

Journal section: Oral Medicine and Pathology  
Publication Types: Research

doi:10.4317/medoral.27658

## Descriptive transcriptomic profiling differentiates oral leukoplakia from proliferative verrucous leukoplakia and reveals distinct molecular signatures

Mario Pérez-Sayáns <sup>1,2,3</sup>, Fábio França Vieira e Silva <sup>1,2</sup>, Ceres Fernández-Rozadilla <sup>4</sup>, Ángel Carracedo <sup>4</sup>, Silvia Carlés-González <sup>4</sup>, Alejandro Ismael Lorenzo-Pouso <sup>1,2</sup>, Alba Pérez-Jardón <sup>1,2</sup>, Pilar Gándara-Vila <sup>1,2</sup>, Abel García-García <sup>1,2</sup>, José Manuel Suárez-Peñaranda <sup>1,2</sup>, Andrés Blanco-Carrión <sup>1,2</sup>, Cintia Micaela Chamorro-Petronacci <sup>1,2</sup>

<sup>1</sup> Oral Medicine, Oral Surgery and Implantology Unit (MedOralRes), Faculty of Medicine and Dentistry, University of Santiago de Compostela, San Francisco Street, s/n, 15782 Santiago de Compostela, Spain

<sup>2</sup> Health Research Institute of Santiago de Compostela (FIDIS), ORALRES Group, Santiago de Compostela University Clinical Hospital, University of Santiago de Compostela, Choupana Street, s/n, 15706 Santiago de Compostela, Spain

<sup>3</sup> Materials Institute of Santiago de Compostela (iMATUS), Mestre Mateo Avenue, 25, 15782 Santiago de Compostela, Spain

<sup>4</sup> Health Research Institute of Santiago de Compostela (FIDIS), Genomic Medicine Group, s/n, Choupana Street, 15706 Santiago de Compostela, Spain

Correspondence:  
Santiago de Compostela University Clinical Hospital  
University of Santiago de Compostela, Choupana Street  
s/n, 15706 Santiago de Compostela, Spain.  
fabio.frana@rai.usc.es

Received: 10/07/2025  
Accepted: 19/10/2025

Please cite this article in press as: Pérez-Sayáns M, Silva FFV, Fernández-Rozadilla C, Carracedo A, Carlés-González S, Lorenzo-Pouso AI, Pérez-Jardón A, Gándara-Vila P, García-García A, Suárez-Peñaranda JM, Blanco-Carrión A, Chamorro-Petronacci CM. Descriptive transcriptomic profiling differentiates oral leukoplakia from proliferative verrucous leukoplakia and reveals distinct molecular signatures. Med Oral Patol Oral Cir Bucal. 2025.  
doi:10.4317/medoral.27658

Article Number: 27658 <http://www.medicinaoral.com/>  
© Medicina Oral S. L. C.I.F. B 96689336 - pISSN 1698-4447 - eISSN: 1698-6946  
eMail: [medicina@medicinaoral.com](mailto:medicina@medicinaoral.com)  
**Indexed in:**  
Science Citation Index Expanded  
Journal Citation Reports  
Index Medicus, MEDLINE, PubMed  
Scopus, Embase and Emcare  
Indice Médico Español

### Abstract

**Background:** Oral leukoplakia and proliferative verrucous leukoplakia represent oral potentially malignant disorders. Oral leukoplakia typically presents as solitary lesions, while proliferative verrucous leukoplakia manifests as multifocal lesions with higher malignant potential. This study aimed to investigate the genetic heterogeneity between these disorders through differential gene expression, genetic variants, and microRNA profiling to identify potential biomarkers for diagnosis and prognosis.

**Material and Methods:** Biopsies and peripheral blood samples were obtained from 20 patients. Subsequently, RNA extraction, RNA-Seq libraries preparation, and bioinformatic analyses were conducted to ascertain differential gene expression, genetic variants, and microRNA expression.

**Results:** In mRNA analysis, overexpressed genes in proliferative verrucous leukoplakia are primarily associated with inflammation and immune regulation, while underexpressed genes relate to skin barrier maintenance. Pathway analysis reveals underexpressed genes related to impaired keratinization in proliferative verrucous leukoplakia and with keratin envelope formation in oral leukoplakia, while overexpressed genes are linked to synaptic processes and protein-protein interactions. Somatic mutation drivers in proliferative verrucous leukoplakia include variants in NRXN3, SRGAP2B, INIP, MYO18A, and ATF7IP genes. Regarding variant analysis, two variants in the Syndecan 3 (SDC3) gene identified in proliferative verrucous leukoplakia have demonstrated enormous value and indicate an important biomarker for a differential diagnosis and to predict prognosis. Proliferative verrucous leukoplakia shows in miRNA analysis MIR1246 and MIR767 overexpression, with MIR135B being the most underexpressed.

**Conclusions:** Our findings emphasize the intricate transcriptomic profiles in oral leukoplakia and proliferative verrucous leukoplakia development, laying the groundwork for future studies to enhance clinical management and patient outcomes in oral oncology. Syndecan 3 gene polymorphisms may represent a key point in proliferative verrucous leukoplakia differential diagnosis and prognostic prediction.

**Keywords:** Oral leukoplakia, proliferative verrucous leukoplakia, oral squamous cell carcinoma, transcriptomic profiles.

