Journal Pre-proof

Whole genome DNA methylation and mutational profiles identify novel changes in Proliferative Verrucous Leukoplakia

E.M. Okoturo , D. Green , K. Clarke , T. Liloglou , M.T. Boyd , R.J. Shaw , J.M. Risk

PII: S2212-4403(23)00081-0
DOI: https://doi.org/10.1016/j.oooo.2023.03.004
Reference: OOOO 5040

To appear in: *Oral Surg Oral Med Oral Pathol Oral Radiol*

Received date: 10 October 2022
Revised date: 9 January 2023
Accepted date: 6 March 2023

Please cite this article as: E.M. Okoturo , D. Green , K. Clarke , T. Liloglou , M.T. Boyd , R.J. Shaw , J.M. Risk , Whole genome DNA methylation and mutational profiles identify novel changes in Proliferative Verrucous Leukoplakia, *Oral Surg Oral Med Oral Pathol Oral Radiol* (2023), doi: https://doi.org/10.1016/j.oooo.2023.03.004

Whole genome DNA methylation and mutational profiles identify novel changes in Proliferative Verrucous Leukoplakia

E.M. Okoturo*[,a,1], D. Green*[b], K. Clarke[b], T. Liloglou[a,2], M.T. Boyd[a], R.J. Shaw[a], J.M. Risk[a#]

*: these authors contributed equally to this manuscript

a: Liverpool Head & Neck Centre, William Henry Duncan Building, University of Liverpool. L7 8TX. UK.

b: Computational Biology Facility, Liverpool Shared Research Facilities, University of Liverpool, Crown Street, Liverpool. L69, 7ZB. UK.

Present addresses:

1: Head and Neck Cancer Division, Oral & Maxillofacial Surgery Department, & Molecular Oncology Program, Medical Research Centre, Lagos State University College of Medicine, Ikeja, Lagos, Nigeria.

2: Edge Hill University, Medical School, St Helens Road, Ormskirk, Lancashire, L39 4QP

#: Corresponding author. Dr. J.M. Risk, Liverpool Head and Neck Centre, 3rd Floor William Henry Duncan Building, University of Liverpool, L7 8TX. UK. E: J.M.Risk@liverpool.ac.uk

Word counts:

Abstract: 197 Manuscript: 3640 Number of references: 41

Number of figures/tables: 3/3 Number of supplementary elements: 6

Declaration of interests: none

Funding: TO was supported by: Academic Staff Training and Development Programme (TETFund), Lagos State University, Ojo, Nigeria (LASU/VC/TETF/AST&D/001); the British Association of Oral and Maxillofacial Surgeons (BAOMS); The University of Liverpool PhD bursary; University of Liverpool Technology Directorate Voucher (40428830). The funding sources had no role in the study design, collection and analysis of data, writing of the manuscript or decision to publish.

Statement of clinical relevance

PVL, an oral potentially premalignant disorder with a high rate of transformation, shows a novel spectrum of mutations and differences in genome methylation compared with conventional oral SCC which may provide insight into its discrete aetiology.

Keywords

proliferative verrucous leukoplakia (PVL); carcinoma, verrucous; oral squamous cell carcinoma; oral potentially malignant disorder (OPMD); leukoplakia, oral; DNA methylation; whole exome sequencing

Abbreviations:

PVL-OSCC: PVL-associated OSCC

cOSCC: conventional (i.e. non-PVL related) OSCC

Abstract

Objectives: Proliferative verrucous leukoplakia (PVL) is a rare form of oral leukoplakia with a relatively high transformation rate to oral squamous cell carcinoma (OSCC). Molecular analysis of this disorder at the genome level is limited and only identifies molecular similarities between PVL and OSCC. However, the clinical profile suggests that molecular differences may be more important.

Study Design: Whole exome sequencing of five PVL-associated OSCC and paired blood samples identified somatic mutations common to the tumours. Whole methylome analysis of four PVL-associated OSCC and three OSCC of non-PVL origin used Infinium MethylationEPIC "850k" BeadChip, and differential methylation was explored.

Results: In contrast to conventional OSCC, PVL-associated OSCC showed infrequent *TP53* mutation and altered spectra of *PIK3CA* and *NOTCH1* mutations. Unsupervised hierarchical clustering identified 63 probes that discriminated between PVL-associated OSCC and OSCC of non-PVL origin. Differences in methylation were most significant for divalent metal ion transport, particularly calcium movement.

Conclusion: Specific differences in mutation and methylation profiles between PVL-derived OSCC and OSCC of non-PVL origin suggest differences in their transformation pathways. Further studies of early PVL lesions may identify markers of transformation that are also applicable to more common oral premalignant disorders such as oral epithelial dysplasia.

Introduction

Proliferative verrucous leukoplakia (PVL) is a rare form of oral leukoplakia first described in 1985 by Hansen *et al*[1]. The clinical course begins with an isolated oral white plaque that becomes multifocal and exophytic, often inexorably progressing over decades. It is persistent, irreversible and refractory to treatment, eventually undergoing malignant transformation to verrucous, then invasive, oral squamous cell carcinoma (OSCC) in more than 50% of cases[1,2], although overall survival in PVL-OSCC is significantly better than for conventional OSCC[3]. The diagnosis of PVL is based on a number of clinical and histopathological criteria and is frequently established retrospectively after malignancy has been diagnosed[1].

The molecular pathogenesis of PVL represents an intriguing and largely uncharted field of investigation. There are several critical clinical features that distinguish PVL from other head & neck mucosal lesions at risk of malignant transformation[4]. The very high rate of malignant transformation in PVL, well over 50% in most series[5], contrasts with oral epithelial dysplasia (OED) where the risk is only around 10% for mild to moderate dysplasia and 24% for severe dysplasia and carcinoma in situ[6]. We have speculated that PVL would transform in all patients if given sufficient natural lifespan and observation[7]. Further, PVL generally occurs independently of known risk factors for OSCC, and is specifically unrelated to smoking, alcohol consumption or high risk HPV[7]. The clinical and pathological appearance and progression of PVL to OSCC is highly stereotyped, following a predictable stepwise continuum. As such, it is possible that PVL follows a distinct molecular origin and pathway to conventional OSCC or OED, with critical genetic or epigenetic determinants that may be unique to PVL, or perhaps simply much less commonly seen in comparable TCGA metadata relating to head and neck SCC. Given the inexorable progression to OSCC, it is also hypothesised that the critical event to malignant transformation is present even in early lesions, prior to development of epithelial dysplasia.

Recent systematic reviews of the molecular pathogenesis of PVL[8,9] concluded that only aneuploidy showed promise as a putative marker of malignancy and highlighted the candidate gene approach of most previous studies. Evidently, these previous studies have been limited by their methodology to targeting aberrations commonly found in OSCC and only recently have there been reports of genome-wide studies of PVL that

exploit the power of contemporary sequencing approaches[10,11]. These two, related, studies investigated differential gene expression and DNA methylation in a small cohort of PVL patients in comparison to normal oral mucosa. The authors then validated their results *in silico* against the expression data for 314 specimens and 31 adjacent normal samples from the TCGA OSCC sample set (2015) in which they demonstrated similar expression profiles[10].

In contrast to the two studies that compared PVL with normal oral mucosa, our aim was to interrogate the genome-wide and epigenome-wide characteristics of PVL-associated OSCC and identify where the pattern of DNA methylation or somatic mutations contrasted with conventional OSCC. We hypothesised that this may give novel insight into the molecular mechanism of malignant transformation in OSCC.

