

RESEARCH

Open Access



Total RNA sequencing reveals gene expression and microbial alterations shared by oral pre-malignant lesions and cancer

Mohammed Muzamil Khan^{1,2}, Jennifer Frustino³, Alessandro Villa^{4,5}, Bach-Cuc Nguyen⁶, Sook-Bin Woo⁷, William Evan Johnson⁸, Xaralabos Varelas⁹, Maria Kukuruzinska^{6*} and Stefano Monti^{1,2,10*}

Abstract

Head and neck cancers are a complex malignancy comprising multiple anatomical sites, with cancer of the oral cavity ranking among the deadliest and the most disfiguring cancers globally. Oral cancer (OC) constitutes a subset of head and neck cancer cases, presenting primarily as tobacco- and alcohol-associated oral squamous cell carcinoma (OSCC), with a 5-year survival rate of ~65%, partly due to the lack of early detection and effective treatments. OSCC arises from premalignant lesions (PMLs) in the oral cavity through a multi-step series of clinical and histopathological stages, including varying degrees of epithelial dysplasia. To gain insights into the molecular mechanisms associated with the progression of PMLs to OSCC, we profiled the whole transcriptome of 66 human PMLs comprising leukoplakia with dysplasia and hyperkeratosis non-reactive (HkNR) pathologies, alongside healthy controls and OSCC. Our data revealed that PMLs were enriched in gene signatures associated with cellular plasticity, such as partial EMT (p-EMT) phenotypes, and with immune response. Integrated analyses of the host transcriptome and microbiome further highlighted a significant association between differential microbial abundance and PML pathway activity, suggesting a contribution of the oral microbiome toward PML evolution to OSCC. Collectively, this study reveals molecular processes associated with PML progression that may help early diagnosis and disease interception at an early stage.

Author summary

Patients harboring oral premalignant lesions (PMLs) have an increased risk of developing oral squamous cell carcinoma (OSCC), but the underlying mechanisms driving transformation of PMLs to OSCC remain poorly understood. In this study, Khan et al., analyzed a newly generated dataset of gene expression and microbial profiles of oral tissues from patients diagnosed with PMLs from differing histopathological groups, including hyperkeratosis not reactive (*HkNR*) and dysplasia, comparing these profiles with OSCC and normal oral mucosa. Significant similarities between PMLs and OSCC were observed, with PMLs manifesting several cancer hallmarks, including oncogenic and immune pathways. The study also demonstrates associations between the abundance of multiple microbial species and PML groups, suggesting a potential contribution of the oral microbiome to the early stages of OSCC development. The study offers insights into the nature of the molecular, cellular and microbial heterogeneity of oral PMLs

*Correspondence:

Maria Kukuruzinska

mkukuruz@bu.edu

Stefano Monti

smonti@bu.edu

Full list of author information is available at the end of the article

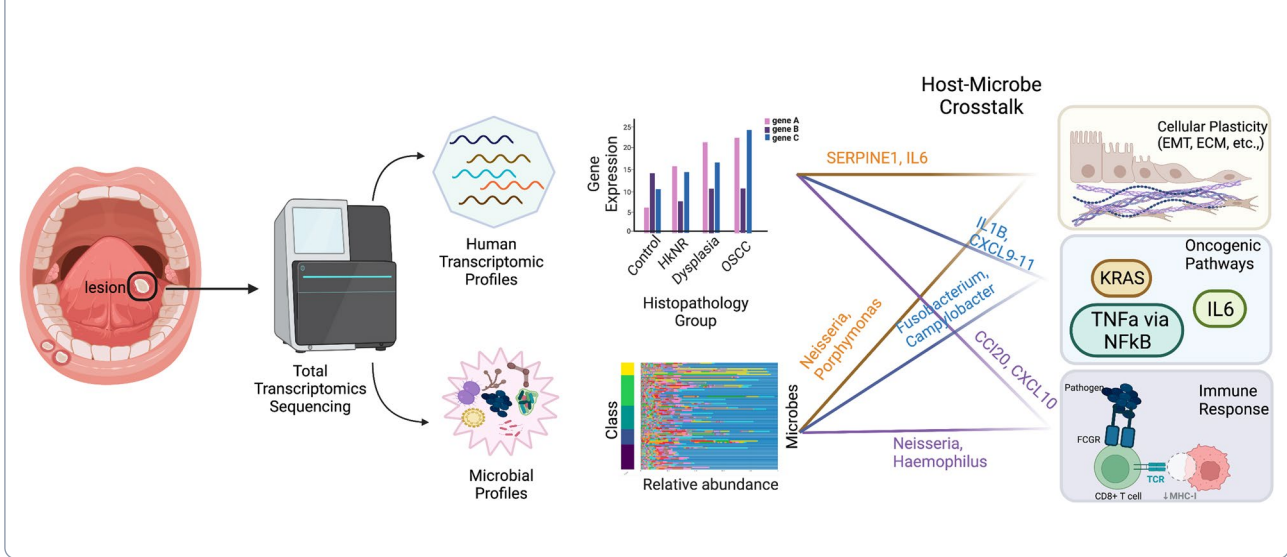


© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

and suggests that molecular and clinical refinement of PMLs may provide opportunities for early disease detection and interception.

Keywords Pre-malignant lesions, Oral cancer, Oral microbiota, Early detection

Graphical abstract



Introduction

Head and neck cancer of the oral cavity ranks among the top ten most diagnosed cancers globally, with tobacco- and alcohol-associated oral squamous cell carcinoma (OSCC) pathology. OSCC arises from oral epithelial mucosa in a multi-step process in which normal cells are transformed into preneoplastic cells and then to cancer [1–3]. These premalignant lesions (PMLs)—also referred to as oral potentially malignant disorders (OPMD) [4], but mostly presenting as leukoplakia and variants, and erythroplakia—transform to OSCCs through various histopathological stages, from hyperkeratosis/hyperplasia (or hyperkeratosis not reactive, *HkNR*, known previously as *keratosis of unknown significance(KUS)*), to various degrees of dysplasia [5], carcinoma in situ (CIS), and finally to invasive OSCC [6]. The presence of dysplastic areas in the epithelium of the oral cavity has been associated with progression to cancer. Although oral dysplasia may progress to OSCC, the initial cause of their malignant transformation remains unclear. While many leukoplakias exhibit histopathology of dysplasia or invasive SCC at the time of biopsy, many non-dysplastic keratotic lesions (HkNR) also transform to invasive OSCCs over time, suggesting that such lesions may represent early dysplasia before the histologic phenotype of dysplasia is established [1, 3]. The genetic signatures of *leukoplakia* without dysplasia and other early oral lesions have not

been well characterized. Thus, a key step to improve outcomes of OSCC is to identify the molecular factors driving disease initiation and progression, as these factors represent attractive candidates for disease interception through targeted therapies.

Here, we sought to uncover the mechanisms involved in OSCC development from PMLs by employing total RNA-sequencing and advanced computational analyses. A comprehensive knowledge of global alterations in gene expression programs and the detailed characterization of cancer-associated transcriptional and microbial signatures will provide new insights into the underlying molecular networks involved in OSCC initiation and progression, allowing the development of prognostic models of early OSCC pathobiology and the identification of markers for cancer risk assessment to improve prediction accuracy and disease outcome. Gene expression profiling represents a powerful tool to investigate progression in early oral dysplasia and HkNR [7, 8]. Importantly, adoption of a total RNA-sequencing technique allows us to simultaneously characterize the microbiota of the biopsies and to correlate microbial differences with host gene/pathway activity and clinical-pathologic features. While the role of the gut microbiome in inflammation and colorectal cancer has received much attention, little is known about the role of the oral microbiome in pre-malignant

oral lesions [9], and our study contributes to filling this knowledge gap.