## Introduction

Oral leukoplakia (OL) and proliferative verrucous leukoplakia (PVL) are oral potentially malignant disorders (OPMD), representing a significant concern in the oral pathology field [1]. PVL corresponds to the OPMD with the greatest potential for malignancy (66-70%) compared to OL (7.9-11.7%). These lesions share remarkable similarities in clinical presentation and histopathology, highlighting challenges in differentiating between them and underscoring the importance of searching for novel biomarkers [1]. This study employed RNA-Seq to analyze the differential gene expression profiles and genetic variations in OL and PVL to elucidate the molecular mechanisms driving their origin and clinical behavior.

## Material and Methods

### *Patients*

Samples were obtained from 20 patients diagnosed clinically and histopathologically with OL (n=15) or PVL (n=5); From each patient, one fresh biopsy and one peripheral blood sample were collected. Patients were referred to the Oral Medicine Unit of the Faculty of Medicine and Dentistry at the University of Santiago de Compostela and had at least 24 months of follow-up. Incisional biopsies of 5×5mm from different lesions were taken using a scalpel and stored in Allprotect Tissue Reagent (Qiagen, Hilden, Germany) at -80°C until analysis. Simultaneously, 10 mL of peripheral venous blood was obtained from each patient according to the standard protocol.

In the PVL group, 20% of patients were male and 80% female, while in the OL group, 64.3% were male and 35.7% female. The mean age at diagnosis was 67.9 years in PVL (range 52.5-82.8) and 62.68 years in OL (range 26.5-85.2). All PVL cases were non-homogeneous, whereas in OL, 64.3% were homogeneous and 35.7% non-homogeneous. Lesions in PVL occurred on the tongue (40%), palate (20%), and gum (40%). In OL, the tongue was affected in 50%, buccal mucosa in 35.7%, palate in 7.1%, and gum in 7.1%. Dysplasia in PVL was absent in 60% and low-grade in 40%. In OL, dysplasia was absent in 50%, low-grade in 35.7%, high-grade in 7.1%, and inconclusive in 7.1%. Clinical evolution in PVL showed cure in 80% and stability in 20%. In OL, 28.6% were cured, 42.9% stable, 21.4% showed malignization, and 7.1% had recurrence. All PVL patients

were alive at last follow-up, while in OL, 92.9% were alive and 7.1% had died.

### *Ethical Aspects*

Before collecting samples, each patient received the study's information sheet and signed an informed consent form, agreeing to participate in the study. They were treated in accordance with the Declaration of Helsinki and its later amendments. This study was approved by the clinical research and drug ethics committee of Galicia (Ref. 2018/503).

### *RNA Extraction*

The RNA extraction was performed using the mirVana miRNA Isolation Kit (Ambion, Austin, TX, USA), and RNA was treated with DNase I, using the DNA-free DNA Removal Kit (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA), following the manufacturer's protocols. The RNA quality was analysed by the RNA integrity number (RIN) values with an Agilent 2100 Bioanalyzer system.

### *RNA-Sequencing (RNA-Seq)*

TruSeq Stranded Total RNA libraries were constructed by MacroGen Inc. (MacroGen Europe, Amsterdam, The Netherlands) to obtain Illumina paired-end reads (151 bp read length).

RNA-Seq quality control was performed using Fastqc for each FASTQ file and Multiqc to integrate sample results (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Per Sequence Quality Scores, Sequence Duplication Levels, and Overrepresented Sequences were considered. Reads were aligned to the human genome (GRCh38) using STAR [2], and gene counts (GC) were estimated using featureCounts [3]. Differential expression gene (DGE) analyses were conducted using R Software (R Core Team 2021, Vienna, Austria). Count values were imported and processed using edgeR [4]. Expression values were normalized using the trimmed mean of M values (TMM) method, and lowly-expressed genes were filtered out.

### *Genetic Variant Evaluation*

Mutations were described by mRNA variant, associated gene, cDNA position, and nucleotide mutation. Scale-Invariant Feature Transform (SIFT) and Polymorphism Phenotyping (PolyPhen) were performed as bioinformatic tools for functional impact predictive assessment.

### *DGE Analysis*

DGE was identified using linear models (Limma-

Voom [5], and P values were adjusted for multiple comparisons by applying the Benjamini-Hochberg correction. Heatmaps and volcano plots were generated using the heatmap3 [6], EnhancedVolcano (<https://github.com/kevinblighe/EnhancedVolcano>), and Glimma [7] packages.

#### Pathway Analyses

Gene Set Enrichment Analysis (GSEA) techniques, specifically Gene Ontology Biological Process (GOBP) and REACTOME (molecular pathways) were used to indicate major functional enrichments for different over- and underexpressed genes (OE and UE).

#### Gene Functional Annotation

Gene functional annotation was performed using clusterProfiler [8]. Receiver Operating Characteristic (ROC) analyses were conducted using verification (<https://cran.r-project.org/web/packages/verification/index.html>).

#### DNA/Somatic Variant Analyses

Variant calling was performed using Samtools, and Bcftools, as previously reported [9]. Variants in positions with coverage below 10 for either diseased or matched control samples were discarded. Disease-associated variants were established by comparing the variants of each diseased sample with its matched healthy control. Odds ratios (OR) were evaluated using Fisher's exact tests, and P values were adjusted for multiple comparisons by applying the Benjamini-Hochberg correction. The Ensemble Variant Effect Predictor (VEP) was used for variant annotation [10].