Materials and Methods

Patients

Patients with PVL-associated OSCC (PVL-OSCC) were identified from cases under observation at the Oral Dysplasia Clinic of the Liverpool University Dental Hospital over the period 2008-2017 (Table 1). A diagnosis of PVL for this study was defined as a progressive clinical course based on Hansen's histological grading[1] with numerous biopsy episodes, post excisional recurrence and a propensity to develop into OSCC. Patients were included in this REC approved study (EC 47.01 & H/10/1002) after giving written, informed consent. The histopathology of the biopsied material was re-examined and cases with uncertainty about a diagnosis of PVL were excluded from this research, leaving a total of 22 patients.

Control cases were taken from a previously used cohort of OSCCs (REC number 07/Q1505/15[12]) (Table 1). Archival, formalin-fixed paraffin embedded (FFPE) tissue was preferred for single gene, confirmatory assays because this was comparable to the format of the majority of the PVL material: it was available for 20 cases. These 20 plus an additional 5 cases were available as fresh tissue. These cases will be referred to as conventional OSCCs (cOSSCs).

Samples

Following surgery, 5mm^3 tissue were obtained from within the tumour of PVL-OSCC and cOSCC, outwith any necrotic area or tumour margins. Tissue with histopathologically low Hansen grade (Grade 1-3) was obtained from the same or separate resections undertaken on the same day and termed low-grade PVL (LG-PVL). Tissue was snap frozen and stored at -80^{0}C. Paired blood samples were taken as controls for whole exome analysis. Genomic DNA from all fresh frozen samples was extracted using a DNEasy blood and tissue kit (Qiagen) according to manufacturer's protocol.

Archival, formalin-fixed paraffin embedded (FFPE) tissue from PVL and cOSCC cases was obtained from the local Pathology department. LG-PVL were taken from areas with Hansen's grade 1-3 within PVL-OSCC tumour specimens. Genomic DNA was extracted using a DNA FFPE Tissue kit (Qiagen) according to the manufacturer's protocol.

Whole Exome Sequencing (WES)

Genomic DNA from 5 fresh, frozen PVL-OSCC cases (Table 1) with paired blood samples were subject to Whole Exome sequencing at the Centre for Genomic Research (GRC), University of Liverpool. Briefly, target capture and library construction were undertaken using SureSelect Human All Exon v7 (Agilent). The libraries were sequenced on an Illumina Hiseq 4000 platform.

All reads were aligned to the reference genome (WG38) using BWA (Burrows-Wheeler Alignment tool)[13]. Somatic variants were identified by subtracting variants also detected in the paired blood sample genome sequence using Strelka2[14] utilising a base call quality score of 90% and a minor allele frequency of ≤1.0.

Whole Genome Methylation Profiling (WGMP)

Genomic DNA was isolated from fresh frozen samples obtained from 4 PVL-OSCC, 3 cOSCC and 4 premalignant PVL samples, 2 of which were obtained from tissue adjacent to the PVL-OSCC samples (Table 1). DNA samples were transferred to Edinburgh Genomics for bisulphite treatment and genome-wide microarray hybridisation using Infinium MethylationEPIC "850k" BeadChip. Duplicate hybridisation was undertaken for 3/4 PVL-OSCC, 1/3 cOSCC and 2/4 LG-PVL samples. Arrays were scanned on Illumina's iScan technology and the data transferred to the Computational Biological Facility (CBF), University of Liverpool.

Preprocessing and normalisation

EPIC arrays were analysed using functionality within the *minfi* package[15]. Briefly, samples were checked for global hybridisation quality based on an average detection p value threshold of < 0.05 for all probes. Individual CpG probes were discarded from the array if their detection p value was > 0.01, indicating a failed position. Probes were also discarded if they were located on chromosomes X & Y, within 2 base pairs of a single nucleotide polymorphism (SNP) with a minor allele frequency greater than 0.05 or if the probe was previously found to cross hybridise to multiple genomic locations[16,17]. Sample wise normalisation was performed using the regression of correlated probes (RCP) algorithm to correct for biases between type I and type II probe distributions[18].

Differential methylation

Univariate statistical analysis for each CpG was performed using the linear models for microarray data (limma) package[19]. The duplicateCorrelation function was used to estimate the consensus correlation between arrays originating from the same biological donor and was subsequently used as a parameter for lmFit. To implement the linear models, a contrast matrix was built with coefficients for 'PVL-OSCC and cOSCC' and 'PVL-OSCC and LG-PVL'. Significance was determined based on a Benjamini Hochberg (BH) adjusted FDR < 0.05

Differential methylation was assessed in the context of regions using DMRcate[20]. Regions were defined as blocks of 1000 nucleotides for which a Gaussian kernel smoothed function is fit. A region was considered to be differentially methylated if its BH corrected FDR was < 0.1 and had an absolute mean beta fold change > 0.1.

Analysis

All principal component analysis (PCA) were performed in R (Version 4.0.2). M values were centred and scaled prior to matrix decomposition using the *prcomp* function from the stats package[21]. Heatmaps were generated using the *pheatmap* function within the pheatmap package[22]. All hierarchical clustering was performed using the ward.D2 method. Statistical analysis to test association between principal components and variables in the data was undertaken using one-way ANOVAs. If there was a significant difference for

the main effect (p value < 0.05) a Tukey's Honest Significant Difference (TukeyHSD) post-hoc comparison was applied. Significance for post-hoc testing was accepted as an adjusted p value < 0.05.

Single gene assays

Somatic variants

Pyrosequencing assays for selected, PVL-OSCC- and cOSCC-specific, somatic changes were designed using Oligo 7.0 software (MBI). Oligonucleotides were synthesized by Eurofins Genomics, Ebersberg, Germany (Supplemental Table 1). Pyrosequencing was undertaken on 3 DNA samples from the WES cohort plus an additional 12 PVL-OSCC, 13 LG-PVL and 20 cOSCC FFPE specimens (Table 1) using Pyromark Gold reagents (Qiagen) on a Pyromark Q96 workstation (Qiagen) and the resulting pyrograms analysed for the presence of variants at the given site.

Promoter methylation

DNA was converted with sodium bisulphite (EZ DNA methylation-Gold, Zymo Research). Quantitative methylation specific PCR (qMSP) and pyrosequencing methylation (PMA) assays were designed using Oligo 7.0 (MBI) and PyroMark® Assay Design SW 2.0 software (Qiagen), respectively. Oligonucleotides were synthesized by Eurofins Genomics, Ebersberg, Germany (Supplemental Table 1). Lymphocyte DNA artificially methylated to 20%, 10%, 5% and 2.5% were used as standards as previously described[23].

For qMSP of the *CDKN2A* (p16) promoter, a previously published method was used[23]. DNA from 12 PVL-OSCC and 11 cOSCC were analysed. Assays were multiplexed with *ACTB* as an internal normalisation control and run in duplicate. The mean C_t value from the duplicates was used to calculate relative quantitation (RQ) values using the ΔΔCt method with the unmethylated lymphocyte DNA being used as the reference and RQ was compared to the standards to obtain a semi-quantitation of DNA methylation.

For PMA of the *MGMT* promoter, PCR products from 20 PVL-OSCC and 20 cOSCC FFPE specimens were subject to pyrosequencing using Pyromark Gold reagents (Qiagen) on a Pyromark Q96 workstation (Qiagen) and the resulting pyrograms analysed for the presence of methylation as previously described[27].