Materials and methods

Subject population

A cohort of 66 individuals presenting with early oral lesions was chosen for this study (Table 1). Participants were enrolled prospectively from the Brigham and Women’s Hospital in Boston, MA, and Erie County Medical Center in Buffalo, NY, with protocols approved by the institutional review boards at the study sites. All study participants provided written informed consent. All patients underwent an oral biopsy and histopathological examination. The histopathology groups included normal mucosa (controls, n = 18, 27%), hyperkeratosis not reactive (HkNR, n = 17, 26%), moderate or severe dysplasia (dysplasia, n = 22, 33%), and oral squamous cell carcinoma (OC, n = 9, 14%). The patients with progression status ‘Progressed to Dysplasia’ (n = 2; 25%) in the dysplasia group transformed to

an advanced stage of the dysplasia pathology (e.g., mild to moderate and moderate to severe status). Smoking status for the missing patients was imputed using a random forest classification approach (see Data Analysis section below). In our study, the HkNR and dysplastic groups are collectively referred to as PML.

Laboratory techniques

Sample preparation

All samples (n = 66) were processed for RNA purification using miRNeasy Mini Kit (Qiagen Cat # 217004) according to manufacturer’s instructions. Briefly, frozen tissue samples were homogenized in QIAzol lysis reagent using the TissueLyser II for 30 s and stored at – 70 °C until all samples were collected. Next, tissue homogenates were brought up to room temperature and mixed with chloroform to extract RNA followed by centrifugation at 12,000×g at 4 °C for 15 min. The top aqueous phase (~ 350 µl) was mixed extensively with 1.5 volumes of 100% ethanol by pipetting. Samples were then loaded onto RNeasy Mini spin columns and centrifuged at

Table 1 Patient summary by histopathology and clinical factors

Characteristic	Histopathology group			
	Control N = 18 (27%) ¹	Hyperkeratotic; Not Reactive (HkNR) N = 17(26%) ¹	Dysplasia N = 22 (33%)*	OSCC N = 9(14%) ¹
Age	66 (51,73)	67 (30, 81)	66 (44, 92)	68 (49, 80)
Sex				
Female	11 (61%)	11 (65%)	11 (50%)	5 (56%)
Male	7 (39%)	6 (35%)	11 (50%)	4 (44%)
Race				
Asian	0 (0%)	0 (0%)	1 (14%)	0 (0%)
Black	2 (18%)	0 (0%)	1 (14%)	0 (0%)
White	9 (82%)	2 (100%)	5 (71%)	3 (100%)
Unknown	7	15	15	6
Smoking_status				
No	11 (69%)	7 (54%)	10(53%)	7 (88%)
Yes	5 (31%)	6 (46%)	9 (47%)	1 (12%)
Unknown	2	4	3	1
Alcohol status				
Current	14(88%)	9 (69%)	13(72%)	4 (50%)
Former	0 (0%)	0 (0%)	3(17%)	1 (12%)
Never	2 (12%)	4 (31%)	2 (11%)	3 (38%)
Unknown	2	4	4	1
Progression status				
Stable	7 (70%)	5(71%)	3 (38%)	2 (100%)
Progressed to dysplasia	1 (10%)	2 (29%)	2 (25%)	0 (0%)
Progressed to SCC	2 (20%)	0 (0%)	3 (38%)	0 (0%)
Unknown	8	10	14	7

Biopsy from patients (n = 66) are separated into four histopathology groups and segregated byage, smoking, and disease progression status

¹ Median (Range); n (%)

8,000×g for 15 seconds at room temperature. The flow through was discarded, and this step was repeated once. RNA was further purified by DNA digestion on RNeasy Mini spin columns (Cat # 7925) and RNA was eluted by centrifugation at 8,000×g in 40 µl of RNase-free water. RNA concentrations and purity were determined using a NanoDrop 2000 spectrophotometer.

RNA sequencing

Total transcriptomic sequencing was performed on biopsy samples from 66 patients at BU Microarray and Sequencing Resource Core Facility. Libraries were prepared from 100 ng of total RNA from each sample using KAPA RNA HyperPrep Kit with RiboErase (HMR) (Kapa Biosystems, USA) according to the manufacturer's protocol and sequenced on an Illumina NextSeq 500 instrument with 1.45 pM input and 1% PhiX control library spiked in (Illumina, USA) targeting 30–40 Million reads pairs per sample.

Preprocessing and quality control

Human The fastq files obtained from sequencing runs were aligned and mapped using the *align()* and *featureCounts()* from the Rsubread [10] package. This method uses the 'seed and vote' strategy to map seeds/subreads (small read fragments) to a given genomic location and picks the highest voted aligned subread and is much faster than other aligners. The human reference genome version used was GRCh38.99. Samples with a mapping quality of >40% were retained. Genes with counts greater than one count-per-million (cpm) in at least 9 samples (the lowest number of samples in a phenotypic group) were preserved and used for downstream analyses.

Microbiome To obtain the metagenomic profiles, the same sequencing files that were used for human alignment and mapping were processed with Pathoscope 2.0 [11]. The tool first removes all reads that align to the host transcriptome, with the remainder aligned and mapped to reference microbial taxonomies such as bacteria, fungi, and viruses. Bowtie alignment (the aligner used within Pathoscope) was run with parameters '-local -k' (multiple local alignments option) set to 20, '-min-score' set to 90% (to extract only fragments that map to microbial genomes with high confidence) and read length, *L*, set to 70 (from fastQC). The microbial reference genome used was the 2020 version of the RefSeq library that includes reference genomes for bacteria, virus, and fungi. These reference genomes include the taxonomic levels from Phylum to Species and yielded 948 operational taxonomical units (OTU) consisting of 323 genera (across all samples) with a read count of ~607K per sample on average. The whole analysis was carried out using the *animalcules* R package. [12]

Data analysis

Signature projection

Several transcriptional signatures were used to interrogate our dataset (Figs. 1C and 4). For each signature of interest, either paired sets of up- and down-regulated genes, or a one-directional geneset were defined. For each geneset, an aggregate "activity" score was computed by *Gene Set Variation Analysis* (GSVA) [13]. For bidirectional signatures, a combined score was computed based on the difference of the up- and down-regulated GSVA scores. The following signatures were derived from publicly available datasets: (1) a signature of the comparison of Adjacent Epithelial (AE, *n* = 44) with Head and Neck Squamous Cell Carcinoma (HNSCC, *n* = 500) from the TCGA [14], consisting of 1317 up- and 1330 down-regulated markers in HNSC; (2) a signature comparing malignant-transforming (MT, *n* = 10) with not-transforming (NT, *n* = 10) *oral potentially malignant disorders* (OPMD) [15], consisting of 9 up- and 28 down-regulated markers. OPMD were there defined as oral lesions diagnosed as leukoplakia or erythroleukoplakia, and carrying a potential risk of malignant transformation [15]; 3) a one-directional signature of "partial epithelial-to-mesenchymal transition" (pEMT), a molecular phenotype characteristic of cells spatially localized to the leading edge of primary tumors and shown to be an independent predictor of nodal metastasis [16].

Dimensionality reduction and visualization

UMAP dimensionality reduction analysis was performed based on the 2nd through 10th principal components (PC) derived from the top 500 highly variable genes in the dataset. Probabilistic clustering as implemented in the *mclust* [17] package was performed to estimate the number of clusters, which yielded *k* = 3 clusters, followed by application of *k*-means clustering, which yielded the cluster assignments depicted in Fig. 1B. Cluster enrichment for any of the lesion groups was assessed based on a three-group *chi-square* test, with the group labeling a cluster identified as the one with a positive and higher *chi-square* score than the other groups.