## Results

The quality control and normalization are demonstrated in Supplementary Figure 1-4 ([http://www.medicina.oral.com/carpeta/suppl1\\_27658](http://www.medicina.oral.com/carpeta/suppl1_27658)).

#### Differential Expression mRNA

##### Differential Gene Expression (DGE)

Principal Component Analysis (PCA) reflects a clear separation between biological groups (Figure 1A). Differential expression of the PVL group versus the OL group is calculated, with 138 OE genes and 93 UE genes. Genes showing differential expression are represented in red (OE) and blue (UE) in a volcano plot (Figure 1B), demonstrating clear clustering among different samples and in a heatmap (Figure 1C). For OE genes, the main altered biological pathway and function would be Inflammation and Immunity regulation, represented by PYDC1, TRIM49, and KLK1 genes. For UE genes, the skin barrier formation and maintenance would be affected, with SPRR2D, SPRR1A, SPRR2A, SPRR4, and SPRR3 genes, belonging to the proline-rich proteins family, being UE. No statistically significant association has been found with dysplasia degree. Regarding clinical evolution, two genes with differential expression have been found. Comparing patients who progressed to malignancy with those who remained stable, the C1orf226 gene ( $\log_{2}FC=2.473$ ,  $p=6.39E-06$ ) was

significantly OE, and the CYB5A gene ( $\log_{2}FC=-3.209$ ,  $p=8.732E-06$ ) was UE.

#### Pathway analyses

GSEA techniques indicate the main functional enrichments for different genes. For UE genes, the main biological processes involved were keratinization, peptide cross-linking, keratinocyte differentiation, epidermal cell differentiation, and epidermis development, suggesting the impaired keratinization role in PVL lesions. Regarding molecular pathways, the cornified envelope and keratinization formation are confirmed as the primary pathways involved in UE genes. For OE genes, synaptic processes, and protein-protein interaction pathways are related, with SUMOylation being one of the most prevalent (ratio 0.0625) (Supplementary Figure 5 - [http://www.medicina.oral.com/carpeta/suppl1\\_27658](http://www.medicina.oral.com/carpeta/suppl1_27658)).

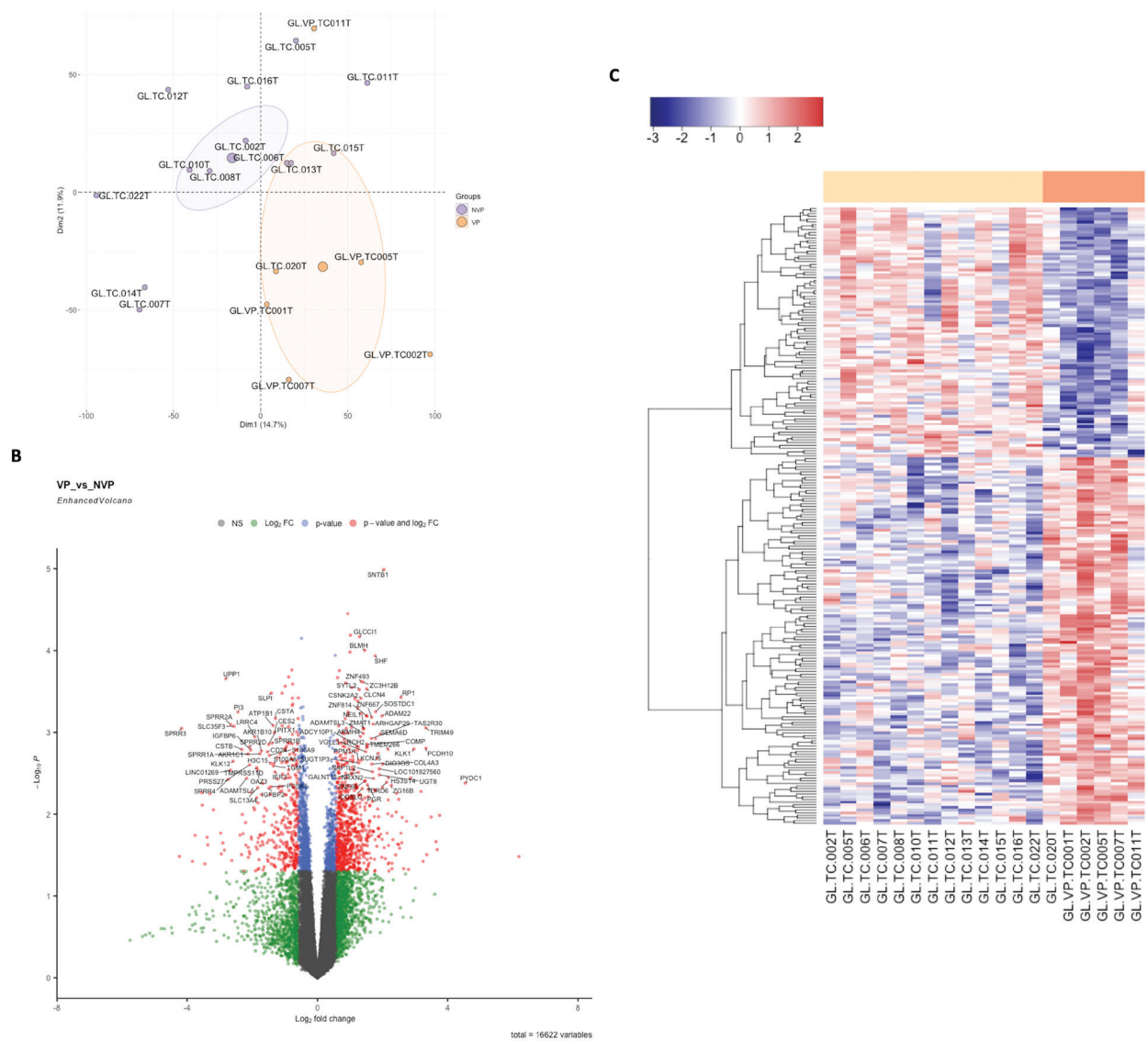
#### Somatic mutation drivers identification for PVL