Statistical analysis

A Students t test was used for comparisons between PVL-OSCC and cOSCC variant frequencies

Results

Whole exome sequencing

Seven genes were identified as having PVL-OSCC specific mutations (present in at least 3/5 samples: Table 2). Two of these are also commonly mutated in conventional OSCC (*NOTCH1* and *PIK3CA*). An interrogation of the WES data for alterations in other genes mutated at more than 15% in conventional OSCC revealed changes in PVL-OSCC samples at lower frequencies (Table 2).

The genetic changes observed in the *TTN, LRP2, LRP5, JMJD1C* and *EPG5* genes in PVL-OSCC were classified as of moderate-low impact because they produced amino acid changes with limited structural effect or no amino acid change (Supplemental Table 2). However, the changes observed in *NOTCH1 a*nd *PIK3CA* were classified as of high impact, producing amino acid changes that altered protein secondary/tertiary structure or inserted a stop codon. Analysis of mutation profiles demonstrated the presence of mostly novel genetic alterations in the PVL-OSCC samples and absence of more commonly observed mutations (Columns 1 & 3, Table 3). The only two mutations observed in both PVL- and conventional OSCC were at *PIK3CA* p.E545K, a mutation found in around 9% of all cancers annotated in TGCA, and *NOTCH1* p.R353C, a mutation found at low frequency in both PVL- and conventional OSCC. Pyrosequencing assays were designed to cover *PIK3CA* p.E545K and p.E542K mutations, the *PIK3CA* p.H1047R mutation, and *NOTCH1* p.G326D and p.C387X/p.C387Y mutations. Samples for pyrosequencing included DNA from specimens used to generate the primary WES data, LG-PVL, to assess early/late stage appearance and conventional OSCC samples. In this larger cohort, the frequency of mutation for *PIK3CA* p.E545K fell to levels comparable to those observed in conventional OSCC (P>0.05)(Table 3). Notably, 2/3 of the LG-PVL samples demonstrating *PIK3CA* p.E545K mutations also showed that mutation in the corresponding PVL-OSCC tissue.

Whole Genome Methylation Profiling (WGMP)

Global PCA

Principal component analysis (PCA) was used to assess variance based on methylation profiles between PVL-OSCC, cOSCC and LG-PVL tissue. PCA based on unbiased, genome-wide methylation profiles of ~785,000 probes demonstrated significant differences in variance within principal component 1 (PC1) between tissue from LG-PVL and tissue from cOSCC ($p = 0.026$) and PVL-OSCC ($p = 0.005$) (Figure 1).

However, no significant differences in variance were observed between cOSCC and PVL-OSCC in PC1 or later components, indicating similar global methylation patterns in the two tissues.

Univariate models

A total of 63 differentially methylated probes between PVL-OSCC and cOSCC were identified using a univariate model in limma with a Benjamini Hochberg FDR < 0.05. (Supplementary Table 3). Proportional to the distribution of probes tested on the EPIC array, the 63 differentially methylated probes contain 1.7 fold enrichment for probes located within CpG islands with 42/63 probes located within a coding region of a gene. A further 19 regions were identified based on defined blocks of 1000 nucleotides within the 63 CpGs to be differentially methylated (DMRs) based on an FDR of 10%. (Supplementary Table 4). Interestingly, the promoters of genes that are commonly methylated in cOSCC, such as *CDKN2A*, were absent from both of these lists.

Targeted multivariate clustering and PCA

Multivariate unsupervised hierarchical clustering and principal component analysis showed that all 3 tissues; LG-PVL, PVL-OSCC and cOSCC could be discriminated by the methylation profiles of the 63 differentially methylated loci (Figure 2A). Interestingly, this analysis grouped LG-PVL more closely with cOSCC than with PVL-OSCC. Moreover, using only the 63 probes identified from the univariate model, PC1 and PC2 were shown to capture ~90% variance and to be significantly associated with tissue type (Figure 2B), with PC1 separating PVL-OSCC from cOSCC and PC2 separating LG-PVL from both OSCC tissue types. LG-PVL was grouped more closely to cOSCC than PVL-OSCC in PC1. However, 16 of the 63 differentially methylated loci showed little or no difference in methylation levels between PVL-OSCC and LG-PVL and may be early events in the development of PVL-OSCC (Supplementary Table 5). Seven of the 16 DMPs are located in a CpG island and 4 of these are associated with known genes: *MACF1, SFRP2, LMX1A* and *ME3*.

Gene Ontology

The 63 differentially methylated probes were analysed for overlapping biological processes. Differences in methylation were most significant for divalent metal ion transport, particularly calcium movement

(Supplementary table 3). Interestingly, these gene promoters were generally less methylated in PVL-OSCC compared with cOSCC.

Analysis of gene promoters commonly methylated in OSCC

None of the gene promoters that are commonly hypermethylated in cOSCC were identified in the list of probes that were differentially methylated between PVL-OSCC and cOSCC (Supplementary Tables 3 & 4). This implied that the methylation levels at these promoters are similar in PVL-OSCC and cOSCC. In order to investigate this further, multivariate unsupervised hierarchical clustering was undertaken using 315 probes representing 24 genes shown to have significantly altered promoter methylation in OSCC[25,26] (Supplementary Table 6). This confirmed the separation of LG-PVL tissue from all cancer tissue and suggested that, although there were many similarities in the methylation profiles of PVL-OSCC and cOSCC, some differences were observable (Figure 3).

In addition, the methylation status of the *CDKN2A* (p16) promoter was investigated by RT-qMSP[23] and the promoter of *MGMT* by PMA[27] in DNA prepared from FFPE tissue. None of the PVL-OSCC tumours demonstrated *CDKN2A* promoter methylation compared with 5/9 (56%) cOSCC tumours (P=0.013). Similarly, only 3/18 (17%) PVL-OSCC showed *MGMT* promoter methylation compared with 5/18 (28%) cOSCC, but this was not statistically significant (P=0.33).

Discussion

PVL transforms into OSCC at a frequency of more than 50% in comparison with 10% for mild to moderate OED and 24% for severe OED and carcinoma in situ[5,6]. Historically, molecular studies have attempted to show that the molecular profile of PVL-OSCC is similar to cOSCC. However, the current study aimed to investigate the differences between these two entities in an effort to explain the differences in their transformation rates.

Whole exome sequencing identified 7 genes presenting with mutations in ≥60% of samples, only 2 of which (*PIK3CA* and *NOTCH1*) had previously been identified as commonly mutated in cOSCC[24]. In contrast, *TP53* mutations were less common in this cohort. Of course, these data may reflect the small sample size used,

but the *TP53* data reflect previous observations in PVL[28]. The *PI3CA* mutations identified in PVL-OSCC tumours were different to those most commonly observed in cOSCC, but have been previously shown to have pathogenic potential in other cancers, most notably breast and endometrium[29]. The p.T1025A alteration located in the PI3K/PI4K domain of the protein has been shown to increase transforming ability in cell lines in culture and is predicted to be a gain of function mutation[30]. The p.E545K alteration is located in the helical domain, is comonly observed in a number of cancers and results in an amino acid substitution of opposite charge that uncouples the catalytic and regulatory subunits of the protein[31], as does the p.E542K mutation that is more commonly observed in head and neck cancer. Interestingly, the p.E545K mutation has also been shown to alter fatty acid metabolism, including the upregulation of arachadonic acid that is needed for the synthesis of prostaglandins, a pro-inflammatory molecule implicated in cancer development[32] and may thus be more important in the pathogenesis of transformation. The p.E545K mutations of *PI3CA* were also observed in early (LG-PVL) lesions, which suggests that this oncogenic driver is an early occurring event. However, there is no support in the literature for a role for early mutation of *PI3CA* in OED. *NOTCH1* alterations can be oncogenic or tumour suppressive in HNSCC[33]. The mutations observed in this study are all located in the extracellular subunit and may affect receptor binding, with 2 of them changing cysteine residues and potentially affecting secondary and tertiary structure that is reliant on disulphide bonding.