Smoking status imputation

Smoking status, when missing, was estimated based on imputation. To this end, multiple classifiers—random forest (RF), support vector machine (SVM), and *K* nearest neighbors (KNN)—were evaluated by tenfold cross validation on the dataset restricted to the top 1000 genes with highest median absolute deviation (MAD) and the

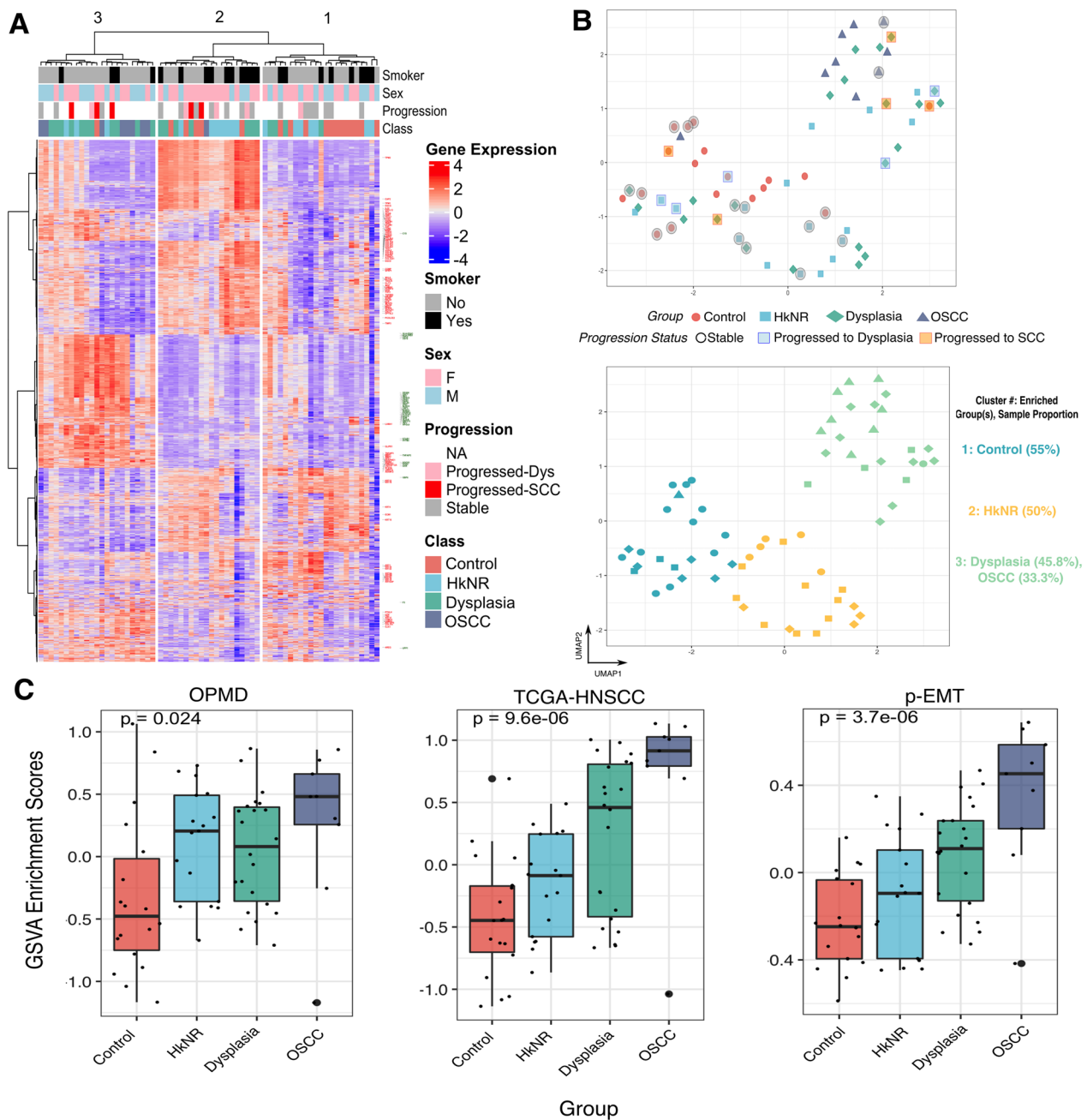


Fig. 1 PMLs show transcriptional heterogeneity. **A** Heatmap of top 2000 highly varying genes, with rows (genes) and columns (patients) sorted by hierarchical clustering. The columns are split into three groups as determined by a 3-way cut of the dendrogram. Row annotation highlights epithelium-related genes in red and immune-related genes in green. **B** (top) UMAP analysis of gene expression profiles from all samples colored by histopathological group. The progression status of 27 patients is denoted with black circles (stable), blue squares (progressed to dysplasia), and red squares (progressed to SCC). **B** (bottom) Clustering analysis using K-means, with $k=3$. Within each cluster, the group positively enriched by chi-square test is highlighted. **C** Boxplots of GSVA-based enrichment scores. Displayed from left to right are: score differences of up- and down-regulated OPMD signatures ($n=37$ genes; up=9, down=27); score differences of up- and down-regulated TCGA HNSCC signatures ($n=2647$; up=1317, down=1330); and scores from a unidirectional p-EMT signature ($n=100$). P-values are obtained using an ANOVA test

samples with available smoking status, with the RF classifier yielding the best area under the curve (AUC=0.79). A RF classifier trained on all complete data (n=56) was then applied to the prediction of the missing smoking status (n=10).

Differential analysis

Differential gene expression analysis was performed based on DESeq2 [18]. Age, sex, and smoking status were included as covariates in the analysis. Differential signatures were defined as the sets of transcripts with FDR-corrected q-value ≤ 0.05 and log₂ fold-change ≥ 1.5 for

up-regulated signatures and log₂ fold-change ≤ -1.5 for down-regulated signatures.

Pathway enrichment analysis

Differential signatures were annotated by over-representation analysis (ORA) based on hyper-geometric test using the hypeR package [19] and the HALLMARK and REACTOME geneset compendia available through MSigDB [20, 21]. hypeR was run with the complete set of genes in the dataset as background (n=21,500), and the significant pathways were defined as those with an FDR-corrected q-value ≤ 0.1. The hierarchy of pathways as

