Oral Oncology 145 (2023) 106480

Contents lists available at ScienceDirect

Oral Oncology

journal homepage: www.elsevier.com/locate/oraloncology

Detecting salivary host and microbiome RNA signature for aiding diagnosis of oral and throat cancer

Guruduth Banavar [a,*], Oyetunji Ogundijo [a], Cristina Julian [a], Ryan Toma [a], Francine Camacho [a], Pedro J. Torres [a], Lan Hu [a], Tarun Chandra [h], Andrew Piscitello [h], Liz Kenny [b], Sarju Vasani [c], Martin Batstone [d], Nevenka Dimitrova [e], Momchilo Vuyisich [a], Salomon Amar [e], Chamindie Punyadeera [f,g,*]

[a] Viome Research Institute, Viome Life Sciences Inc, New York City, NY, and Seattle, WA, USA
[b] Royal Brisbane and Women's Hospital and The School of Medicine, University of Queensland, Brisbane, QLD, Australia
[c] Department of Otolaryngology, Royal Brisbane and Women's Hospital and Faculty of Medicine, University of Queensland, Herston, QLD, Australia
[d] Oral and Maxillofacial Surgery Department, Royal Brisbane and Women's Hospital, Herston, Queensland, Australia
[e] New York Medical College, Valhalla, NY, USA
[f] The Saliva and Liquid Biopsy Translational Laboratory, Griffith Institute for Drug Discovery, Griffith University, Nathan, QLD, Australia
[g] Menzies Health Institute Queensland, Griffith University, Gold Coast, Australia
[h] EmpriQA LLC, Long Grove, IL, USA

ARTICLE INFO

Keywords:
Oral cancer
Throat cancer
RNA
Saliva
Metatranscriptomics
Diagnosis

ABSTRACT

Objective: Oral squamous cell carcinoma (OSCC) and oropharyngeal squamous cell carcinoma (OPSCC) can go undetected resulting in late detection and poor outcomes. We describe the development and validation of CancerDetect for Oral & Throat cancer™ (CDOT), to detect markers of OSCC and/or OPSCC within a high-risk population.

Material and methods: We collected saliva samples from 1,175 individuals who were 50 years or older, or adults with a tobacco use history. 945 of those were used to train a classifier using machine learning methods, resulting in a salivary microbial and human metatranscriptomic signature. The classifier was then independently validated on the 230 remaining samples prospectively collected and unseen by the classifier, consisting of 20 OSCC (all stages), 76 OPSCC (all stages), and 134 negatives (including 14 pre-malignant).

Results: On the validation cohort, the specificity of the CDOT test was 94 %, sensitivity was 90 % for participants with OSCC, and 84.2 % for participants with OPSCC. Similar classification results were observed among people in early stage (stages I & II) vs late stage (stages III & IV).

Conclusions: CDOT is a non-invasive test that can be easily administered in dentist offices, primary care centres and specialised cancer clinics for early detection of OPSCC and OSCC. This test, having received FDA's breakthrough designation for accelerated review, has the potential to enable early diagnosis, saving lives and significantly reducing healthcare expenditure.

Introduction

Oral cancer is the seventh-most common neoplasm and the ninth most common cause of cancer related death globally [1]. The American Cancer Society estimates about 54,000 new cases of oral cancer, leading to 11,230 deaths, in the United States in 2022 [2]. More than half of oral cancers in the world occur in Asia, in South/Southeast Asia, oral cancer is one of the top three cancers [3]. Oral squamous cell carcinoma (OSCC) is the most common oral cancer, accounting for 2 % of all cancers and with a high recurrence rate even with treatment [4]. Oropharyngeal squamous cell carcinoma (OPSCC), commonly known as throat cancer, is currently emerging in the developed world, and shares a similar clinical management than OSCC in primary care centres. For example, dentists perform a visual and tactile examination for oral cavity cancer,

* Corresponding authors at: Viome Life Sciences, Inc, 241 W 37th St, Suite 800, 10018 New York, NY, USA (G.B.); and Saliva and Liquid Biopsy Translational Laboratory, Griffith University, Building N75, Nathan, QLD 4111, Australia (C.P.).
E-mail addresses: guru@viome.com (G. Banavar), c.punyadeera@griffith.edu.au (C. Punyadeera).

https://doi.org/10.1016/j.oraloncology.2023.106480
Received 10 April 2023; Received in revised form 16 June 2023; Accepted 28 June 2023
Available online 14 July 2023

and in the same session, also palpate the neck, chin, tongue, and scan for lumps in the throat, which are indicative of oropharyngeal cancer. Following that exam, primary care providers will, typically, send the patient to an ENT specialist or an oral surgeon, who will do the definitive diagnostic and treatment planning. While OPSCC shares similar etiologic factors with OSCC like smoking history or alcohol consumption, OPSCC is also highly associated with HPV, which makes this cancer biologically and clinically different [5–7].

Survival rates of OPSCC and OSCC patients vary based on stage at the time of diagnosis and disease progression [8]. The five-year overall survival rate in the U.S. for OSCC is 84 %, if diagnosed in the early stages of the disease (i.e., Stage I or II). However, more than 70 % of OSCC diagnoses are not made until the disease is in stage III or IV. At these later stages, the five-year survival rate, for OSCC specifically, drops to less than 50 % [9]. Research has shown that the reasons for late diagnosis are layered and complicated, including under-utilisation of dental and primary care, and the lack and poor quality of oral cancer screening in patients that do seek general care [10]. Most importantly, in the earliest, most treatable stages, many oral cancers have little to no symptoms and may not be easily visible [11,12].