Methylation profiling comparing PVL-OSCC with non-malignant LG-PVL identified 157,963 differentially methylated probes and 26,498 differentially methylated regions. If the list is restricted to promoters (located -1500 to -200bp upstream from the start site and within the 5'UTR) then 4,871 islands are differentially methylated. This is of a similar scale to that described by Herreros-Pomares et al who identified 4,647 DMRs[10] in a series of 10 PVL when compared with 5 healthy tissues. However, there were no common targets in the top 50 DMRs from the two datasets. Herreros-Pomares then showed similarities in profiles between PVL data and OSCC on TCGA for 13 differentially methylated, differentially expressed probes. In contrast, the data presented in the current paper identifies *differences* between the methylation profiles of PVL-OSCC and cOSCC. Unsurprisingly, none of the 63 DMPs that are differentially methylated were identified by Herreros-Pomares as having similar methylation profiles in PVL and OSCC.

It may be postulated that DMPs that are differentially expressed in PVL-OSCC compared with cOSCC but that show similar methylation levels in a comparison of LG-PVL with PVL-OSCC indicate early events in the transformation process. Of the four known genes that fell into this group, only *MACF1* showed a methylation pattern consistent with tumour development. This microtubule associated protein is implicated in proliferation, migration and cell signalling in several cancers[34] and showed reduced methylation in PVL-OSCC. In contrast, reduced expression of *SFRP2* and *LMX1A* have been previously associated with colorectal and bladder cancer, respectively[35,36], but both showed decreased methylation in PVL samples in our study. Similarly, *ME3* shows increased expression in pancreatic cancer[37], but was characterised by increased methylation in our PVL samples. It is well known that methylation status does not always negatively correspond with expression and it should be remembered that the relative levels of methylation reported in the current study are in comparison with cOSCC and not with normal tissue. Nevertheless, the implication is that *MACF1* is a gene worthy of further study in the development of PVL-OSCC.

Global hypomethylation coupled with targeted promoter hypermethylation is a key feature of OSCC[38,39]. Subtle differences between OSCC populations have been documented[40], so it is not surprising that we have identified differences between the methylation patterns in PVL-OSCC and cOSCC. None of these differences in methylation were in gene promoters that are commonly methylated in cOSCC, so we hypothesized that these were critical, and common, to the transformation process in all oral cancer. However, multivariate unsupervised hierarchical clustering of 315 probes representing 24 genes shown to have significantly altered promoter methylation, either in our previous study[25] or in several candidate gene studies[26] did show differences between PVL-OSCC and cOSCC methylation patterns. Interestingly, comparatively lower levels of methylation were observed in PVL-OSCC compared with cOSCC for a number of these probes, and we validated this in the *CDKN2A* gene promoter. This suggests that different mechanisms are involved in the development of PVL-OSCC compared with cOSCC.

Gene ontology analysis of the 63 differentially methylated probes identified divalent metal ion transport, particularly calcium movement, and other signalling pathways as being affected, although the gene

promoters were generally less methylated in PVL-OSCC compared with cOSCC, suggesting increased expression of these pathways. However, the pattern was by no means uniform, with some pathway members being hypermethylated in cOSCC and others in PVL-OSCC. Calcium movement is associated with intracellular signalling which is often activated in cancer, but this is not the only role of this cation. Increased calcium in the mitochondria can stimulate the calcium-sensitive dehydrogenases of the Kreb's cycle leading to increased ATP production and increased release of superoxides[41], which in turn may stimulate proliferation and DNA damage, respectively.

In conclusion, we have identified specific differences in mutation and methylation profiles between PVL-associated OSCC and conventional OSCC, suggesting differences in the transformation pathway in these two entities. However, it was not possible to identify a single genetic or epigenetic pathway / gene unique to PVL. Although the data is based on few samples, these were carefully curated for a clinical and pathological diagnosis of PVL and augmented with samples from premalignant, low grade PVL. Further studies should concentrate on analysis of these earlier lesions in order to identify markers of transformation that may also be applicable to more common oral premalignant disorders such as oral epithelial dysplasia[42].

References

[1] Hansen LS, Olson JA, Silverman SJ. Proliferative verrucous leukoplakia. A long-term study of thirty patients. Oral Surgery, Oral Medicine, and Oral Pathology. 1985;60:285-98.

[2] Silverman SJ, Gorsky M. Proliferative verrucous leukoplakia: a follow-up study of 54 cases. Oral Surgery,Oral Medicine, Oral Pathology, Oral Radiology and Endodontology. 1997;84:154-7.

[3] Gonzáles-Moles MÁ, Warnakulasuriya S, Ramos-García P. Prognosis parameters of oral carcinomas developed in proliferative verrucous leukoplakia: A systematic review and meta-analysis. Cancers (Basel). 2021;13:4843.

[4] Akrish S, Ben-Izhak O, Sabo E, Rachmiel A. Oral squamous cell carcinoma associated with proliferative verrucous leukoplakia compared with conventional squamous cell carcinoma - a clinical, histologic and immunohistochemical study. Oral Surgery, Oral Medicine, and Oral Pathology and Oral Radiology. 2015;119:318-25.

[5] Ramos-García P, Gonzáles-Moles MÁ, Mello FW, Bagan JV, Warnakulasuriya S. Malignant transfromation of oral proliferative verrucous leukoplakia: A systematc review and meta-analysis. Oral Dis. 2021;27:1896-907.

[6] Mehenna HM, Rattay T, Smith J, McConkey CC. Treatment and follow-up of oral dysplasia - a systematic review and meta-analysis. Head & Neck. 2009;31:1600-9.

[7] Borgna SC, Clarke PT, Schache AG, Lowe D, Ho MW, McCarthy CE, et al. Management of proliferative verrucous leukoplakia: justification for a conservative approach. Head & Neck. 2017;39:1997-2003.

[8] Okoturo EM, Risk JM, Schache AG, Shaw RJ, Boyd MT. Molecular pathogenesis of proliferative verrucous leukoplakia: a systemic review. British Journal of Oral and Maxillofacial Surgery. 2018;56:780-5.

[9] Rintala M, Vahlberg T, Salo T, Rautava J. Proliferative verrucous leukoplakia and its tumor markers: Systematic review and meta-analysis. Head & Neck. 2019;41:1499-507.

[10] Herreros-Pomares A, Llorens C, Soriano B, Bagan L, Moreno A, Calabuig-Fariñas S, et al. Differentially methylated genes in proliferative verrucous leukoplakia reveal potential malignant biomarkers for oral squamous cell carcinoma. Oral Oncol. 2021;116:105191.

[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. Clinical Oral Investigation. 2021;25:2645-57.

[12] Shaw RJ, Hobkirk AJ, Nikolaidis G, Woolgar JW, Triantafyllou A, Brown JS, et al. Molecular staging of surgical margins in oral squamous cell carcinoma using promoter methylation of $p16^{INK4A}$, cytoglobin, E-cadherin, and *TMEFF2.* Annals of Surgical Oncology. 2013;20:2796-802.

[13] Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25:1754-60.

[14] Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nature Methods. 2018;15:591-4.

[15] Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30:1363-9.

[16] Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;8:203-9.

[17] Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. 2016;17:208.