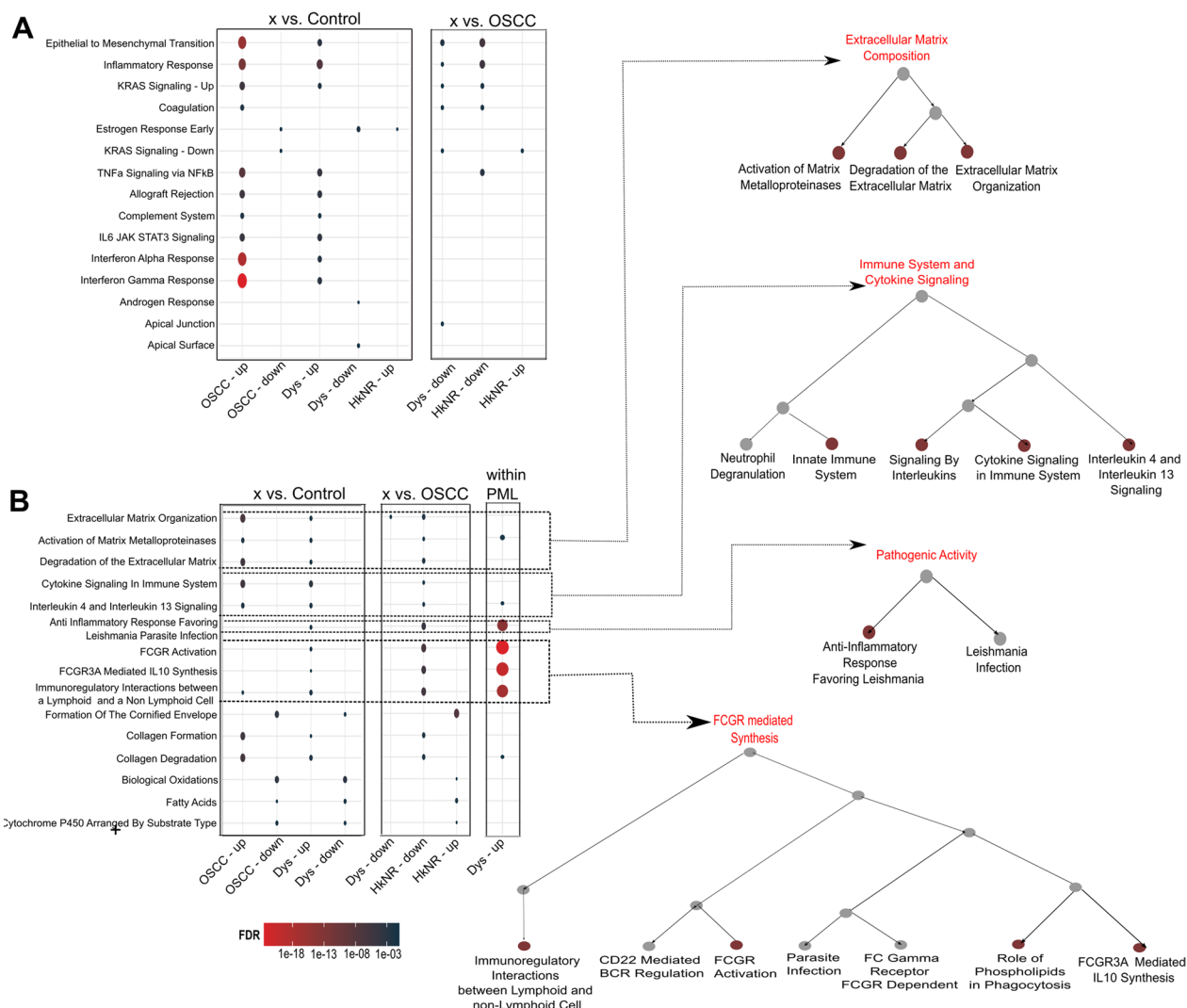


Fig. 2 Pathway enrichment of the PML and OSCC signatures. **A** Over-representation-based enrichment results based on the Hallmarks compendium of differentially expressed genes from pairwise analysis with control and OSCC groups. The size of the dot is proportional to the number of overlapping genes, and the color coding represents the FDR-corrected q-val. **B** Over-representation-based enrichment results based on the Reactome compendium. The hierarchical organization of related pathways is shown on the right

depicted in Fig. 2B was derived using the hierarchicalSets [22] method.

Immune cell type deconvolution

Using TPM-normalized (transcripts per million) counts, *CIBERSORT* [23] was applied with reference signatures from immune cell types derived from human hematopoietic cells consisting of 22 main immune cell-types.

Microbiome analysis

The microbiome analysis was carried out using *animalcules* [12], which provides various functions to create a multi-assay experiment object that contains the relative abundance of microbes detected in the study and several visualization methods. The relative abundance and diversity plots in Additional file 1: Figures S5 A-C were generated using this package. DESeq2-based [18] differential analysis of microbes abundance between groups was also performed using an *animalcules* wrapper. An FDR-corrected q -value ≤ 0.05 was used to select differentially abundant microbes.

Microbe-set enrichment analysis and pathway activation

The microbe-set enrichment analysis (MSEA) [24] python package was used to establish significant associations between microbes and genes. MSEA defines gene-labeled microbe-sets by tallying gene-microbe co-occurrences in PUBMED publications, and by then grouping microbes that co-occur with the same gene. For example, the microbe-set for gene UHRF2 consists of the 4 microbes (*Porphyomonas*, *Haemophilus*, *Fusobacterium*, and *Campylobacter*) found to “significantly” co-occur with that gene across publications based on a Jaccard index criteria. A library of $\sim 1,300$ microbe-sets thus defined, corresponding to as many genes, was used for our analysis [24].

Next, microbial genera found to be differentially abundant in OSCC and PML compared to the control group were tested for enrichment against this microbe-set library, and the results visualized as a bipartite graph consisting of microbe-gene interactions with a significant FDR-corrected q -value ≤ 0.05 and a positive MSEA combined score.

Finally, the union of mset-labeling genes was tested for enrichment of the significant Hallmark genesets found in the host analyses (OSCC vs. control and PML vs. control, Fig. 2A). Genes in the overlap between Hallmarks and the host differential signatures were tested against the mset-labeling genes, and those significant at a $q \leq 0.25$ were reported.

Results

Patient population

A cohort of 66 individuals selected for this study included patients with oral lesions representing controls (normal mucosa, 27%), hyperkeratosis not reactive (HkNR, 26%), moderate or severe dysplasia (dysplasia, 33%), and oral squamous cell carcinoma (OSCC, 14%) (Table 1). The mean age of patients was 55. Clinical characteristics such as smoking status—current or former, and follow-up information on the progression of the lesions were included, when available. Progression status for 27 of the 66 patients was available, with 10 patients progressing to the next stage of pre-cancer or OSCC pathology. In particular, 2 patients with HkNR histopathology transformed to dysplasia, and 3 patients with dysplastic phenotype progressed to OSCC, accounting for $\sim 8\%$ of the PML population. Additionally, 2 patients in the control group progressed to OSCC.

The transcriptional profiles of PMLs display heterogeneity

We selected the top 2000 highly variable genes based on median absolute deviation (MAD) and visualized their log-transformed and scaled expression profiles on a heatmap in Fig. 1A, with the rows (genes) and columns (samples) clustered across histopathological groups in an unsupervised manner. The heatmap was partitioned into the three top branches identified by hierarchical clustering, which was independent of the probabilistic clustering results shown in Fig. 1B (bottom panel). Importantly, the groupings displayed in Fig. 1A, B captured similar salient features of the histopathological groups, attesting to the robustness of the transcriptionally driven patient segregation. In particular, Cluster 1 predominantly contained control samples, cluster 2 consisted primarily of PMLs and some of the control samples, and cluster 3 included all but one OSCC samples along with some PMLs. Of note, cluster 3 also contained the control sample that progressed to OSCC. Several genes within pathways that were the focus of subsequent sections are highlighted. For example, genes associated with epithelial and extracellular matrix-related functions, such as COL7A1, TGFB1, LAMA3, and KRT10, are labeled in red, and immune-related genes, such as HLA-A, HLA-B, and STAT1, are labeled in green. From this heatmap a pattern emerged where all the OSCC samples and some of the dysplastic and HkNR samples showed high expression in most of the genes related to these processes, with the remaining samples showing patterns that matched the control group. Importantly, some of the samples from the non-cancer groups that progressed to OSCC showed transcriptional patterns similar to those of OSCC.

While the heatmap shows transcriptional heterogeneity, the unsupervised organization of the PML groups determined using a UMAP dimensionality reduction approach showed the clear separation of the control and OSCC groups, with the PML groups distributed along a gradient between the two extremes (Fig. 1B, top panel). Within these groups, most of the dysplastic samples were located proximal to the OSCC group region on the upper right side of the plot, while the HkNR samples were more scattered in-between suggesting more heterogeneous phenotypes. Model-based probabilistic clustering [17] was applied to estimate in an unbiased, data-driven fashion the number of clusters and their composition, which yielded a 3-cluster partition. The resulting cluster assignments are shown in Fig. 1B (bottom panel), with cluster 1 comprising a majority of control samples, cluster 2 comprising some control samples and a significant proportion of HkNR, and cluster 3 including all the OSCC samples and a significant proportion of the dysplastic samples.

Figure 1B (bottom panel) also shows the progression status of patients ($n=27$) obtained through follow-up visits. Progression status is denoted by red and blue shaded boxes, while black denotes a stable condition. Patients with dysplasia that transformed to OSCC segregated toward the cancer cluster (cluster 3). Of note, a control sample that progressed to OSCC was clustered with the cancer group (cluster 3) by the UMAP analysis, suggesting clinical relevance of the associated molecular profile. This sample had a unique transcriptional profile that shared similarities with the SCC group as shown in Fig. 1A.

Taken together, these unsupervised analyses highlight the considerable transcriptional heterogeneity of PMLs. The limited availability of lesions with long-term follow-up precluded us from drawing stronger conclusions, but we note that PMLs likely to progress share greater transcriptional similarity with OSCCs than non-progressing PMLs.

PMLs are enriched for malignant-transforming and partial-EMT signatures

To evaluate the transcriptional patterns associated with progressive malignant transformation in our datasets, we leveraged multiple signatures from published studies capturing salient features of such a transformation. These included: a tumor versus normal signature derived from the cancer genome atlas (TCGA) of HNSCC RNAseq dataset [14]; a signature of malignant-transforming (MT) vs. not-transforming (NT) OPMD [15]; and a signature of “partial-EMT” (pEMT)—a molecular phenotype shown to be an independent predictor of nodal metastasis [16].

For each of the signatures, GSVA-based enrichment scores corresponding to the up- and down-regulated genes were estimated for each of the samples, and the difference scores (up–down) were then computed. The scores were then stratified by histopathology group (Control, HkNR, Dysplasia, OSCC), as shown in Fig. 1C.

When stratifying the samples by the OPMD signature, HkNR showed higher and more variable enrichment scores among the PML. Stratification by the TCGA HNSCC “tumor vs. normal” signature showed a clear upward trend tracking with progression, as did stratification by the p-EMT signature, a defining characteristic of more aggressive HNSCs [16].