In parallel to gene expression analyses, an RNA-based variant was performed to elucidate somatic alterations that are driving OL and PVL formation, as well as those that can differentiate both lesion types. Regarding PVL, our findings demonstrate some variants with frequent occurrence, considering those with a frequency higher than 0.6. The results demonstrate that the variants rs31437 in the NRXN3 gene (Frequency: 0.66, OR for NVP=0;  $p=0.0003$ ); rs1478770644 in the SRGAP2B gene (Frequency: 0.50); rs557699893 in the INIP gene (Frequency: 0.50); rs8076604 in the MYO18A gene (Frequency: 0.50) and rs2231909 in the ATF7IP gene (Frequency: 0.50) are exclusive to PVL. ROC analysis identified three highly predictive genes: ENSG00000172164 (SNTB1), AUC=1, adjusted  $p$ -value=0.002137697, ENSG00000182700 (IGIP), AUC=1, adjusted  $p$ -value=0.002137697, and ENSG00000116396 (KCNC4), AUC=1, adjusted  $p$ -value=0.002137697. The ROC curve analysis comparing malignancy and non-malignancy groups has no significant results for the studied genes.

#### Different Variant Analysis (DVA)

The DVA is carried out similarly to DGE, revealing differentially enriched variants between the PVL and OL groups. Hierarchical clustering reveals distinct groupings based on variants (Figure 2A). We observe two clusters, one on the left, formed by samples GLVP.02/GLVP.05/GL.015/GL.011/GL.020, with similar variants and consisting of OL that did not malign from diagnosis throughout the follow-up. On the right would be another cluster with all the other OLs, including the 3 that evolved into a malignant tumor and the one that presented a previous in situ carcinoma before the study (GLVP.01). Within this right cluster, there is a small sub-cluster that groups two PVLs, GLVP.01 and GLVP.07, with similar height to the other PVLs.

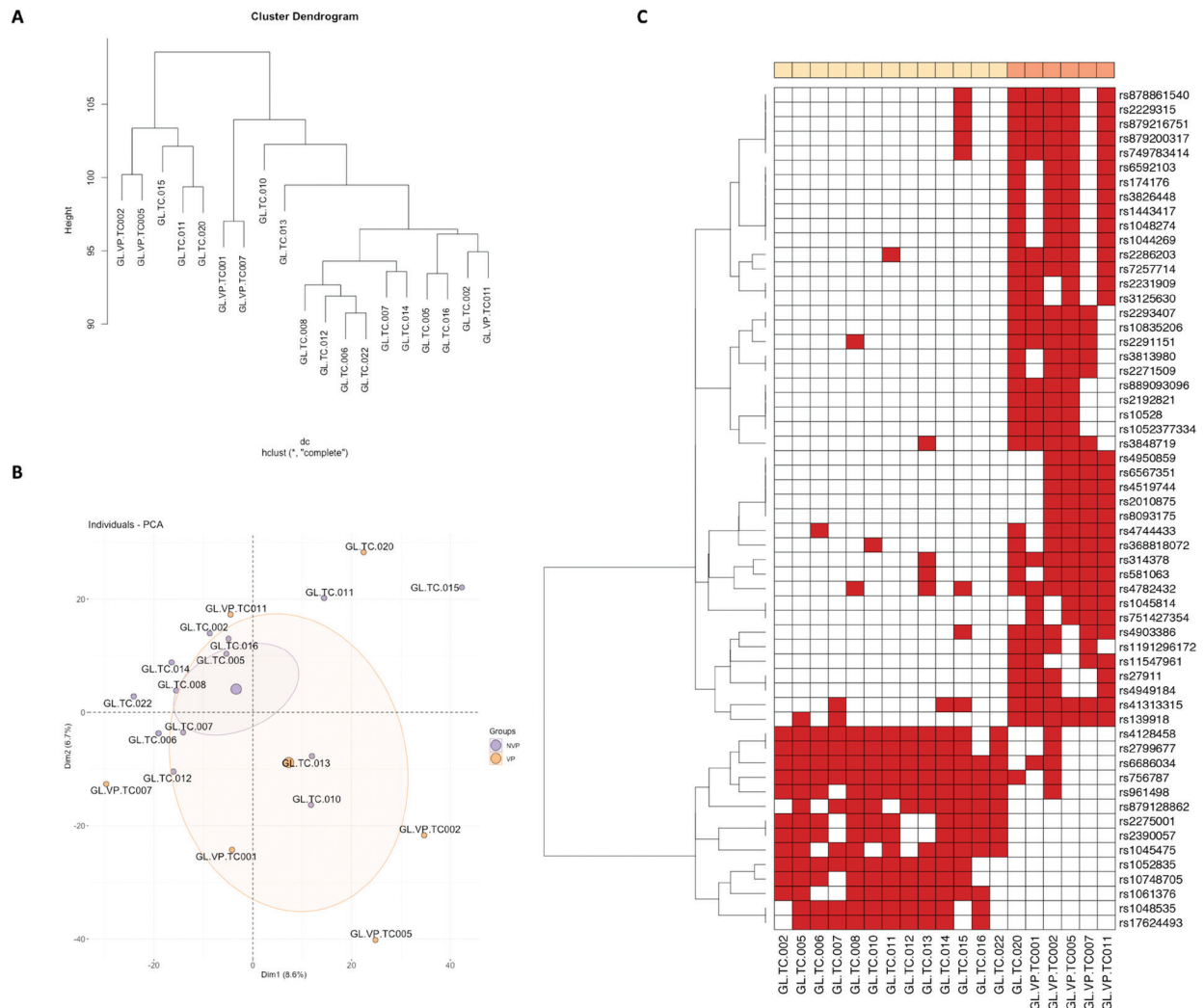
In the PCA for DVA, there is a greater sample dispersion in the PVL group, while a very high percentage of OLs