[18] Niu L, Taylor JA. RCP: A novel probe design bias correction method for Illumina Methylation BeadChip. Bioinformatics. 2016;32:2659-63.

[19] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analysis for RNA-sequencing and microarray studies. Nuc Acids Res. 2015;43:e47.

[20] Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, Lord RV, et al. De novo identification of differentially methylated regions in the human genome. Epigenetics & Chromatin. 2015;8:6.

[21] RCoreTeam. R: A language environment for statistical computing.: R Foundation for Statistical Computing, Vienna, Austria; 2021.

[22] Kolde R. pheatmap: Pretty heatmaps. R package version 1.0.12. 2019.

[23] Nikolaidis G, Raji O, Markopoulou S, Gosney J, Bryan J, Warburton C, et al. DNA methylation biomarkers offer improved diagnostic efficiency in lung cancer. Cancer Res. 2012;72:5692-701.

[24] TCGA. NCI GDC Data Portal. Data Release 29.0 ed.

[25] Jithesh PV, Risk JM, Schache AG, Dhanda J, Lane B, Liloglou T, et al. The epigenetic landscape of oral squamous cell carcinoma. Br J Cancer. 2013;108:370-9.

[26] Mascolo M, Siano M, Ilardi G, Russo D, Merolla F, De Rosa G, et al. Epigenetic disregulation in oral cancer. International Journal of Molecular Sciences. 2012;13:2331-52.

[27] Shaw RJ, Hall GL, Lowe D, Bowers NL, Liloglou T, Field JK, et al. CpG island methylation phenotype (CIMP) in oral cancer: Associated with a marked inflammatory response and less aggressive tumour biology. Oral Oncol 2007;43:878-886

[28] Gopalakrishnan R, Weghorst CM, Lehman TA, Calvert RJ, Bijur G, Sabourin CL, et al. Mutated and wild-type p53 expression and HPV integration in proliferative verrucous leukoplakia and oral squamous cell carcinoma. Oral Surg Oral Med Oral Pathol Oral Radiol Endodontics. 1997;83:471-7.

[29] Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nuc Acids Res. 2019;47:D941–D7.

[30] Ng PK, Li J, Jeong KJ, Shao S, Chen H, Tsang YH, et al. Systemic functionnal annotation of somatic mutations in cancer. Cancer Cell. 2018;33:450-62.

[31] Leontiadou H, Galdadas I, Athanasiou C, Cournia Z. Insights into the mechanism of the *PI3CA* E545K activating mutation using MD simulations. Scientific Reports. 2018;8:15544.

[32] Koundouros N, Karali E, Tripp A, Valle A, Inglese P, Perry NJS, et al. Metabolic fingerprinting links oncogenic *PIK3CA* with enhanced arachidonic acid-derived eicosanoids. Cell. 2020;181:1596-611.

[33] Fukusuki T, Califano J. The NOTCH Pathway in Head and Neck Squamous Cell Carcinoma. J Dent Res. 2018;97:645-53.

[34] Miao Z, Ali A, Hu L, Zhao F, Yin C, Chen C, et al. Microtubule actin cross-linking factor 1, a novel potential target in cancer. Cancer Sci. 2017;108:1953-8.

[35] Zhang H, Qi J, Wu Y-Q, Zhang P, Jiang J, Wang Q-X, et al. Accuracy of early detection of colorectal tumours by stool methylation markers: A meta-analysis. World J Gastroenterol. 2014;20:14040-50.

[36] Zhao Y, Guo S, Sun J, Huang Z, Zhu T, Zhang H, et al. Methylcap-Seq reveals novel DNA methylation markers for the diagnosis and recurrence prediction of bladder cancer in a Chinese population. PLoS ONE. 2012:7:e35175.

[37] Zhang Q, Li J, Tan XP, Zhao Q. Effects of ME3 on the proliferation, invasion and metastasis of pancreatic cancer cells through epithelial-mesenchymal transition. Neoplasma 2019;66:896-907.

[38] Foy J-P, Pickering CR, Papadimitrakopoulou V, Jelinek J, Lin SH, William Jr WN, et al. New DNA methylation markers and global DNA hypomethylation are associated with oral cancer development. Cancer Prev Res. 2015;8:1027-35.

[39] Chatterjee R, Das S, Chandra A, Basu B. Epigenome-Wide DNA Methylation Profiles in Oral Cancer. In: Wei LK, editor. Computational Epigenetics and Diseases: Academic Press; 2019. p. 213-31.

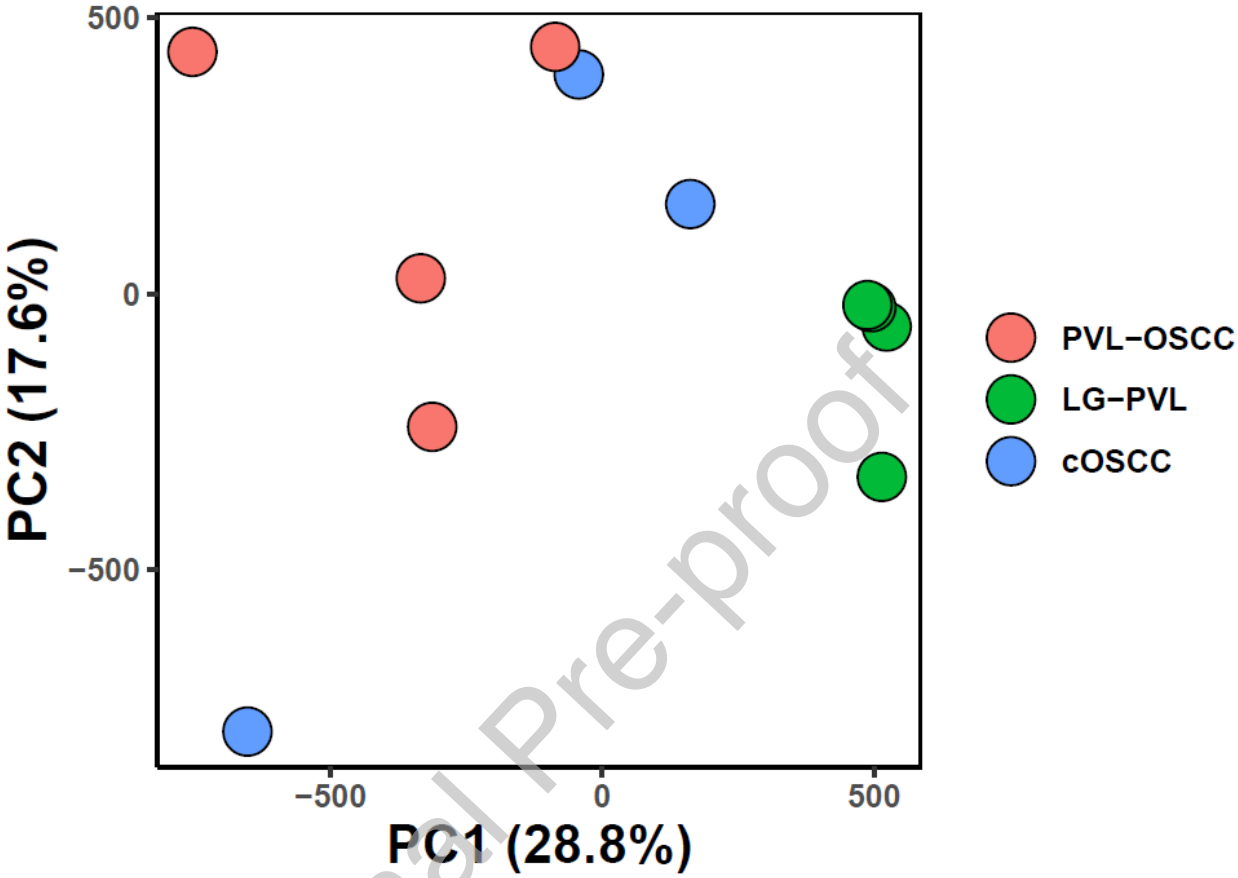
[40] Basu B, Chakraborty J, Chandra A, Katarkar A, Baldevbhai JRK, Chowdhury DD, et al. Genome-wide DNA methylation profile identified a unique set of differentially methylated immune genes in oral squamous cell carcinoma patients in India. Clinical Epigenetics. 2017;9:13.