In summary, these results confirm that our data capture salient features of transformation identified in previous studies and help further refine the transcriptional programs shared by different oral lesions. In particular, known signatures of premalignant and malignant transformation and tumor aggressiveness tracked with the canonical lesion progression represented by our histopathology groups, with interesting exceptions. Notably, the HkNR group displayed a higher level of the OPMD signature (but also higher variability) than the putatively more advanced dysplasia.

PML signatures are associated with oncogenic and immune-altering pathways

The observed highly varying transcriptional profiles of PMLs (Fig. 1A–B) may contribute to current challenges in their evaluation and clinical treatment even with histopathology information. To gain further insight into the transcriptional programs defining the different groups, we performed pairwise differential gene expression and pathway analyses using *DESeq2* and *hypeR*, respectively. Differential analysis was performed by comparing each remaining group to controls, to OSCC, and by pairwise comparing the PML sub-groups, while controlling for sex and smoking status. The number of significant markers along with differential analysis results for each pairwise comparison is reported in the Additional file section (Additional file 1: Figure S1 and Additional file 2, Additional file 3, Additional file 4: Tables S1–S3). As shown in Fig. 2, the signature genes with a $\log_{2}FC$ cutoff of ± 1.5 and $q\text{-value} \leq 0.05$ were found to be significantly enriched for several immune response pathways defined in the Hallmark and Reactome compendia. Hallmark pathways that were significantly enriched in the OSCC and dysplastic groups compared to controls included key cancer progression and transformation pathways, such as epithelial-to-mesenchymal transition (EMT) and KRAS signaling, as well as inflammatory response pathways such as TNF- α signaling via NF- κ B, interferon gamma (IFN- γ) and alpha (IFN- α) response, IL2/STAT5,

and IL6/JAK-STAT3 signaling. Genes driving the EMT enrichment in both groups included extracellular matrix-related genes such as ADAM12, TGFB1, LAMA3, MMP1 and MMP3, and in the OSCC group multiple collagen genes (COL11A1, COL12A1, COL4A1) and serine proteinase inhibitors SERPINE1 and SERPINH1, among others, while genes driving the IFN- γ and IFN- α enrichment in both groups included multiple chemokine ligands (CXCL9, CXCL10, CXCL11), and in the OSCC group multiple interferon induced proteins (IFI1-3, IFI35, IFI44, etc.) among others (see Additional file 3: Table S2). The comparisons to the cancer group revealed similar enrichment patterns of immune response pathways in the cancer group compared to the dysplastic and the HkNR groups. The signatures derived from comparing HkNR and dysplasia groups were not significantly enriched for any of the Hallmark pathways, while they were enriched for multiple pathways from the Reactome collection.

The differential signatures were then tested for enrichment with pathways in the Reactome compendium to evaluate potential metabolic and pathogenic associations. Multiple pathways from the same biological processes were found to be enriched and were thus organized and visualized with their hierarchical structure in Fig. 2B. This analysis showed a significant enrichment, in PMLs and OSCCs when compared to control, of tumor microenvironment-related activities, such as extracellular matrix (ECM) organization, and activation of matrix metalloproteinases. Confirming the Hallmarks-based analysis, the Reactome-based analysis also identified and further resolved several enrichments of immune and pathogenic pathways. Signaling by interleukins, the FC gamma receptor (FCGR) mediated regulation, and FCGR phagocytosis were enriched in the OSCC and dysplasia groups when compared to controls and HkNR groups. Figure 2B shows their hierarchical organization, with multiple FCGR-related pathways, as well as multiple immune-related pathways.

The strong enrichment of p-EMT and EMT signatures in the PML groups raised the question of whether this enrichment stems from increased plasticity within the epithelial compartment or from alterations in adjacent stroma. To investigate this, we tested our series for enrichment with respect to a cancer-associated fibroblast (CAF) signature previously described [25], inclusive of genes COL1A1, COL1A2, COL3A1, and PDGFRB, among others. Interestingly, the signature was significantly enriched in the non-control groups, with the strongest enrichment observed in the PML groups (Additional file 1: Figure S2B). Furthermore, CAF enrichment was significantly positively associated with p-EMT enrichment (Additional file 1: Figure S2C), with the

association stronger in PMLs. These observations suggest that CAF infiltration and increased plasticity co-occur in PMLs, and single cell-based studies will be needed to further elucidate their interplay.

PMLs and OSCC have significantly higher levels of immune activity than controls

Our differential gene and pathway analyses showed that the PMLs shared significant similarities with the OSCC group. Given the observed up-regulation of multiple immune pathways, we next focused on the immune landscape by investigating changes in immune cell proportions across healthy and disease groups. To this end, we performed immune cell type deconvolution using a gene signature compendium comprising 22 immune cell-types derived from human hematopoietic cells. The cell proportions estimated with CIBERSORT, visualized as boxplots in Fig. 3, were stratified by histopathology group at different levels of resolution. Immune cell types were first divided into the innate and adaptive immune components in the top panel of Fig. 3. These were further subdivided into their main subtypes, with the innate component subdivided into macrophages (M0-like, M1-like and M2-like), monocytes, mast cells, NK cells and neutrophils (Fig. 3, bottom left panel), and the adaptive immune component subdivided into B- and T-cells (Fig. 3, bottom right panel). Of note, the abundance of all innate immune cell subtypes significantly increased in non-control groups, including PMLs and cancer groups. Also, B- and T-cells showed an increase in non-control groups but did not reach significance. A detailed breakdown of the cellular activities in these immune cells is shown in Additional file 1: Figure S4 A–B. For example, CD8+ T-cells, Tregs and CD4 memory resting/activated cells showed an increased abundance in PMLs and cancer. Sample level breakdown is visualized as a heatmap in Additional file 1: Figure S4C.

Oral PMLs show transcriptional similarities with bronchial PMLs

Given the contiguity of the oral cavity in the upper airway with the lungs in the lower airway, we next assessed the transcriptional similarities between our oral PMLs and bronchial PMLs. Beane et al. [26] identified four distinct PML bronchial subgroups, including a proliferative and an inflammatory response sub-groups, and nine gene co-expression modules recapitulating their transcriptional programs. Significantly, the gene modules corresponding to cell-cycle (proliferative sub-group), inflammatory response and IFN signaling (inflammatory response sub-group) showed enrichment in our non-control groups (Additional file 1: Figure S3), suggesting that oral lesions