**Figure 1:** Differential Gene Expression (DGE) of mRNA in Oral Leukoplakia (OL) Versus Proliferative Verrucous Leukoplakia (PVL). (A) Principal Component Analysis (PCA) reflects a clear separation between biological groups, although intra-group dispersion is notable; (B) Volcano Plot; (C) and heatmap showing DGE represented in red for overexpressed genes and in blue for underexpressed genes, demonstrating clear clustering among different samples.

are grouped in a restricted area. Note the distance to the subcluster of samples GL.015, GL.011, and GL.020, which had a favorable clinical course (Figure 2B). In the DVA heatmap, there is a large number of common or nearly specific genetic PVL variants, among the most significant are rs314378, rs139918, rs10835206, rs2293407, rs7257714, while for the OL group, there are smaller variants number but grouping almost all non-verrucous leukoplakias. GL.020 presents a variant profile most similar to PVLs, which can be clinically correlated as the patient has 3 lesions in different locations (Figure 2C). Regarding variant annotation among the variants showing differential enrichment, In the PVL group, a missense variant in SDC3 (rs4949184, 1:30874552) was identified. The A allele resulted in an aspartic acid to tyrosine substitution (D/Y) with moderate predicted

impact. Two transcripts were affected (protein positions 245 and 303), both classified as deleterious by SIFT (score 0) and predicted as probably damaging (PolyPhen 0.962) or possibly damaging (PolyPhen 0.888). In the OL group, a missense variant in FBLN2 (rs1061376, 3:13637835) led to an aspartic acid to glutamic acid substitution (D/E). This variant was detected in three transcript positions (protein positions 1157 and 1204), consistently predicted as deleterious by SIFT (0.01-0.02) and probably damaging by PolyPhen (0.94-0.998). Additionally, a missense variant in LAD1 (rs2799677, 1:201386815) was observed, corresponding to a lysine to asparagine substitution (K/N). It was found in three transcript positions (protein positions 182 and 196) and classified as deleterious by SIFT (score 0, one low-confidence) and possibly damaging by PolyPhen (0.582-0.736).



**Figure 2:** Different Variant Analysis (DVA). (A) Hierarchical clustering (dendrogram) reveals distinct groupings based on variants. Two main clusters demonstrate, one on the left, formed by samples with similar variants and consisting of OL that did not malign from diagnosis throughout the follow-up, and another on the right with all the other OLs including the 3 that evolved into a malignant tumor, and the one that presented a previous in situ carcinoma before the study, with a subcluster that groups two PVLs, GLVP.01 and GLVP.07, with a similar height to the other PVLs. (B) Principal Component Analysis (PCA) for DVA demonstrates a greater dispersion of samples in the PVL group, while a very high percent-age of OLs are grouped in the same area. (C) Heatmap for DVA demonstrates many common or nearly specific genetic variants of PVL compared to OL number.

### miRNA Analysis

#### Differential miRNA expression

PCA analysis (Figure 3A) reflects high intra-group dispersion, especially in the OL group. Genes showing differential expression are indicated in red (OE) and blue (UE). These genes are represented in a volcano plot heatmap (Figure 3B) and a heatmap (Figure 3C). It is noteworthy MIR1246 and MIR767 OE. No UE miRNA with an FC <2 is described, with MIR135B being the last on the list. Regarding follow-up, some OE miRNAs are described in cases that undergo malignization: MIR767 (FC=3.56;  $p=0.0358$ ), MIR208B (FC=3.29;  $p=0.033$ ), MIR208A (FC=2.54;  $p=0.0042$ ), MIR518C (FC=2.07;  $p=0.0253$ ), MIR4798 (FC=2.06;  $p=0.0150$ ). One UE miRNA with FC <2 is described, MIR365B (FC=-2.21;  $p=0.0084$ ).