[41] McCormack JG, Halestrap AP, Denton RM. Role of calcium ions in regulation of mammalian intramitochondrial metabolism. Physiology Reviews. 1990;70:391-425.

[42] Morandi L, Gissi D, Tarsitano A, Asioli S, Gabusi A, Marchetti C, et al. CpG location and methylation level are crucial factors for the early detection of oral squamous cell carcinoma in brushing samples using bisulfite sequencing of a 13-gene panel. Clinical Epigenetics. 2017;9:85.

Figure Captions

Figure 1. Unbiased analysis of genome-wide methylation differences between cOSCC, LG-PVL and PVL-OSCC. A: principal component analysis; B: box & whisker plot of data from principal component 1 (PC1).



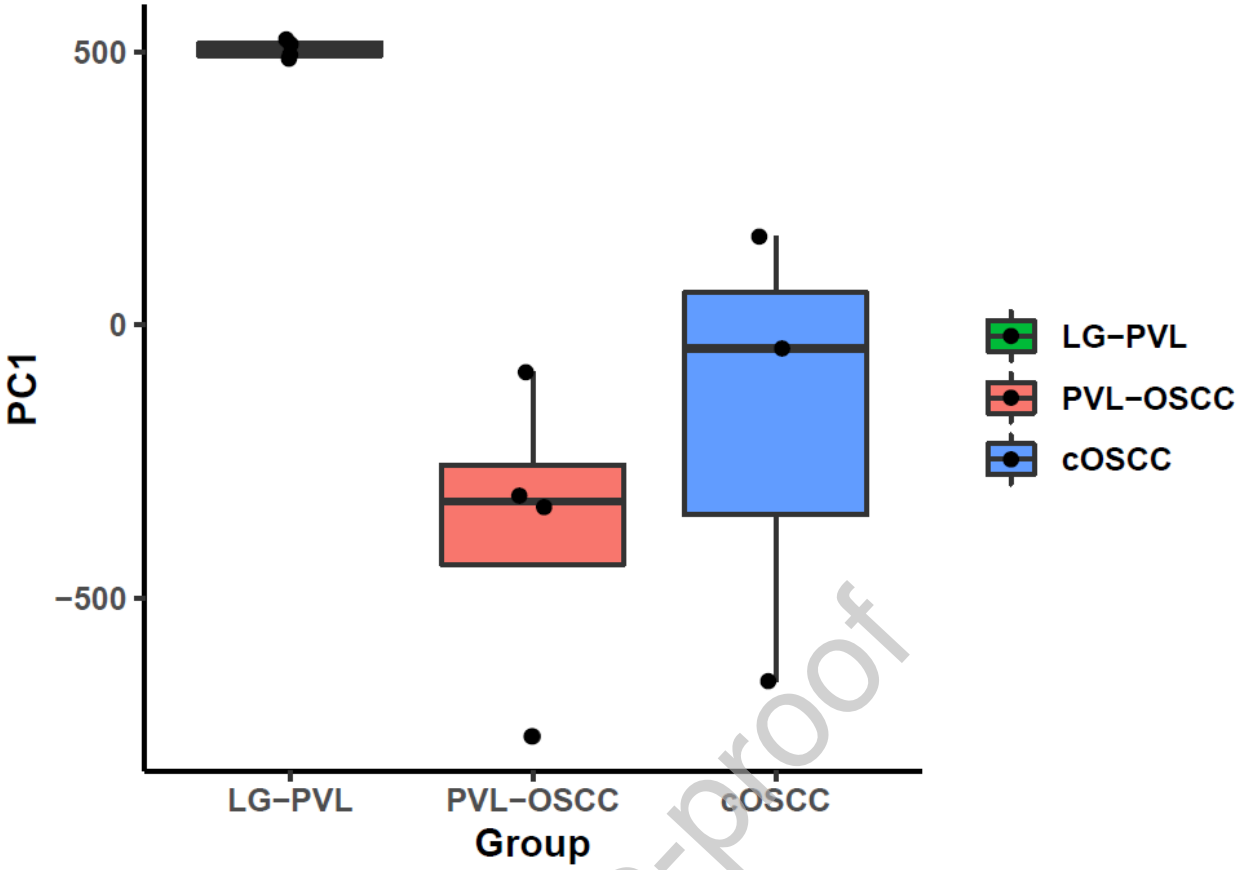
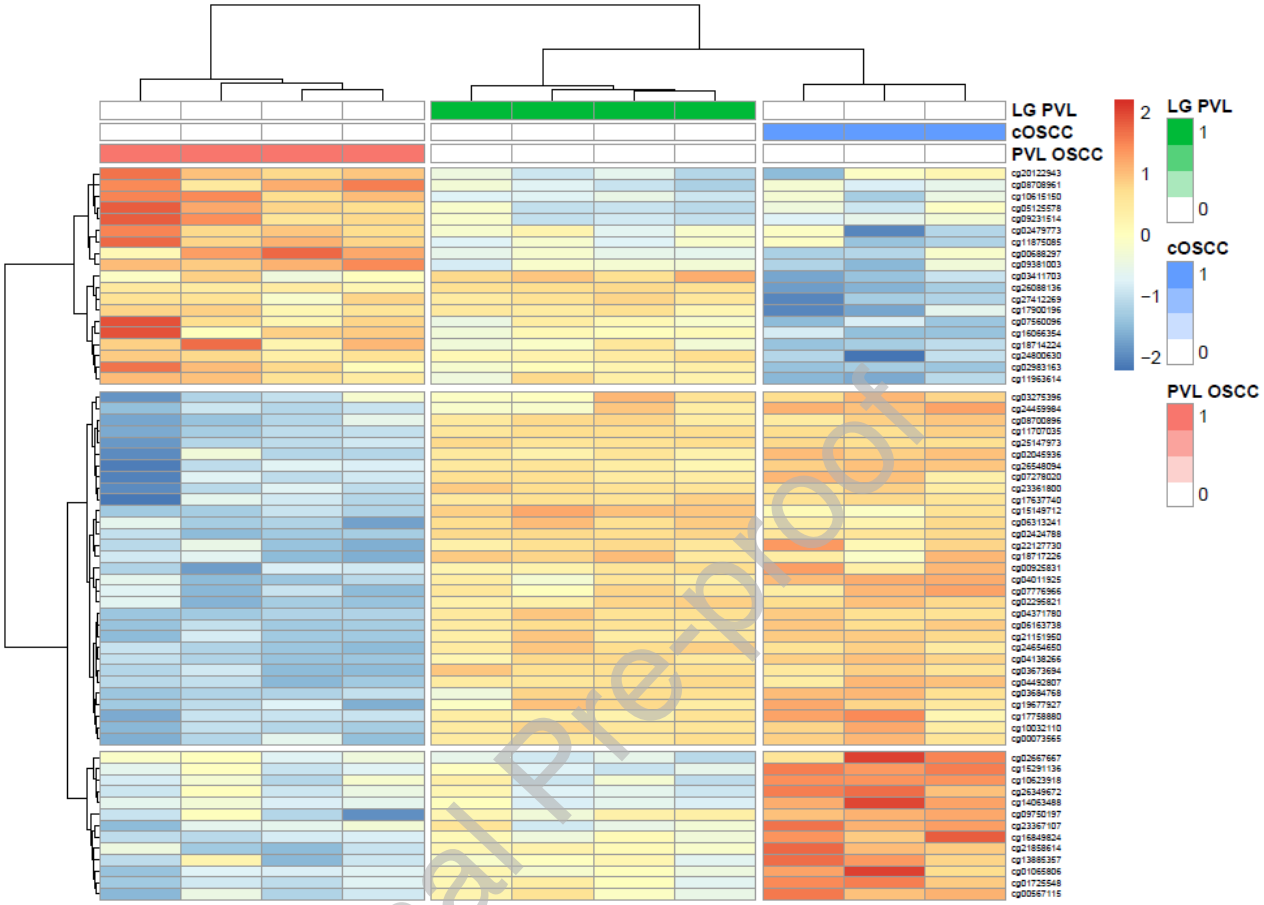