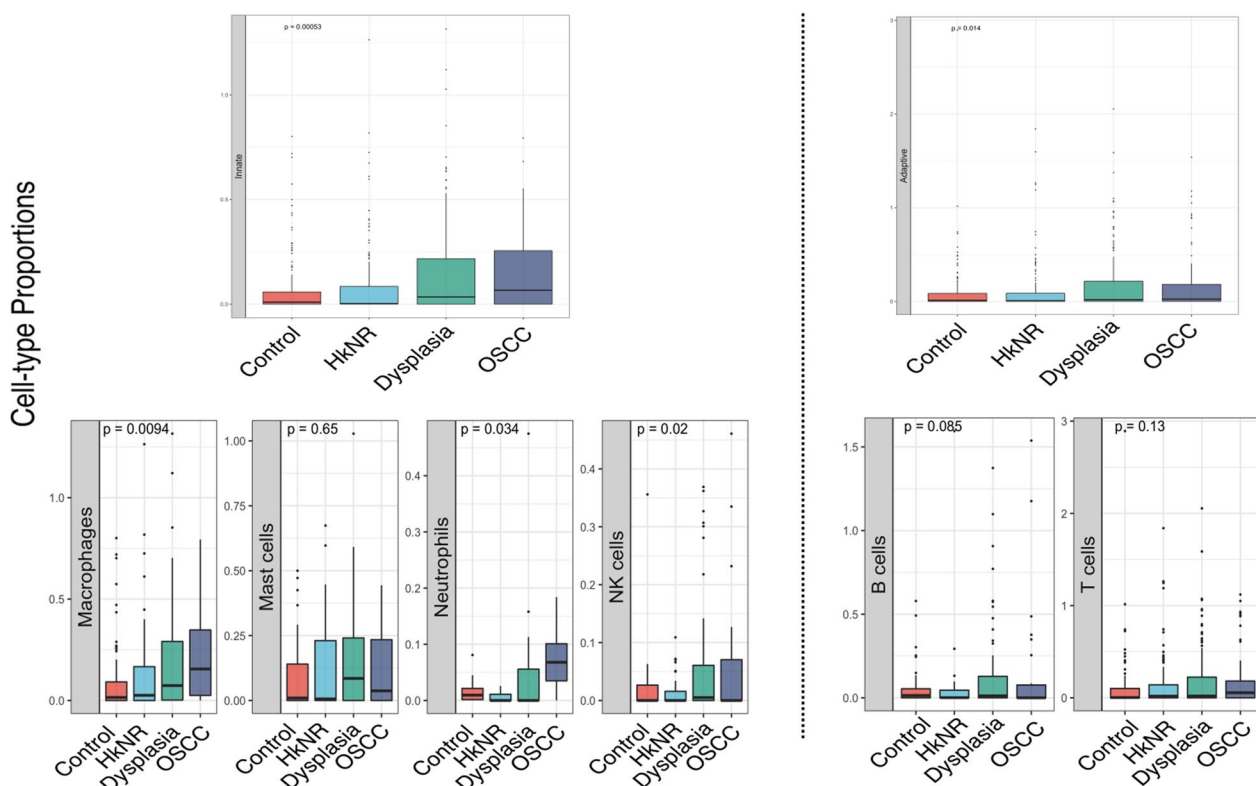


Fig. 3 Cell-type proportions Relative abundance of immune cell-types stratified by histopathology. The distribution of immune cell-type proportions inferred from CIBERSORT is plotted stratified by histopathological groups at different levels of resolution. (Top) Cell type proportions of the total innate (left) and adaptive (right) compartments. (Bottom) Specific cell sub-type proportions within the two compartments. In all plots, proportions were estimates relative to the total within each sample. P-values are obtained from ANOVA test

from the upper airway region and lesions from the main airway compartment share common processes likely to be implicated in their progression to malignancy.

Differentially abundant microbes in PMLs and OSCC are associated with pathways that promote malignancy and immune activity

The relative abundance of the microbial taxa stratified by the phylum, genus and species levels are plotted in Fig. 4A. Overall, the microbial abundance across groups varied, with the genus of *Fusobacterium* dominating the OSCC group, (Additional file 1: Figure S5A; *Fusobacterium* in green). The alpha and beta diversity stratified by the histopathological group were higher in the OSCC and PML groups compared to the healthy individuals (Additional file 1: Figure S5 B–C). To identify microbes (at the genus level) that were differentially abundant in PML and OSCC compared to the control group, we performed differential abundance analysis. With an FDR-corrected q-value ≤ 0.05 , 27 microbes were more abundant in the OSCC group than the control group, 8 in HkNR and 28 in dysplasia (Additional file 5: Table S4). Some of the microbial genera that were significantly more abundant

in both PML and OSCC groups than in the control group were *Fusobacterium*, *Shewanella*, and the fungus *Candida* (Additional file 1: Figure S5D). These have all been shown to be associated with cancer in previous studies [27–31].

To explore potential associations between differentially abundant microbial genera and the host transcriptome, we performed microbe-set enrichment analysis (MSEA) [24], whereby the differentially abundant microbes we identified were tested for enrichment against a library of ~1,300 gene-labeled microbe-sets, and the significant microbe-gene associations were visualized in Fig. 4 and tabulated in Additional file 6: Table S5. The union set of genes labeling significantly enriched microbe-sets was in turn tested for enrichment against the Hallmark compendium using hypeR (Additional file 7: Table S6) The significant genesets overlapping with those found in the host analysis in Fig. 2A are also displayed in Fig. 4 (center).

The genesets corresponding to EMT, KRAS signaling, IL6/STAT/JAK3, IL2/STAT5, IFN Gamma Response, Inflammatory Response, and TNF α via NFkB, were found to be significantly enriched in both analyses, suggesting

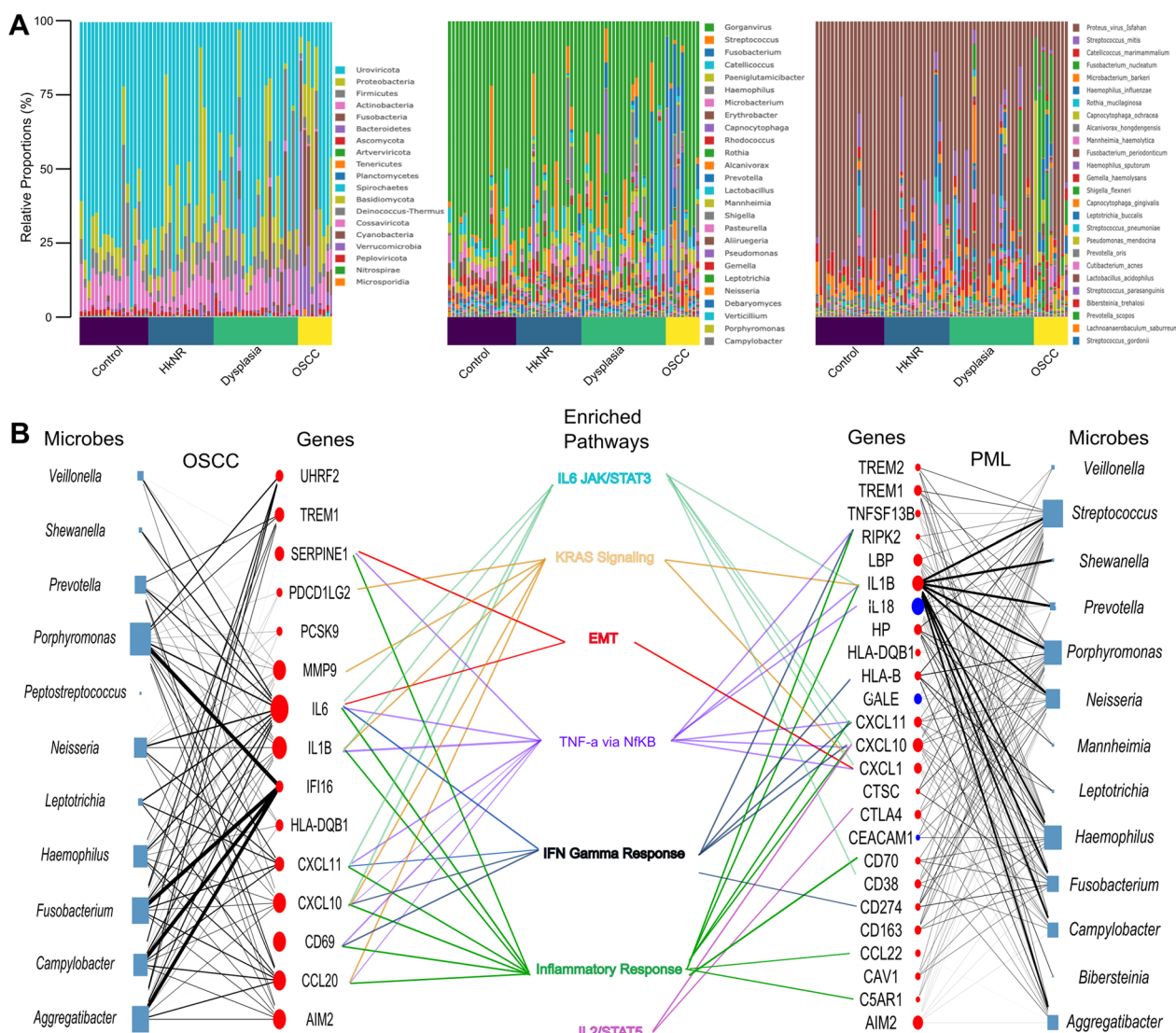


Fig. 4 Differential microbial abundance and their association with PML and OSCC. **A** Log-CPM normalized relative proportions of microbial profiles of Phylum, Genus and Species level stratified by histopathological grouping. **B** Associations of differentially abundant microbes with human genes and host-response pathways are outlined. The node size denotes the total number of associations for a microbe or a gene, and the edges' thickness is proportional to the enrichment score between a microbe and a gene. The node color indicates whether the gene was up- (red) or down-regulated (blue) in the corresponding host transcriptome differential analysis. The gene-pathway edges denote enrichment by over-representation analysis

a potential mediation by the microbiome of the association between OSCC's and PML's phenotypes and selected cancer- and immune-related pathways. For example, when comparing OSCC to control, *Veillonella*, *Aggregatibacter*, *Prevotella*, *Porphyromonas*, and *Neisseria* drove the enrichment with respect to both the SERPINE1- and IL6-indexed microbe-sets, which in turn are included in the Hallmarks' EMT geneset (Additional file 7: Table S6). Similarly, *Fusobacterium*, *Campylobacter*, *Aggregatibacter* and *Porphyromonas* drove the enrichment with respect to the IFI16-indexed microbe-set. IFI16 is an

interferon gamma inducible gene that activates the pro-inflammatory pathway IFN- γ . Differentially abundant microbes were also enriched for the microbe-sets labeled by CXCL10, CXCL11, and CCL20, among others. Similar patterns were observed for the PML group.