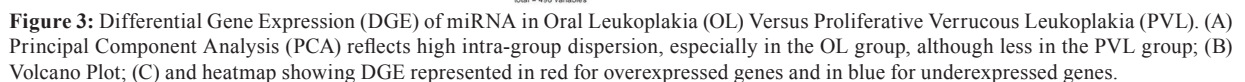
### Functional Annotation

Functional enrichments for different UE and OE miRNAs are indicated in Supplementary Figure 6 ([http://www.medicina.oral.com/carpeta/suppl1\\_27658](http://www.medicina.oral.com/carpeta/suppl1_27658)).

For UE genes, the main biological process involved was the translation regulation, mainly blocking protein production. For OE genes, muscle apoptosis processes and negative vascular endothelial growth factor regulation are primarily related.

### Discussion

Our findings on mRNA DGE reveal that OE genes in PVL were mainly related to inflammation and immune regulation, in contrast, UE genes were associated with forming and maintaining the skin barrier. Llorens *et al.* also found significant upregulation of genes associ-



with synaptic density and maturation, impacting neurological development [16]; rs557699893 in INIP, crucial for genome stability and disease susceptibility [17]; rs8076604 in MYO18A, involved in maintaining muscle cell stability [18]; rs2231909 in ATF7IP, related to immune functions and cancer [19].

Through ROC analysis in identifying somatic mutation drivers for PVL, three genes with high predictive capacity were identified: SNTB1, IGIP, and KCNC4. Zhang *et al.* demonstrated that SNTB1 positively regulates the expression of YAP1, affecting tumor growth and metastasis in colorectal cancer [20]. Austin *et al.* dissected IGIP functions and described that this gene regulates IgA by acting as a switch or differentiation factor [21]. In head and neck squamous cell carcinoma (HNSCC), it is reported that there is an increase of KCNC4 mRNA levels in tumors compared to the corresponding normal epithelia [22]. In our analysis related to DVA analysis, two variants in the SDC3 gene identified in patients with PVL are of particular interest. Syndecans are a family of four trans-membrane heparan sulfate proteoglycans, including

syndecan -1, -2, -3, and -4. SDC3 is mainly expressed in brain and nervous tissue, with a key role in development, cell adhesion, and migration [23]. Although some Syndecan family members are already widely debated in the literature, SDC3 is a gene that needs more information, with no association with OL or PVL previously described. In contrast, Syndecan 1 has been associated with oral dysplastic lesions [24], OSCC [25], and oral verrucous carcinoma [26].

Hudák *et al.* identified SDC3 expression of monocytes as a novel biomarker for early Alzheimer's disease detection [27]. Brito *et al.* described the power of SDC3 in enhancing WNT signaling in osteoblasts, making SDC3 an attractive target for novel bone anabolic drug development [28]. Kehoe *et al.* demonstrated the role of SDC3 in inflammation, acting as a pro-inflammatory agent in the endothelial presentation of chemokines and the leukocytes recruitment and in cartilage damage in a rheumatoid arthritis model, unlike in the skin and cremaster, where its action is anti-inflammatory [29].

In cancer research, Prieto-Hernández *et al.* demonstrated that SDC3 expression is upregulated in several types of cancer, especially in hypoxic solid tumors [30]. Hillemeyer *et al.* suggest that SDC3 upregulation promotes ovarian cancer pathogenesis by modulating stemness-associated pathways [31]. Yamada *et al.* showed that the antitumor miR-144-5p/oncogenic SDC3 axis was deeply involved in renal cancer pathogenesis [32].

In the OL group, we found variants in two genes, FBLN2 and LAD1. FBLN2 is involved in carcinogenesis through interaction with other extracellular matrices (ECM) proteins, including integrins and syndecans, with pathways related to the elastic fibers formation and ECM organization, being related to diseases such as Osteochondritis Dissecans and Cutis Laxa [33]. LAD1 is a gene classified as collagen-anchored filaments in the basement membrane of mammalian epidermal cells that contributes to the stability of epithelial layers associated with the underlying mesenchyme, being correlated with diseases such as Epidermolysis Bullosa Acquisita and Iga Linear Disease [34]. Farah *et al.* explored the biological implications of the DGE between OL with and without dysplasia, describing that the negatively regulated genes were significantly enriched considering the biological processes related to collagen fibrils organization and ECM [13].

In our miRNA analysis, the PVL group DGE compared to the non-PVL group highlighted that MIR1246 and MIR767 are OE, with MIR135B being the most UE. miR-1246 has been reported to regulate several pathways, and miR-767 is highly expressed in senescent vascular endothelial cells and their exosomes [35]. Choi *et al.* demonstrated that blocking miR135b improves the microenvironment and basal cell proliferation in psoriasis, demonstrating a central role for miR135b in epidermal keratinocytes [36].

When analyzing the malignancy process, our results demonstrate that MI767, MIR208B, MIR208A, MIR518C, and MIR4798 are OE, while MIR365B is UE. In contrast, Zhu *et al.* identified a series of other 8 miRNAs related to OL to OSCC evolution [37]. Chen *et al.* also demonstrated different results with miR-129-5p, miR-296-5p, and miR-450b-5p commonly regulated between OL and OL-OSCC [38].