The current standard of care for oral cancer screening and diagnosis relies on a physical exam, identification of lesion(s), followed by imaging, invasive biopsy and histopathological evaluation. Biopsies will only sample a limited amount of cancer tissue and heterogeneity within the cancer is not accounted for. There are no oral cancer screening guidelines published either from the American Cancer Society, the National Comprehensive Cancer Network (NCCN), or the National Cancer Institute. The only recommendation that exists for oral cancer is in the form of a resolution passed by the American Dental Association in 2019 recommending dentists to conduct routine visual and tactile examinations for oral and oropharyngeal cancer for all patients [13], but with no objective criteria. However, only 29.4 % of adults in the United States reported ever having received a visual and tactile examination for OSCC or OPSCC [14]. Patients referred for biopsies run the risk of hematoma at the biopsy site, or worse, increased risk of metastasis if cancer cells are disseminated into the bloodstream [15]. Moreover, the standard biopsy techniques may not be appropriate for all patients, including those with conditions that preclude the safe use of local anaesthetic and those with severe bleeding diathesis or coagulopathies [14].

Even though tobacco consumption, alcohol abuse and poor oral hygiene remain the major risk factors for oral cancer, there has been increasing evidence to suggest that people who are not exposed to these risk factors are also affected. Dysbiosis in the oral microbiome leads to a chronic inflammatory state, suppresses anti-tumor immunity, and leads to the creation of novel mutagens [16]. One of the examples supporting this evidence is periodontitis, which is associated with an increased risk for cancer and poor survival in many studies [17,18]. Streptococcus, Fusobacterium, Capnocytophaga, Prevotella, among other bacteria are shown to be increased in OSCC [19–22]. Changes in microbiota have been observed in throat cancer patients as well [16]. This provides the scientific evidence to further explore microbial organisms and functions in the saliva as a means of developing a tool to evaluate oral and throat cancers.

Previously we developed a classifier for the detection of OSCC using only microbial expression on a smaller cohort [23]. In this current study, we incorporate both OSCC and OPSCC, include human gene expression in addition to microbial expression, and expand the studied cohort significantly. The resulting test, CancerDetect for Oral & Throat Cancer™ [CDOT], built using salivary metatranscriptomics and validated with an independent cohort, was granted breakthrough designation for accelerated review by the Food and Drug Administration (FDA) in April 2021. While this test can fit different intended uses, for this validation study we have only recruited people older than 50 or with a history of tobacco use, targeting people at a high risk for developing oral or throat cancer. This study was reported in format according to the Standards for Reporting Diagnostic Accuracy (STARD) 2015 statement.

Materials and methods

Test description

We have developed a simple cancer detection test as shown in Fig. 1, consisting of the following elements: (i) Sample collection/transport, (ii) Lab Processing (iii) Data Processing, and Test Report.

(i) Sample collection and transport. Unstimulated whole mouth saliva samples were collected as published previously [23]. The collection tube contains a preservative that dissolves cell membranes and penetrates all cells, denatures nucleases, and prevents RNA self-cleaving by preventing deprotonation of the 2′-OH. The use of this proprietary preservative enables ambient temperature transportation for saliva samples. Saliva sample collection, preservation, transportation and lab preparation are described in Banavar et al.

(ii) Lab processing. Our CLIA-certified lab receives the saliva samples and processes it to extract and sequence the RNA from the saliva sample. Our test extracts and sequences all mRNA molecules in a non-discriminatory fashion, after eliminating the non-informative rRNA molecules. After sample preparation is completed in the Lab, total RNA is extracted from clarified lysate using a custom silica bead-based protocol, which includes on-bead DNA removal by DNase. Total RNA is quantified using the RiboGreen method and diluted when necessary. Bacterial and human rRNAs are physically removed from the specimen using a subtractive hybridization method. The remaining RNAs are converted into Illumina directional sequencing libraries [24]. Library pools are then sequenced on Illumina NovaSeq 6000 to produce sequencing data.

(iii) Data processing. The sequenced data is processed through our bioinformatics pipeline and an OSCC/OPSCC classifier. The bioinformatics pipeline maps sequenced reads to human genes (or HG), as well as microbial species (or SP) and microbial gene clusters annotated as KEGG Orthologs (or KO) [23]. For HG detection, paired-end reads are mapped to the human transcriptome. Gene expression levels are computed by collecting the transcript-level abundance (transcript per million - TPM) and then aggregating them to the gene level using Salmon version 1.1.0 [25]. For taxonomic classification, reads are mapped to a custom catalogue derived from genomic sequences from all domains of the phylogenetic tree, namely, bacteria, archaea, eukaryota, and viruses. Taxonomies are identified and their relative activities are calculated at three different taxonomic ranks (genus, species, and strain). To identify and quantify transcriptionally active genes in the microbial community, functional assignments are obtained through alignment of the sequencing reads to another custom curated catalogue of genes and the KEGG databases [26]. Further details of the bioinformatics processing can be found in Banavar et al. [23]. All the detected molecular features (HG, SP, and KO features) are then used for downstream analyses including classifier development, validation, and eventually, classification of new samples.

The OSCC/OPSCC classifier is a machine-learning (ML) model that uses the HG, SP, and KO features and classifies the sample as belonging to the “OSCC/OPSCC class” or the “Not OSCC/OPSCC class” within prespecified performance criteria. The overall workflow of classification model development and independent validation is shown in Fig. 2. Model development is described in detail in the Supp. Table 1 and Supp. Fig. 1a and 1b. The threshold was selected during the model development and further applied during the independent validation. We performed cross validation with several hyperparameters, including the threshold. In particular, we looked at 20 evenly spaced values between 0.001 and 0.99 and the threshold that maximised the sum of specificity and specificity was considered. The final trained model coming out of the development phase was determined to be capable of inferring, at a high probability, whether a participant’s sample has OSCC and/or OPSCC or not, and used for independent validation with unseen samples, on the other side of the “firewall” in Fig. 2.