Discussion

OSCC are thought to arise from PMLs in the oral cavity, which have the potential to transform to cancer through a multi-step series of clinical and histopathological stages. Histologically, PMLs may represent varying

degrees of oral epithelial dysplasia, although very little is known about pre-OSCC biology [1]. Characterization of the transcriptional programs defining these premalignant stages may help identify important mechanisms of early transformation and support the design of novel interception strategies. To this end, we generated total RNA sequencing profiles from biopsy samples of 66 patients comprising healthy controls, PML lesions that included HkNR and dysplasia, as well as OSCC, with the goal of defining molecular profiles associated with dysplasia pathobiology. We performed statistical and bioinformatics analysis of the transcriptional profiles to define and annotate signatures corresponding to the distinct patient groups, and to investigate the biological pathways associated with progressive transformation. Importantly, availability of the total transcriptome allowed us to query and overlay the oral microbial gene signatures onto the host transcriptome.

Our results indicate that PMLs show considerable heterogeneity, as highlighted in Fig. 1. Of particular interest was the large heterogeneity manifested by the HkNR group based on hierarchical clustering as well as on dimensionality reduction and model-based clustering analyses (Fig. 1A–C), with subjects from that group included in all three clusters (although predominantly in the middle cluster). HkNR has been recognized as an early dysplastic lesion with significant malignant transformation potential [35], and its transcriptional heterogeneity described here suggests the need for a molecular refinement of this broad phenotype. The availability of follow-up information for a subset of the patients, and their segregation tracking with progression placement in the control-to-cancer gradient (Fig. 1B) also suggests the feasibility of developing prognostic biomarkers based on expression profiles. Still, the latter will require a larger PML patient cohort than the one available in this study.

Differential signature and pathway analyses across the different groups showed a significant early activation of cancer-associated pathways (EMT and KRAS signaling) as well as of immune and inflammatory pathways, confirming the important role played by the immune system in malignant transformation [32, 33]. Further mechanistic studies will be necessary to elucidate whether the activation of these pathways is a response to, or driver of, such transformation. It should be emphasized that, within the PML group, these enrichments were mainly found in the dysplastic samples, reflecting the higher heterogeneity of the HkNR sub-group. Also relevant was a significant enrichment of Fc gamma receptor (FCGR)-mediated regulation and FCGR phagocytosis in OSCC and dysplasia compared to control. Fc gamma receptors

are a family of heterogeneous molecules that play both activating and inhibitory tumor activity, and can bind to immunoglobulin (Ig) classes and subclasses, including those present in infected cells and pathogens [34]. Once bound, the FCGR phagocytosis process functions to eliminate the unwanted pathogenic activity [35, 36]. This may suggest that to prevent PMLs from progressing to malignancy, these inflammatory pathways are recruited to remove any of the pathogens or abnormal cells.

Another relevant finding was the enrichment in cancer and PMLs in pathways associated with extracellular matrix (ECM) organization, and activation of matrix metalloproteinases. The former has known roles in cell proliferation and migration, and the latter has a role in wound healing but adverse effects in cancer when activated by certain growth factors [37].

The oral microbiota is the second most diverse microflora in the human body after the gut. The relative abundance of several microbial genera in Additional file 1: Figure S5 is an indication of how varied the microbiota of the PML groups is when compared to the normal oral mucosa (control group), with a higher alpha diversity measure (Additional file 1: Figure S5A). Varying levels of certain microbes can be correlated with disease progression, and our results confirm this association. For instance, *Fusobacterium*—a gram-negative bacterium significantly more abundant in PML and OSCC than in control in our series (Additional file 1: Figure S5C and Additional file 3: Table S2)—has been shown to be abundant in patients diagnosed with oral cancer [27, 31]. *Fusobacterium nucleatum*, a sub-species of the *Fusobacterium* genus, has adhesion properties such that it may latch onto other bacteria and infected cells and promote an EMT-like phenotype and metastasis in some cases [28]. This role was supported by our MSEA analysis which showed a significant association with the EMT phenotype (Fig. 4B). Similarly, the genus of *Neisseria*, also more abundant in PML and OSCC than in control in our series (Additional file 1: Figure S5C and Additional file 3: Table S2), has been hypothesized to play a role in alcohol-related oral carcinogenesis. Taken together, our results support the conclusion that an increase in the activity of these microbes is associated with PML progression [30], and functional studies will be needed to decode the details of the observed associations.

In summary, our results indicate that higher levels of gene expression in PMLs and OSCCs are associated with selected oncogenic pathways, including EMT and KRAS signaling, as well as with pro- and anti-inflammatory pathways, which in turn are associated with the differential abundance of multiple microbial genera [38].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40246-023-00519-y>.

Additional file 1: Figure S1: Total number of markers per pairwise analysis. Total number of differentially expressed genes in the pairwise analysis with $\log_{2}FC \pm 1.5$ and $q\text{-val} \leq (0.01, 0.05, 0.1)$. **Figure S2:** Enrichment of cancer-associated fibroblasts. **A.** Gene expression of PDGFB1, COL1A1, COL1A2, COL3A1. **B.** Top 50 genes from fibroblast signatures shows enrichment in PML and OSCC groups. P-values are obtained by ANOVA test. **C.** Association of p-EMT and fibroblasts GSVA scores. **Figure S3:** Enrichment of lung bronchus modules. GSVA enrichment scores from modules 5, 8, and 9 from lung bronchus premalignant lesion study in our oral PML series. P-values are obtained by ANOVA test. **Figure S4:** Cell-type deconvolution scores of immune sub-types. **A.** innate, **B.** adaptive types. P-values are obtained by ANOVA test. **C.** heatmap of abundances stratified by histopathology along with smoking and progression statuses. **Figure S5:** Microbial diversity analysis. **A.** Relative abundance of microbial genera across groups with the genus of *Fusobacterium* in green and *Streptococcus* in orange. **B.** Alpha diversity stratified by histopathological groups, p-value obtained by Kruskal–Wallis test. **C.** Beta diversity stratified by histopathological groups, p-value obtained by PERMANOVA test. **D.** Relative abundance in DESeq2-normalized and \log_2 -transformed counts of the top three differentially abundant species in any of the comparisons of HkNR, Dysplasia, and OSCC with control. Q-values obtained by DESeq2-based analysis.

Additional file 2: Table ST1. Differential expression analysis results from DESeq2 for pairwise comparisons between the histopathological groups.

Additional file 3: Table ST2. Pathway enrichment analysis of differentially expressed genes pairwise using hyper enrichment analysis using Hallmark compendium.

Additional file 4: Table ST3. Pathway enrichment analysis of differentially expressed genes pairwise using hyper enrichment analysis using Reactome compendium.

Additional file 5: Table ST4. Microbial differential abundant analysis results from DESeq2 for pairwise comparisons between the histopathological groups.

Additional file 6: Table ST5. MSEA results of microbe-set associations with genes.

Additional file 7: Table ST6. Hyper enrichment analysis of microbe-set genes from MSEA on enriched pathways from host analysis.