Our research involves identifying RNA variants from mucosal and blood samples instead of DNA. Consequently, the primary limitation is the potential presence of variants in the mucosa that may not be found in other tissues due to the differing expression profiles between tissues rather than mutations necessarily associated with the pathology. Ideally, variants should be identified from DNA or, failing that, use a germline control from the same tissue if RNA (healthy mucosa rather than blood) is desired. However, in OPMD patients, the oral mucosa behaves like a precancerous field. Therefore, the use of healthy mucosa as a control is also not ideal. Despite some limitations, our study offers important multi-omic insights into distinguishing OL from PVL beyond clinical characteristics.

## Conclusions

Differential mRNA expression analysis revealed distinct OE and UE genes between PVL and OL, with key pathways involving inflammation, immune response, and epithelial barrier maintenance. Enrichment analyses emphasized keratinization-related processes in PVL and cornified envelope formation in OL. Somatic mutation analysis uncovered PVL-specific variants, including in the ATF7IP and MYO18A genes, suggesting their relevance to pathogenesis and diagnostic potential. Notably, SNTB1, IGIP, and KCNC4 showed high predictive value for PVL. Variant analysis also identified two significant SDC3 variants in PVL, supporting their role as potential diagnostic biomarkers. miRNA profiling revealed differentially expressed miRNAs, some associated with malignant transformation. Functional annotation highlighted biological processes such as muscle apoptosis and translational repression. Altogether, these findings offer molecular insights into PVL and OL, laying the groundwork for future diagnostic and therapeutic strategies.

## Acknowledgement

Declared none.

## Institutional Review Board Statement

Before collecting samples, each patient received the study's information sheet and signed an informed consent form, agreeing to participate. They were treated under the Declaration of Helsinki and subsequent modifications. This study was approved by the clinical research and drug ethics committee of Galicia (Ref. 2018/503).

## Data availability statement

Data is available upon request from the authors.

### Author Contributions

MPS and CMCP were responsible for designing the study; CFR and AC were responsible for carrying out the analyses; MPS, CFR, ABC, AILP, and AC were re-sponsible for analyzing data; MPS, CFR, CMCP, and FFVS were responsible for interpreting the results, writing the report, and creating the figures, graphs, and tables; JMSP, AGG, APJ, and PGV were responsible for providing feedback for the report.

### Funding

This work was supported by Asociación Española Contra el Cáncer [Grant number: GC16173720CARR].

### Conflicts of Interest

The authors declare no conflicts of interest.