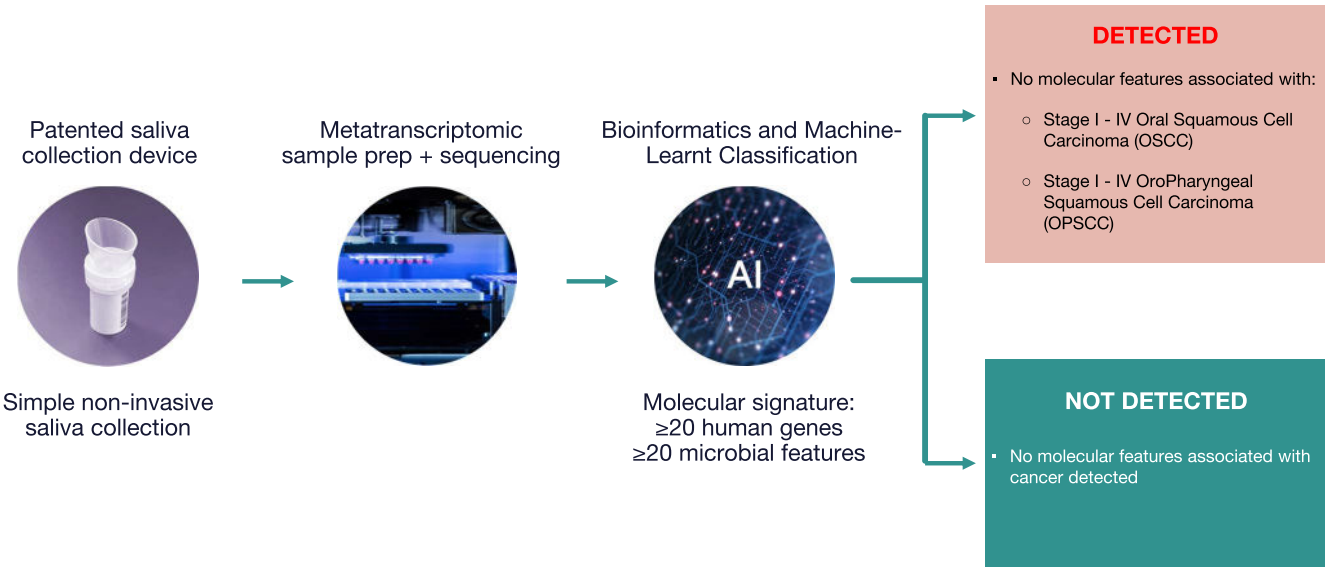

Fig. 1. Overview of salivary RNA metatranscriptomic signature based cancer detection system.

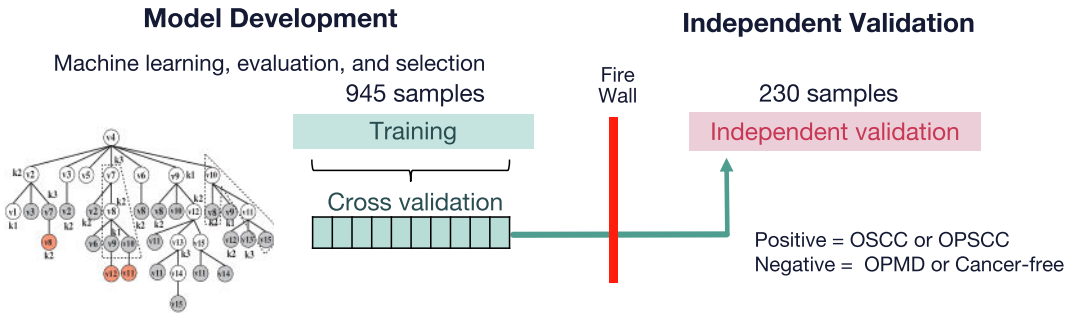

Fig. 2. Model development, cross validation and independent validation workflow.

Patient population

The CDOT test was developed with samples from a cohort of 945 individuals (details in the supporting information), and validated in an unseen independent cohort of 230 individuals in order to evaluate its performance in the proposed intended use.

Validation study participants were either 50 years or older or had a history of tobacco use. Having a "history of tobacco use" included being a current or former tobacco user per the Affordable Care Act (ACA) definition. A current tobacco user was defined as someone who uses tobacco products four or more times per week in the past six months. A former tobacco user is a person who has quit using tobacco products at the current time but had previously used tobacco products four or more times per week for six months or more, within the last 20 years. For purposes of this study, we defined tobacco broadly consistent with definitions for multiple health organisations, including the World Health Organization (WHO), which defines "tobacco use" to include smoking, sucking, chewing or snuffing any tobacco product.

The validation cohort of 230 included 101 samples prospectively collected from the Royal Brisbane Women Hospital between October 2012 and August 2019, and a combination of 129 clinically adjudicated patients with OPMD and cancer free patients from the Viome company customer database in the US (Table 1). Eligible participants had to be free from any active infection, have no cancer in the past, not be pregnant and have no irradiation to the neck and head region. The study was approved by the Queensland University of Technology and University of Queensland Medical Ethical Institutional Boards (HREC no.: 1400000617 and HREC no.: 2017000662 respectively) and the Royal Brisbane and Women's Hospital (HREC no.: HREC/12/QPAH/381) Ethics Review Board. All participants gave their consent to participate in the study.

Table 1
Independent validation cohort (n = 230). OSCC, Oral squamous cell carcinoma; OPSCC, Oropharyngeal squamous cell carcinoma; OPMD, Oral premalignant disorder.