Acknowledgements

We acknowledge the TCGA Research Network (<https://www.cancer.gov/tcga>) for granting access to cancer patient bulk expression data. We also acknowledge the BU Microarray and Sequencing Resource Core Facility for the data generation. We would like to thank David Sherr and Manish Bais for their feedback on the interpretation of the study results.

Author contributions

SM, MAK, JF, AV designed the study; SM, MAK, XV, JF, WEJ acquired funding; JF, AV collected the samples; CBN performed library prep; MMK, SM performed quality control and statistical analysis; MMK, SM, XV, MAK interpreted the data; MMK, SM prepared the manuscript; MMK, SM, XV, MAK, AV, JF, SBW edited the manuscript. All authors reviewed the manuscript.

Funding

This study was supported by NIH/NIDCR grants 5 R01 DE030350 (MAK, SM, XV), R01 DE031831 (SM), NIH/NCATS grant BU-CTSI 1UL1TR001430 (SM), Find the Cause Breast Cancer Foundation (<https://findthecausebcf.org>) (SM), ACS Research Scholar Award RSG-17-138-01-CSM (XV), Research for Health in Erie County, Inc. (JF), and R01 GM127430 and R21 AI154387 (WEJ). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. All authors have read the journal's authorship agreement and policy on disclosure of potential conflicts of interest.

Availability of data and materials

The RNA-sequencing data of the 66 samples here described are available at Gene Expression Omnibus (GEO), accession GSE227919 (provisional release date: 12/31/2023, which will be updated upon publication). The scripts used to perform the analysis can be accessed at https://github.com/montilab/pml_analysis.

Declarations

Competing interests

The authors declare no competing interests.

Author details

¹Section of Computational Biomedicine, Boston University Chobanian & Avedisian School of Medicine, 75 E. Newton St. E609, Boston, MA 02118, USA. ²Bioinformatics Graduate Program, Boston University, Boston, MA, USA. ³Department of Dentistry/Oral Oncology & Maxillofacial Prosthetics, Erie County Medical Center, Buffalo, NY, USA. ⁴Oral Medicine, Oral Oncology and Dentistry, Miami Cancer Institute, Baptist Health South Florida, Miami, FL, USA. ⁵Department of Orofacial Sciences, University of California San Francisco, San Francisco, CA, USA. ⁶Department of Translational Dental Medicine, Boston University School of Dental Medicine, 700 Albany Street, W-201, Boston, MA 02118, USA. ⁷Division of Oral Medicine and Dentistry, Brigham and Women's Hospital and Harvard University, Boston, MA, USA. ⁸Division of Infectious Disease, Center for Data Science, Rutgers New Jersey Medical School, Newark, NJ 07103, USA. ⁹Department of Biochemistry, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, USA. ¹⁰Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA.

Received: 11 April 2023 Accepted: 25 July 2023

Published online: 04 August 2023

References

- Villa A, et al. Oral keratosis of unknown significance shares genomic overlap with oral dysplasia. *Oral Dis.* 2019;25:1707–14.
- Woo S-B. Oral epithelial dysplasia and premalignancy. *Head Neck Pathol.* 2019;13:423–39.
- Stojanov IJ, Woo S-B. Malignant transformation rate of non-reactive oral hyperkeratoses suggests an early dysplastic phenotype. *Head Neck Pathol.* 2022;16:366–74.
- Warnakulasuriya S, et al. Oral potentially malignant disorders: a consensus report from an international seminar on nomenclature and classification, convened by the WHO Collaborating Centre for Oral Cancer. *Oral Dis.* 2021;27:1862–80.
- Lumerman H, Freedman P, Kerpel S. Oral epithelial dysplasia and the development of invasive squamous cell carcinoma. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 1995;79:321–9.
- Warnakulasuriya S, Johnson NW, van der Waal I. Nomenclature and classification of potentially malignant disorders of the oral mucosa. *J Oral Pathol Med.* 2007;36:575–80.
- Banerjee AG, Bhattacharyya I, Vishwanatha JK. Identification of genes and molecular pathways involved in the progression of premalignant oral epithelia. *Mol Cancer Ther.* 2005;4:865–75.
- Saintigny P, et al. Gene expression profiling predicts the development of oral cancer. *Cancer Prev Res (Phila).* 2011;4:218–29.
- Wang H, et al. Microbiomic differences in tumor and paired-normal tissue in head and neck squamous cell carcinomas. *Genome Med.* 2017;9:14.
- Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 2019;47: e47.
- Hong C, et al. PathoScope 2.0: a complete computational framework for strain identification in environmental or clinical sequencing samples. *Microbiome.* 2014;2:33.
- Zhao Y, et al. animalcules: interactive microbiome analytics and visualization in R. *Microbiome.* 2021;9:76.
- Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013;14:7.

14. Lawrence MS, et al. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*. 2015;517:576–82.
15. Sathasivam HP, et al. Gene expression changes associated with malignant transformation of oral potentially malignant disorders. *J Oral Pathol Med*. 2020. <https://doi.org/10.1111/jop.13090>.
16. Puram SV, et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell*. 2017;171:1611–1624.e24.
17. Fraley C, Raftery AE, Murphy TB, Scrucca L. mclust Version 4 for R: normal mixture modeling formodel-based clustering, classification, and density estimation. <https://stat.uw.edu/sites/default/files/files/reports/2012/tr597.pdf> (2012).
18. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
19. Federico A, Monti S. hypeR: an R package for geneset enrichment workflows. *Bioinformatics* (Oxford, England). 2020;36:1307–8.
20. Liberzon A, et al. Molecular signatures database (MSigDB) 3.0. *Bioinform* (Oxford, England). 2011;27:1739–40.
21. Liberzon A, et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1:417–25.
22. Pedersen TL. Hierarchical Sets (2021).
23. Newman AM, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. 2019;37:773–82.
24. Kou Y, Xu X, Zhu Z, Dai L, Tan Y. Microbe-set enrichment analysis facilitates functional interpretation of microbiome profiling data. *Sci Rep*. 2020;10:1–12.
25. Kartha VK, et al. PDGFR β is a novel marker of stromal activation in oral squamous cell carcinomas. *PLoS ONE*. 2016;11: e0154645.
26. Beane JE, et al. Molecular subtyping reveals immune alterations associated with progression of bronchial premalignant lesions. *Nat Commun*. 2019;10:1856.
27. Fujiwara N, et al. Involvement of fusobacterium species in oral cancer progression: a literature review including other types of cancer. *Int J Mol Sci*. 2020;21:6207.
28. Zhang S, et al. Fusobacterium nucleatum promotes epithelial-mesenchymal transition through regulation of the lncRNA MIR4435-2HG/miR-296-5p/Akt2/SNAI1 signaling pathway. *FEBS J*. 2020;287:4032–47.
29. Xu Y, Jia Y, Chen L, Gao J, Yang D. Effect of *Streptococcus anginosus* on biological response of tongue squamous cell carcinoma cells. *BMC Oral Health*. 2021;21:141.
30. Muto M, et al. Acetaldehyde production by non-pathogenic *Neisseria* in human oral microflora: implications for carcinogenesis in upper aerodigestive tract. *Int J Cancer*. 2000;88:342–50.
31. Yost S, et al. Increased virulence of the oral microbiome in oral squamous cell carcinoma revealed by metatranscriptome analyses. *Int J Oral Sci*. 2018;10:32.
32. Ferris RL. Immunology and immunotherapy of head and neck cancer. *J Clin Oncol*. 2015;33:3293–304.
33. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature*. 2017;541:321–30.
34. Cassard L, et al. Fc gamma receptors and cancer. *Springer Semin Immun*. 2006;28:321–8.
35. Feng M, et al. Phagocytosis checkpoints as new targets for cancer immunotherapy. *Nat Rev Cancer*. 2019;19:568–86.
36. Chen S, Lai SWT, Brown CE, Feng M. Harnessing and enhancing macrophage phagocytosis for cancer therapy. *Front Immunol*. 12 (2021).
37. Caley MP, Martins VLC, O'Toole EA. Metalloproteinases and wound healing. *Adv Wound Care (New Rochelle)*. 2015;4:225–34.
38. Sedghi L, DiMassa V, Harrington A, Lynch SV, Kapila YL. The oral microbiome: role of key organisms and complex networks in oral health and disease. *Periodontol*. 2021;2000(87):107–31.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

