2016)</td>
<td>Chemokine signaling pathway,</td>
</tr>
<tr>
<td>CRABP1</td>
<td>cellular retinoic acid binding protein 1</td>
<td>esophageal squamous cell carcinoma</td>
<td>(Tanaka et al., 2007)</td>
<td>-</td>
</tr>
<tr>
<td>DMBT1</td>
<td>deleted in malignant brain tumors 1</td>
<td>oral squamous cell carcinoma</td>
<td>(Imai et al., 2005)</td>
<td>Salivary secretion,</td>
</tr>
<tr>
<td>EPO</td>
<td>erythropoietin</td>
<td>squamous cell carcinoma</td>
<td>(Li et al., 2009)</td>
<td>Cytokine-cytokine receptor interaction</td>
</tr>
<tr>
<td>CEACAM4</td>
<td>carcinoembryonic antigen related cell adhesion molecule 4</td>
<td>colorectal carcinoma</td>
<td>(Koops et al., 1998)</td>
<td>-</td>
</tr>
</tbody>
</table>