	Positives (n = 96)		Negatives (n = 134)	
	OSCC (n = 20)	OPSCC (n = 76)	OPMD (n = 14)	Cancer free (n = 120)
Female (n)	3 (15.0 %)	4 (5.3 %)	7 (50.0 %)	42 (35.0 %)
Age (yrs) mean ± SD	61.3 ± 11.1	60.9 ± 7.6	62.9 ± 15.9	60.5 ± 8.8

Patients with OSCC or OPSCC were clinically diagnosed by the AJCC 8th edition to confirm their cancer status. Clinical data also included histopathology reports after biopsying the patients, spanning early (Stage I/II) and late (Stage III/IV) stage OSCC and OPSCC. OSCC and OPSCC diagnosis was performed with biopsy and examination of formalin-fixed paraffin-embedded (FFPE) tissue sections by routine (Hematoxylin and eosin) stain using standard methodology. Most patients with OPSCC (>97 %) were HPV positive as tested by salivary HPV-16 and published by our team [7,27–29]. Patients with oral premalignant disorders (OPMD) or cancer-free participants could be clinically adjudicated by a primary physician. OPMD included the following conditions: dysplasia, hyperplasia, leukoplakia, erythroplakia, lichenoid lesions, actinic keratosis and lichenoid reaction; as well as canker sores,

gingival enlargement as a result of a dental procedure, lichen planus, keratosis, inflammatory reaction and cheek bites. The list of pre-malignancies was suggested by dental surgeons who see patients with oral lesions in clinical practice. All patients were tested with CDOT before receiving any treatment.

Results

We evaluated the performance of the CDOT test using various metrics. Each participant sample was analysed using the classifier and the results were compared to the participant's known or assumed (cancer-free volunteers were assumed to be cancer-free) cancer status to determine the classifier's performance characteristics (Fig. 2). Specificity and sensitivity were also evaluated by disease stage (early vs late), smoking status (current, former, non-smoker and unknown) and age (<50 and ≥50).

Fig. 3a and 3b show the area under the receiver operating characteristic (ROC) curve (ROC - AUC) and the distributions of the predicted probabilities for the model on the independent validation data set. The AUC is 96 %, indicating a probability of 0.96 that our classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one (assuming 'positive' ranks higher than 'negative').

Further, the OSCC-OPSCC classifier correctly classified 18/20 = 90 % OSCC positive patients (sensitivity to OSCC), 64/76 = 84.2 % OPSCC positive patients (sensitivity to OPSCC) (Table 2a) and 126/134 = 94 % negative participants as no cancer (specificity no negative samples) (Table 2b). Out of the early stage participants with OSCC or OPSCC, the OSCC-OPSCC classifier was able to classify 9 out of the 10 as OSCC and 51 out of 62 as OPSCC positive thereby demonstrating a reasonable expectation of clinical success in identifying participants with OSCC and/or OPSCC, including those with early-stage disease (Table 2a). Also, in Table 2b, we included the breakdown of the model's specificity to negative samples.

We stratified the patient characteristics across the care centres used to source the patient samples for validation, to ensure that there was no bias. When evaluating the performance of the model by smoking status, the OSCC-OPSCC classifier correctly classified 100 % of the current smokers. Among former smokers, 7/8 = 87.5 % OSCC and 35/41 (85.4 %) OPSCC were correctly classified as positives. Among non-smokers, 4/4 = 100 % OSCC and 13/17 (76.5 %) OPSCC were correctly classified as positives (Table 2c).

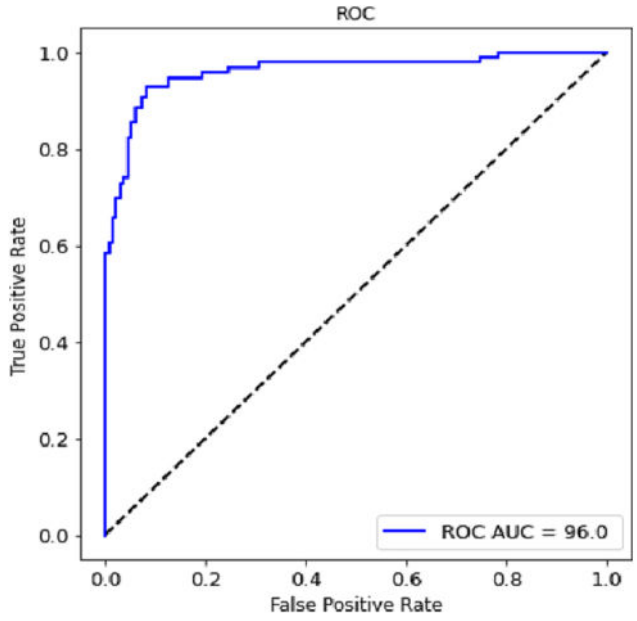

Fig. 3a. ROC plot for the model on independent validation set.

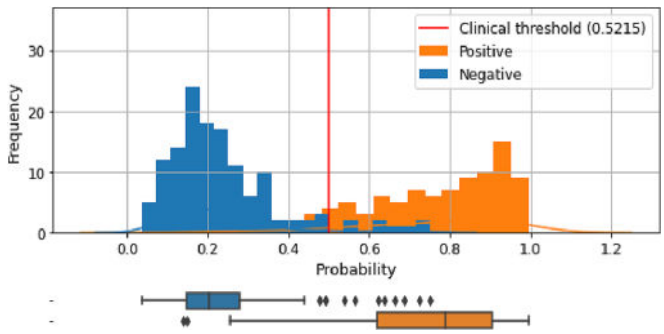

Fig. 3b. Predicted probabilities for all participants in the independent validation set.