Figure 2. Multivariate clustering using 63 differentially methylated probes. A: Unsupervised hierarchical clustering based on scaled M values of methylation at 63 differentially methylated probes; B: Principal component analysis of the 63 differentially methylated probes.



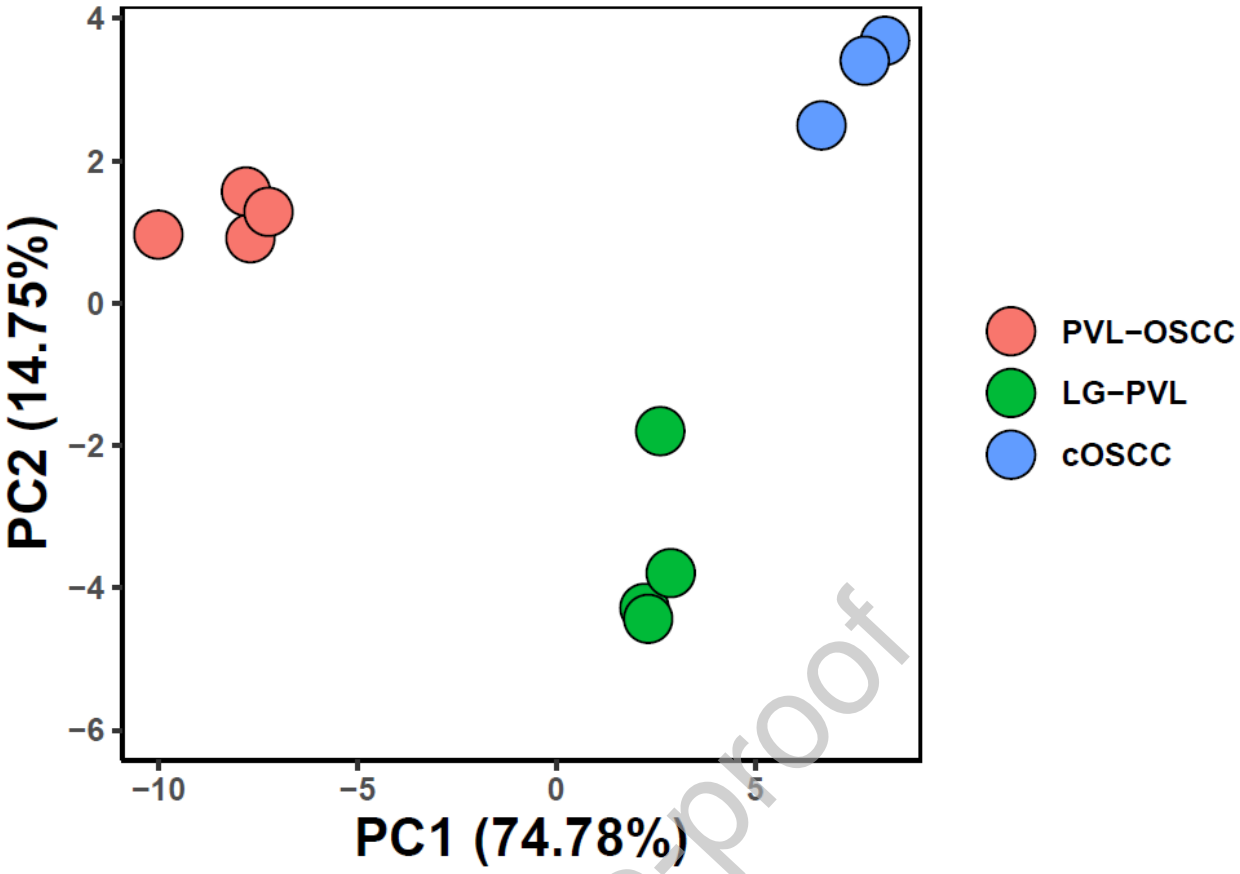


Figure 3. Multivariate clustering using 315 probes associated with differentially methylated genes identified from the literature. Unsupervised hierarchical clustering based on scaled M values of methylation