### References

1. Lafuente Ibáñez de Mendoza I, Lorenzo Pouso AI, Aguirre Urizar JM, Barba Montero C, Blanco Carrión A, Gándara Vila P, *et al.* Malignant development of proliferative verrucous/multifocal leukoplakia: a critical systematic review, meta-analysis and proposal of diagnostic criteria. *J Oral Pathol Med.* 2022;51(1):30-8.
2. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21.
3. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30(7):923-30.
4. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139-40.
5. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 2014;15(2).
6. Zhao S, Guo Y, Sheng Q, Shyr Y. Heatmap3: an improved heatmap package with more powerful and convenient features. *BMC Bioinformatics.* 2014;15(Suppl 10):P16-P.
7. Su S, Law CW, Ah-Cann C, Asselin-Labat ML, Blewitt ME, Ritchie ME. Glimma: interactive graphics for gene expression analysis. *Bioinformatics.* 2017;33(13):2050-2.
8. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OmicS.* 2012;16(5):284-7.
9. Tang X, Baheti S, Shameer K, Thompson KJ, Wills Q, Niu N, *et al.* The eSNV-detect: a computational system to identify expressed single nucleotide variants from transcriptome sequencing data. *Nucleic Acids Res.* 2014;42(22).
10. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17(1).
11. Llorens C, Soriano B, Trilla-Fuertes L, Bagan L, Ramos-Ruiz R, Gamez-Pozo A, *et al.* Immune expression profile identification in a group of proliferative verrucous leukoplakia patients: a pre-cancer niche for oral squamous cell carcinoma development. *Clin Oral Invest.* 2021;25(5):2645-57.
12. García-López R, Moya A, Bagan JV, Pérez-Brocal V. Retrospective case-control study of viral pathogen screening in proliferative verrucous leukoplakia lesions. *Clin Otolaryngol.* 2014;39(5):272-80.
13. Farah CS, Fox SA. Dysplastic oral leukoplakia is molecularly distinct from leukoplakia without dysplasia. *Oral Dis.* 2019;25(7):1715-23.
14. Kyrodimou M, Andreadis D, Drougou A, Amanatiadou EP, Angelis L, Barbatis C, *et al.* Desmoglein-3/ $\gamma$ -catenin and E-cadherin/ $\beta$ -catenin differential expression in oral leukoplakia and squamous cell carcinoma. *Clin Oral Invest.* 2014;18(1):199-210.
15. Kathuria A, Lopez-Lengowski K, Watmuff B, Karmacharya R. Morphological and transcriptomic analyses of stem cell-derived cortical neurons reveal mechanisms underlying synaptic dysfunction in schizophrenia. *Genome Med.* 2023;15(1).
16. Schmidt ERE, Kupferman JV, Stackmann M, Polleux F. The human-specific paralogs SRGAP2B and SRGAP2C differentially modulate SRGAP2A-dependent synaptic development. *Sci Rep.* 2019;9(1).
17. Huang J, Gong Z, Ghosal G, Chen J. SOSS complexes participate in the maintenance of genomic stability. *Mol Cell.* 2009;35(3):384-93.
18. Cao JM, Cheng XN, Li SQ, Heller S, Xu ZG, Shi DL. Identification of novel MYO18A interaction partners required for myoblast adhesion and muscle integrity. *Sci Rep.* 2016;6.
19. Hu H, Khodadadi-Jamayran A, Dolgalev I, Cho H, Badri S, Chiriboga LA, *et al.* Targeting the Atf7ip-Setdb1 Complex Augments Antitumor Immunity by Boosting Tumor Immunogenicity. *Cancer Immunol Res.* 2021;9(11):1298-315.
20. Zhang H, Li Z, Jiang J, Lei Y, Xie J, Liu Y, *et al.* SNTB1 regulates colorectal cancer cell proliferation and metastasis through YAP1 and the WNT/ $\beta$ -catenin pathway. *Cell Cycle.* 2023;22(17):1865-83.
21. Austin AS, Haas KM, Naugler SM, Bajer AA, Garcia-Tapia D, Estes DM. Identification and Characterization of a Novel Regulatory Factor: IgA-Inducing Protein. *J Immunol.* 2003;171(3):1336-42.
22. Menéndez ST, Rodrigo JP, Allonca E, García-Carracedo D, Álvarez-Alija G, Casado-Zapico S, *et al.* Expression and clinical significance of the Kv3.4 potassium channel subunit in the development and progression of head and neck squamous cell carcinomas. *J Pathol.* 2010;221(4):402-10.
23. Oto J, Le QK, Schäfer SD, Kiesel L, Mari-Alexandre J, Gilabert-Estellés J, *et al.* Role of Syndecans in Ovarian Cancer: New Diagnostic and Prognostic Biomarkers and Potential Therapeutic Targets. *Cancers.* 2023;15(12).
24. Lakkam B, Majage B, Astekar M, Gugwad RS, Giri G, Ramasahayam S. Immunohistochemical expression of syndecan-1 in oral dysplastic epithelium. *J Cancer Res Ther.* 2014;10(1):103-6.
25. Woźniak M, Nahajowski M, Hnitecka S, Rutkowska M, Nowak M, Mitelsztet P, *et al.* Expression of syndecan-1 in oral cavity squamous cell carcinoma. *J Histotechnol.* 2021;44(1):46-51.
26. Elhassan AKB, Suleiman AM, El Dawi NIA, Mohamed SB. Detection of micro-invasion in Sudanese oral verrucous carcinoma samples using syndecan-1 stain. *Biomarkers Cancer.* 2019;11.
27. Hudák A, Letoha A, Vizler C, Letoha T. Syndecan-3 as a novel biomarker in Alzheimer's disease. *Int J Mol Sci.* 2022;23(6).
28. Johnson de Sousa Brito FM, Butcher A, Pisconti A, Poulet B, Prior A, Charlesworth G, *et al.* Syndecan-3 enhances anabolic bone formation through WNT signaling. *FASEB J.* 2021;35(4).
29. Kehoe O, Kalia N, King S, Eustace A, Boyes C, Reizes O, *et al.* Syndecan-3 is selectively pro-inflammatory in the joint and contributes to antigen-induced arthritis in mice. *Arthritis Res Ther.* 2014;16(4).
30. Prieto-Fernández E, Egia-Mendikute L, Bosch A, García del Río A, Jimenez-Lasheras B, Antoñana-Vildosola A, *et al.* Hypoxia promotes syndecan-3 expression in the tumor micro-environment. *Front Immunol.* 2020;11.
31. Hillemeier L, Espinoza-Sanchez NA, Greve B, Hassan N, Chelariu-Raicu A, Kiesel L, *et al.* The cell surface heparan sulfate proteoglycan syndecan-3 promotes ovarian cancer pathogenesis. *Int J Mol Sci.* 2022;23(10).
32. Yamada Y, Arai T, Kojima S, Sugawara S, Kato M, Okato A, *et al.* Regulation of antitumor miR-144-5p targets oncogenes: direct regulation of syndecan-3 and its clinical significance. *Cancer Sci.* 2018;109(9):2919-36.
33. Zhang H, Hui D, Fu X. Roles of Fibulin-2 in carcinogenesis. *Med Sci Monit.* 2020;26.
34. Hu L, Liu Y, Fu C, Zhao J, Cui Q, Sun Q, *et al.* The tumourigenic effect of the high expression of ladinin-1 in lung adenocarcinoma and its potential as a therapeutic target. *Mol (')ecules.* 2023;28(3):1103.
35. Ghafouri-Fard S, Khoshbakht T, Hussen BM, Taheri M, Samadian M. A review on the role of miR-1246 in the patho-etiology of different cancers. *Front Mol Biosci.* 2022;8:771835.
36. Choi HR, Nam KM, Park SJ, Kim DS, Huh CH, Park WY, *et al.* Suppression of miR-135b increases the proliferative potential of normal human keratinocytes. *J Invest Dermatol.* 2014;134(4):1161-4.
37. Zhu G, He Y, Yang S, Chen B, Zhou M, Xu XJ. Identification of gene and microRNA signatures for oral cancer developed from oral leukoplakia. *Biomed Res Int.* 2015.
38. Chen H, Liu X, Jin Z, Gou C, Liang M, Cui L, *et al.* A three-miRNAs signature for predicting the transformation of oral leukoplakia to oral squamous cell carcinoma. *Am J Cancer Res.* 2018;8(8):1403-.