Table 2a
Sensitivity (Positive percent agreement). OSCC, Oral squamous cell carcinoma; OPSCC, Oropharyngeal squamous cell carcinoma.

	n/N (%)
Overall sensitivity (TP/TP + FN)	82/96 (85.0 %) 95 % CI [76.7 %, 91.8 %]
OSCC sensitivity	18/20 (90.0 %) 95 % CI [68.3 %, 98.8 %]
OSCC Early stage	9/10 (90.0 %) 95 % CI [55.5 %, 99.7 %]
OSCC Late stage	9/10 (90.0 %) 95 % CI [55.5 %, 99.7 %]
OPSCC sensitivity	64/76 (84.2 %) 95 % CI [74.0 %, 91.6 %]
OPSCC Early stage	51/62 (82.3 %) 95 % CI [70.5 %, 90.8 %]
OPSCC Late stage	13/14 (92.9 %) 95 % CI [66.1 %, 99.8 %]

Table 2b
Specificity (Negative percent agreement). OPMD, Oral Premalignant Disorder.

	n/N (%)
Overall specificity (TN/TN + FP)	126/134 (94.0 %) 95 % CI [88.6 %–97.4 %]
Cancer free	116/120 (96.7 %) 95 % CI [91.7 %, 99.1 %]
OPMD	10/14 (71.4 %) 95 % CI [41.9 %, 91.6 %]

Table 2c
Sensitivity and specificity by smoking status. OSCC, Oral squamous cell carcinoma; OPSCC, Oropharyngeal squamous cell carcinoma.

	Sensitivity (n/N) %		Specificity (n/N) %
	OSCC	OPSCC	
Current	6/6 (100 %) 95 % CI [54.1 %, 100.0 %]	11/11 (100 %) 95 % CI [71.5 %, 100.0 %]	2/4 (50 %) 95 % CI [6.8 %, 93.2 %]
Former	7/8 (87.5 %) 95 % CI [47.3 %, 99.7 %]	35/41 (85.4 %) 95 % CI [70.8 %, 94.4 %]	5/7 (71.4 %) 95 % CI [29.0 %, 96.3 %]
Non-smoker	4/4 (100 %) 95 % CI [39.8 %, 100.0 %]	13/17 (76.5 %) 95 % CI [56.6 %, 96.2 %]	117/121 (97.7 %) 95 % CI [91.8 %, 99.1 %]
Unknown	1/2 (50 %) 95 % CI [1.3 %, 98.7 %]	5/7 (71.4 %) 95 % CI [29.0 %, 96.3 %]	2/2 (100 %) 95 % CI [15.8 %, 100.0 %]

When evaluating the performance of the model by age, among people below 50 years old, 4/4 = 100 % of people with OSCC and 2/3 = 66.7 % of people with OPSCC were correctly classified as positives. Among older people, 15/17 = 88.2 % OSCC and 62/73 (84.9 %) OPSCC were correctly classified as positives (Table 2d). Similarly, when stratifying the data by biological sex, we observed that the distribution of positive and negative samples across the disease state was in concordance between male and female.

An interference evaluation was also performed with 41 cancer free negative participants. Participants were required to chew gum, chew tobacco, and brush their teeth. These analyses determined whether

Table 2d
Sensitivity and specificity by age group. OSCC, Oral squamous cell carcinoma; OPSCC, Oropharyngeal squamous cell carcinoma.

	Sensitivity (n/N) %		Specificity (n/N) %
	OSCC	OPSCC	
<50 years	3/3 (100 %) 95 % CI [29.2 %, 100.0 %]	2/3 (66.7 %) 95 % CI [9.4 %, 99.2 %]	4/5 (80 %) 95 % CI [28.4 %, 99.5 %]
≥50 years	15/17 (88.2 %) 95 % CI [63.6 %, 98.5 %]	62/73 (84.9 %) 95 % CI [74.6 %, 92.2 %]	122/129 (94.5 %) 95 % CI [89.1 %, 97.8 %]

external interference factors influenced the detection power of the model. The probability output of the model did not change based on the presence of the different interfering substances, showing the robustness of the model to interferants (Supp. Fig. 2).

Discussion

This study evaluates the effectiveness of a saliva metatranscriptomic detection test, CDOT™, to identify individuals with OSCC and OPSCC. To evaluate its performance, the test result (negative or positive) was compared with the histopathological diagnosis of all significant lesions discovered during a biopsy. Based on this comparison, CDOT™ sensitivity (true positive fraction) is 90 % for participants with a histopathological diagnosis of OSCC and 84.2 % for participants with a diagnosis of OSPCC. The test can also detect true positives in early and late stages of OSCC with 90 % sensitivity. Furthermore, among participants having a valid oral cancer test result and a self-reported or clinically adjudicated cancer free status, specificity for cancer free patients not including OPMD is 96.7 %. The specificity or ability of the test to designate true negatives is 94 %. Having received FDA breakthrough for accelerated review, this test has the potential to be used for early screening in primary care settings and in secondary care centres, reducing the number of unnecessary biopsies under some scenarios, for example as a rule-in test for recurrence for patients who had oral cancer in the past.

In routine clinical practice today, the diagnostic pathway for oral cancer is dependent on the experience and expertise of different healthcare providers (dentists, dental hygienists and primary physicians) who are responsible for performing the head and neck visual examinations. Oral lesions that may be indicative of oral cancer include heterogeneous appearance such as changes in colour, texture and size; and alterations in the surface, for example, non-healing ulcerations. Several adjunct diagnostic tools are available to aid providers in identification and diagnosis, but there is no general consensus on which, if any, is most reliable. Examples are exfoliative cytology, including liquid-based, scraped and brush cytology [30,31], toluidine blue staining [32], and light-based visual detection systems [33]. The performance of these methods vary widely, with a pooled estimate of 88 % sensitivity and 81 % specificity [34], which is a lower performance when compared to the proposed saliva-based detection test.