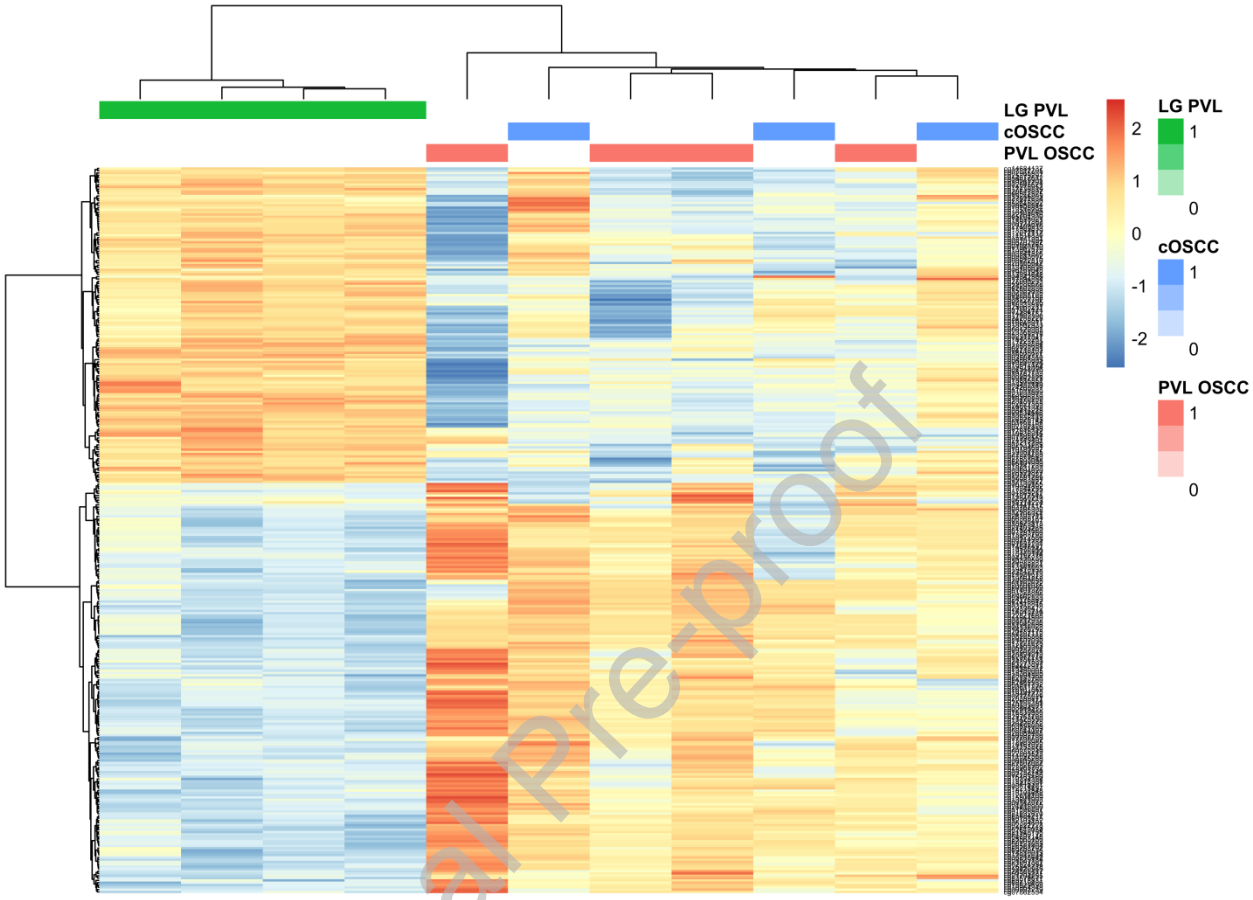


Table 1 Patient cohort

Patient Group	Identifier	OSCC[1] sample	WGA[2] OSCC	LG-PVL[1] sample	WGA[2] LG-PVL	Sex	age at first tumour	Tumour site[3]	PVL Lesion extent
PVL	3328	Y	WES	Y	WGMP	F	64	mandibular alveolus	mandibular gingiva
	4146	Y	WES	Y	WGMP	F	77	FOM	panoral
	4107	Y	WES			F	75	lateral tongue	panoral
	3275	Y	WES/WGMP	Y		F	69	buccal	buccal
	3576	Y	WES/WGMP	Y	WGMP	M	61	FOM	panoral

	3390	Y	WGMP	Y	WGMP	F	70	ventral tongue	hypopharynx
	4244	Y	WGMP			F	95	lateral tongue	panoral
	3002	Y		Y		M	40	FOM	panoral
	3114	Y		Y		F	63	buccal	panoral
	3144	Y		Y		F	80	soft palate	hypopharynx/buccal
	3287	Y		Y		M	42	lateral tongue	lateral tongue
	3294	Y		Y		F	66	hard palate	buccal/hard palate
	3343	Y		Y		F	70	buccal/soft palate	buccal
	3265	Y				F	89	BOT/lingual tonsil	buccal/FOM
	3450	Y				M	60	mandibular alveolus	panora
	3698	Y		Y		M	72	mandibular alveolus	panoral
	3866	Y				F	85	lateral tongue	panoral
	3906	Y				M	35	buccal	buccal
	3935	Y		Y		M	57	FOM	FOM
	3936			Y		M	N/A		FOM
	3406			Y		M	N/A		mandibular alveolus/gingiva
	4224	Y				F	42	lateral tongue	panoral
TOTAL		20		15					
cOSCC n=25	3314	Y	WGMP			M		FOM	
	3269	Y	WGMP			M	66	alveolus	
	3276	Y	WGMP			M	59	FOM	
	3327	Y				M	67	tonsil	
	3329	Y				M	48	FOM	
	3330	Y				M	59	lateral tongue	
	3334	Y				M	64	maxilla	
	3335	Y				M	64		
	3336	Y				M	57	mandibular alveolus	
	3337	Y				M	53	soft palate	
	3338	Y				M	50	ventral tongue	
	3345	Y				M	76	tongue	
	3346	Y				M	76	ventral tongue	
	3347	Y				M	65	mandibular alveolus	
	3350	Y				F	89	mandibular alveolus	
	3352	Y				M	75	FOM	
	3354	Y				F	48	ventral tongue	

3355	Y	M	65	tongue
3361	Y	M	57	tongue
3364	Y	M	63	tongue
3366	Y	F	72	tongue
3367	Y	F	88	maxillary alveolus
3379	Y	M	63	buccal
4012	Y	M	57	tonsil
4035	Y	F	63	mandibular alveolus

[1]: Y in the column indicates that a sample was available for analysis

[2]: Whole Genome Analysis undertaken on this sample (if any). WES: Whole exome sequencing; WGMP: Whole genome methylation profiling

[3]: FOM: floor of mouth; BOT: base of tongue

Table 2 Most frequent genetic changes in PVL-OSCC and conventional OSCC

Gene name	Frequency in PVL-OSCC (this study: n=5)	Frequency in oral SCC[a] (n-311)
LRP2	80% (4)	0
LRP5	80% (4)	0
TTN	60% (3)	0
NOTCH1	60% (3)	21.5% (66)
PIK3CA	60% (3)	16.9% (52)
JMJD1C	60% (3)	0
EPG5	60% (3)	0
TP53	40% (2)	78.2% (240)
FAT1	40% (2)	29.3% (90)
CDKN2A	20% (1)	26.1% (80)

[a] Data generated by the TCGA Research Network[24]; numbers in parenthesis are numbers of samples showing somatic changes

Table 3 Single mutations screened in PVL samples

Gene Name	Mutation	Frequency in PVL-OSCC		Frequency in conventional OSCC		Frequency in LG-PVL
		WES (n=5)	Resequencing	TCGA[a]	Resequencing (n=18)	Resequencing (n=13)
PIK3CA	p.E542K (GAG-GTG)	0	0	8/112 (7%)	0	0
	p.E545K (GAG-AAG)	2 (40%)	2/15 (13%)	10/112 (9%)	3 (17%)	3 (23%)
	p.H1047R (CAG-CGT)	0	0	11/112 (12%)	NT	0
	p.T1025A (ACC-GCC)	1 (20%)	NT	0	NT	NT
NOTCH1	p.G326D (GGC-GAC)	1 (20%)	1/11 (9%)	0	0	NT
	p.R353C (CGT-TGT)	1 (20%)	NT	1/86 (1%)	NT	NT
	p.C387X (TGC-TGA)	1 (20%)	1/14 (7%)	0	0	NT
	p.C387Y (TGC-TAC)	0	1/14 (7%)	0	0	NT
	p.C627G (TGC-GGC)	1 (20%)	NT	0	NT	NT

p.Q1247X (CAG-TAG)	1 (20%)	NT	0	NT	NT

[a] data from TCGA[24]

Numbers in parentheses indicate percentage of samples demonstrating that mutation; NT=not tested

Author contribution

E. Okoturo: Data curation, Formal analysis, Funding acquisition, Investigation, Writing- Original draft, Writing- Review & editing **D. Green**: Data curation, Formal analysis, Visualization, Writing- Original draft, Writing- Review & editing. **K. Clarke**: Data curation, Visualization, Writing- Review & editing. ***T. Liloglou****:* Data curation, Formal analysis, Writing- Review & editing. **M.T. Boyd**: Conceptualization, Writing- Review & editing. **R.J. Shaw:** Conceptualization, Funding acquisition, Resources, Writing- Original draft, Writing- Review & editing. **J.M. Risk:** Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Supervision, Visualization, Writing- Original draft, Writing- Review & editing

Statement of clinical relevance

PVL, an oral potentially premalignant disorder with a high rate of transformation, shows a novel spectrum of mutations and differences in genome methylation compared with conventional oral SCC which may provide insight into its discrete aetiology.