With the advent of AI technologies, there are many new imaging methods introduced in the last decade [3510]. Fluorescent imaging is a non-invasive method supported with confocal laser endomicroscopy which has high magnification power [36] resulting in 92 % specificity, as well as the N2 laser study with 92 % specificity [37]. These laser technologies are still expensive and not readily available for primary care settings. In addition, there is significant upskilling required for thorough examination and the results might be operator dependent. Another example is the Oncogrid surveillance program [38] which uses mobile phones. While these methods are easily accessible, they may only be partially useful for certain low resource setting areas, as they have low sensitivity (around 70 %–85 %) based on limited access to certain areas of the mouth cavity [39,40].

Saliva is in direct contact with the tissues of the oral cavity and represents a biofluid that acts as a great substrate for liquid biopsy. The biomolecules detected by our metatranscriptomic method offer deep resolution and insight into the activity of the human genes as well as all the microbial species. Furthermore, our method uses advanced machine learning modelling to tease out the most distinguishing molecular features associated with OSCC and/or OPSCC. Previous methods have either assessed biomolecules from the human side (e.g. CD44 protein [41] or RNA – 6 markers that include interleukins IL-1Beta, IL-8, OAZ1SAT1S100P, and DUSP1) [42] or difference in species with 16S rRNA gene sequencing and metagenomics [43,44], but their results are less promising, and while they have reported discovery results, they have not been validated in independent cohorts. If a secondary care specialist has a suspicion of cancer, the patient will undergo a biopsy, a common invasive procedure that remains the gold standard for diagnosing premalignant and malignant oral diseases. CDOT™ is non-invasive and can be easily included in secondary care practices to confirm the need for a biopsy. Oral biopsy involves both psychological implications for the patient and technical difficulties for the health practitioner. When lesions are extensive, the most representative areas must be selected to avoid diagnostic errors. In fact, inter- and intra-observer variability of histological diagnosis for dysplasia is well documented [45].

A useful diagnostic tool should be easy to use and cause minimal patient discomfort. Ideally, a diagnostic procedure should be neither time-consuming nor complicated and, in addition to high sensitivity, should have the potential for automation. High specificity also avoids false-positives and, therefore, reduces patient anxiety, additional investigations, and even unnecessary treatment. CDOT™ provides non-invasive information regarding a patient's OSCC or OPSCC disease status that can aid in seeking a definitive diagnosis and treatment planning. It offers significant advantages over existing alternatives because of its high sensitivity and specificity, and it has the potential to identify patients for additional follow-up before their disease has progressed to be apparent in visual/tactile exams (i.e., Stages I/II).

We estimate the prevalence of oral cancer in the United States at 0.4 %. With this prevalence, the Positive Predictive Value (PPV) for our test is 5.4 %, including both OSCC and OPSCC in the intended use. The corresponding negative predictive value (NPV) for our test is 99.9 %. Also, the positive likelihood ratio (LR+) of our test is 14.31 and the negative likelihood ratio (LR-) is 0.16. Since the inverse of negative likelihood ratio (6.25) is less than LR+, based on these data, we conclude that our test will be used as a rule-in test.

An important goal of early detection of oral cancer is to shift from detection at Stage III/IV to detection at Stage I/II. We developed a cancer intercept micro-simulation model to evaluate the use of CDOT-within a large intended use population using SEER data. In one simulation, a cohort of 30-year-olds was generated, and cancers were allowed to develop up to age 50. At age 50, a one-time screen was applied, and the cohort was followed for 3 years. All detected cancers were treated with the standard of care using SEER-derived incidence rates. In this simulation, with a 5-year sojourn time (length of time the cancer remains asymptomatic), we found that the proportion of early and late stage oral cancers in the no-screening scenario was 29 % and 71 % respectively, whereas in the one-time screening scenario using CDOTt was 56 % and 44 % respectively, resulting in 27 % of late-stage cases shifted to early stages.

We recognize that there are some limitations to our study. While the model performs well in patients with OSCC and OPSCC and cancer free patients, the number of participants with pre-malignant diseases is currently too low to discriminate between positives and negatives. Similarly, the number of current and former smokers is low to make any conclusions about the model performance in these populations. The model was validated as well in young and older participants, however other populations at risk such as heavy drinkers and patients with HPV-OPSCC were not evaluated. Lastly, while the study participants were

recruited from across the US and Australia, the positive cases were recruited from a single site in Australia. A larger multi-site validation including centres in the US is forthcoming and is expected to address most of these limitations.

In summary, CDOT™ is a saliva-based detection test for oral cavity cancers and oropharyngeal cancers, with 94 % specificity and sensitivity of 90 % for OSCC and 84 % for OPSCC. The test performs RNA sequencing analysis and uses 270 human and microbial mRNA features as markers associated with oral and throat cancer. Our machine-learning based test was validated on an independent cohort of 230 patients. While future studies with a larger number of patients with pre malignancies and other clinical characteristics are needed, the current method is a practically useful, non-invasive method that has the potential to be incorporated in dentist offices, primary care centres and specialised cancer clinics for early detection of oral and throat cancers.

Funding

Chamindie Punyadeera is currently receiving funds from Cancer Australia (APP1145657), National Health and Medical Research Council (APP 2002576 and APP 2012560), the Garnett Passe and Rodney Williams Foundation, and NIH R21.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Several authors (GB, OO, CJ, RT, FC, PJT, LK, ND, LH and MV) are employees of Viome Life Sciences, Inc. GB, OO, RT, MV, CP are co-inventors on a patent for a metatranscriptomics-based saliva test for OSCC and OPSCC.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.oraloncology.2023.106480.

References

[1] Bosetti C, Carioli G, Santucci C, Bertuccio P, Gallus S, Garavello W, et al. Global trends in oral and pharyngeal cancer incidence and mortality. Int J Cancer 2020;147(4):1040–9.
[2] Key statistics for oral cavity and oropharyngeal cancers. 2022.
[3] Cheong SC, Vatanasapt P, Yi-Hsin Y, Zain RB, Kerr AR, Johnson NW. Oral cancer in South East Asia. Transl Res Oral Oncol 2017;2. https://doi.org/10.1177/2057178x17702921.
[4] Liyanage SS, Rahman B, Ridda I, Newall AT, Tabrizi SN, Garland SM, et al. The aetiological role of human papillomavirus in oesophageal squamous cell carcinoma: a meta-analysis. PLoS ONE 2013;8(7):e69238. https://doi.org/10.1371/journal.pone.0069238.
[5] Ekanayake Weeramange C, Liu Z, Hartel G, Li Y, Vasani S, Langton-Lockton J, et al. Salivary high-risk human papillomavirus (HPV) DNA as a biomarker for HPV-driven head and neck cancers. J Mol Diagnostics 2021;23(10):1334–42.
[6] Lim Y, Tang KD, Karpe AV, Beale DJ, Totsika M, Kenny L, et al. Chemoradiation therapy changes oral microbiome and metabolomic profiles in patients with oral cavity cancer and oropharyngeal cancer. Head Neck 2021;43(5):1521–34.
[7] Tang KD, Vasani S, Menezes L, Taheri T, Walsh LJ, Hughes BGM, et al. Oral HPV16 DNA as a screening tool to detect early oropharyngeal squamous cell carcinoma. Cancer Sci 2020;111(10):3854–61.
[8] Cristaldi M, Mauceri R, Fede OD, Giuliana G, Campisi G, Panzarella V. Salivary biomarkers for oral squamous cell carcinoma diagnosis and follow-up: current status and perspectives. Front Physiol 2019;10:1476. https://doi.org/10.3389/fphys.2019.01476.
[9] Peacock ZS, Pogrel MA, Schmidt BL. Exploring the reasons for delay in treatment of oral cancer. J Am Dent Assoc 2008;139:1346–52. 10.14219/jada.archive.2008.0046.
[10] Warnakulasuriya S, Kerr AR. Oral cancer screening: past, present, and future. J Dent Res 2021;100:1313–20. https://doi.org/10.1177/00220345211014795.
[11] LeHew CW, Epstein JB, Kaste LM, Choi Y. Assessing oral cancer early detection: clarifying dentists' practices. J Public Health Dent 2010;70:93–100. https://doi.org/10.1111/j.1752-7325.2009.00148.x.
[12] ADA expands policy on oral cancer detection to include oropharyngeal cancer. Oral Health Group (Oct. 3, 2019). n.d.
[13] Centers for Disease Control and Prevention. Quickstats: percentage of adults aged >18 years who have ever had an oral cancer examination, by smoking status and age group. United States: National Health Interview Survey; 2008.
[14] Chhabra N, Chhabra S, Sapra N. Diagnostic modalities for squamous cell carcinoma: an extensive review of literature-considering toluidine blue as a useful adjunct. J Maxillofac Oral Surg 2015;14:188–200. https://doi.org/10.1007/s12663-014-0660-6.
[15] Shyamala K, Girish HC, Murgod S. Risk of tumor cell seeding through biopsy and aspiration cytology. J Int Soc Prev Commun Dent 2014;4:5–11. https://doi.org/10.4103/2231-0762.129446.
[16] Vimal J, Himal I, Kannan S. Role of microbial dysbiosis in carcinogenesis & cancer therapies. Indian J Medical Res 2020;152:553–61. https://doi.org/10.4103/ijmr.ijmr_1026_18.
[17] Hajishengallis G, Liang S, Payne M, Hashim A, Jotwani R, Eskan M, et al. Low-abundance biofilm species orchestrates inflammatory periodontal disease through the commensal microbiota and complement. Cell Host Microbe 2011;10(5):497–506.
[18] Chen C, Hemme C, Beleno J, Shi ZJ, Ning D, Qin Y, et al. Oral microbiota of periodontal health and disease and their changes after nonsurgical periodontal therapy. ISME J 2018;12(5):1210–24.
[19] Mager D, Haffajee A, Devlin P, Norris C, Posner M, Goodson J. The salivary microbiota as a diagnostic indicator of oral cancer: a descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. J Transl Med 2005;3:27. https://doi.org/10.1186/1479-5876-3-27.
[20] Su S-C, Chang L-C, Huang H-D, Peng C-Y, Chuang C-Y, Chen Y-T, et al. Oral microbial dysbiosis and its performance in predicting oral cancer. Carcinogenesis 2021;42(1):127–35.
[21] Wang L, Yin G, Guo Y, Zhao Y, Zhao M, Lai Y, et al. Variations in oral microbiota composition are associated with a risk of throat cancer. Front Cell Infect Mi 2019;9. https://doi.org/10.3389/fcimb.2019.00205.
[22] Perera M, Al-hebshi NN, Perera I, Ipe D, Ulett GC, Speicher DJ, et al. Inflammatory bacteriome and oral squamous cell carcinoma. J Dent Res 2018;97(6):725–32.
[23] Banavar G, Ogundijo O, Toma R, Rajagopal S, Lim YK, Tang K, et al. The salivary metatranscriptome as an accurate diagnostic indicator of oral cancer. Npj Genom Med 2021;6(1). https://doi.org/10.1038/s41525-021-00257-x.
[24] Hatch A, Horne J, Toma R, Twibell BL, Somerville KM, Pelle B, et al. A robust metatranscriptomic technology for population-scale studies of diet, gut microbiome, and human health. Int J Genom 2019;2019:1–9.
[25] Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 2017;14:417–9. https://doi.org/10.1038/nmeth.4197.
[26] Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. Nat Biotechnol 2021;39(1):105–14.
[27] Wan Y, Vagenas D, Salazar C, Kenny L, Perry C, Calvopiña D, et al. Salivary miRNA panel to detect HPV-positive and HPV-negative head and neck cancer patients. Oncotarget 2017;8:99990–100001. 10.18632/oncotarget.21725.
[28] Tang KD, Baeten K, Kenny L, Frazer IH, Scheper G, Punyadeera C. Unlocking the potential of saliva-based test to detect HPV-16-driven oropharyngeal cancer. Cancers 2019;11:473. https://doi.org/10.3390/cancers11040473.
[29] Tang KD, Kenny L, Frazer IH, Punyadeera C. High-risk human papillomavirus detection in oropharyngeal cancers: comparison of saliva sampling methods. Head Neck 2019;41:1484–9. https://doi.org/10.1002/hed.25578.
[30] Charanya D, Raghupathy LP, Farzana AF, Murugan R, Krishnaraj R, Kalarani G. Adjunctive aids for the detection of oral premalignancy. J Pharm Bioallied Sci 2016;8:S13–9. https://doi.org/10.4103/0975-7406.191942.
[31] Alsarraf AH, Kujan O, Farah CS. The utility of oral brush cytology in the early detection of oral cancer and oral potentially malignant disorders: a systematic review. J Oral Pathol Med 2018;47:104–16. https://doi.org/10.1111/jop.12660.
[32] Pallagatti S, Sheikh S, Aggarwal A, Gupta D, Singh R, Handa R, et al. Toluidine blue staining as an adjunctive tool for early diagnosis of dysplastic changes in the oral mucosa. J Clin Exp Dent 2013:e187–91.
[33] Nagi R, Reddy-Kantharaj Y-B, Rakesh N, Janardhan-Reddy S, Sahu S. Efficacy of light based detection systems for early detection of oral cancer and oral potentially malignant disorders: systematic review. Med Oral Patol Oral Cir Bucal 2016;21:e447–55. https://doi.org/10.4317/medoral.21104.
[34] Essat M, Cooper K, Bessey A, Clowes M, Chilcott JB, Hunter KD. Diagnostic accuracy of conventional oral examination for detecting oral cavity cancer and potentially malignant disorders in patients with clinically evident oral lesions: systematic review and meta-analysis. Head Neck 2022;44:998–1013. https://doi.org/10.1002/hed.26992.
[35] García-Pola M, Pons-Fuster E, Suárez-Fernández C, Seoane-Romero J, Romero-Méndez A, López-Jornet P. Role of artificial intelligence in the early diagnosis of oral cancer. A scoping review. Cancers 2021;13:4600. https://doi.org/10.3390/cancers13184600.
[36] Aubreville M, Knipfer C, Oetter N, Jaremenko C, Rodner E, Denzler J, et al. Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning. Sci Rep-UK 2017;7(1). https://doi.org/10.1038/s41598-017-12320-8.
[37] Majumder SK, Ghosh N, Gupta PK. N2 laser excited autofluorescence spectroscopy of formalin-fixed human breast tissue. J Photochem Photobiol B: Biol 2005;81:33–42. https://doi.org/10.1016/j.jphotobiol.2005.06.002.
[38] Chigurupati R, Kuriakose M. ONCOGRID: an mHealth approach to prevention and early diagnosis of oral cancer in rural South India; 2017.
[39] Haron N, Zain RB, Nabillah WM, Saleh A, Kallarakkal TG, Ramanathan A, et al. Mobile phone imaging in low resource settings for early detection of oral cancer

and concordance with clinical oral examination. Telemed E-Health 2017;23(3):192–9.
[40] Song B, Sunny S, Uthoff RD, Patrick S, Suresh A, Kolur T, et al. Automatic classification of dual-modalilty, smartphone-based oral dysplasia and malignancy images using deep learning. Biomed Opt Express 2018;9(11):5318. https://doi.org/10.1364/boe.9.005318.
[41] Franzmann EJ, Donovan MJ. Effective early detection of oral cancer using a simple and inexpensive point of care device in oral rinses. Expert Rev Mol Diagn 2018;18:837–44. https://doi.org/10.1080/14737159.2018.1523008.
[42] Martin JL, PENA i SUBIRÀ RN. Validation of reference genes for oral cancer detection panels in a prospective blinded cohort. PLoS ONE 2016;11(7):e0158462. https://doi.org/10.1371/journal.pone.0158462.
[43] Guerrero-Preston R, Godoy-Vitorino F, Jedlicka A, Rodríguez-Hilario A, González H, Bondy J, et al. 16S rRNA amplicon sequencing identifies microbiota associated with oral cancer, human papilloma virus infection and surgical treatment. Oncotarget 2016;7(32):51320–34.
[44] Zhou X, Hao Y, Peng X, Li B, Han Q, Ren B, et al. The clinical potential of oral microbiota as a screening tool for oral squamous cell carcinomas. Front Cell Infect Mi 2021;11:728933. https://doi.org/10.3389/fcimb.2021.728933.
[45] Kujan O, Khattab A, Oliver RJ, Roberts SA, Thakker N, Sloan P. Why oral histopathology suffers inter-observer variability on grading oral epithelial dysplasia: an attempt to understand the sources of variation. Oral Oncol 2007;43:224–31. https://doi.org/10.1016/j.oraloncology.2006.03.009.