



ARTICLE

Molecular Diagnostics

Predicting the clinical outcome of oral potentially malignant disorders using transcriptomic-based molecular pathology

Hans Prakash Sathasivam^{1,2}, Ralf Kist^{1,3}, Philip Sloan^{1,4}, Peter Thomson⁵, Michael Nugent⁶, John Alexander^{1,7}, Syed Haider^{1,7} and Max Robinson^{1,4}

BACKGROUND: This study was undertaken to develop and validate a gene expression signature that characterises oral potentially malignant disorders (OPMD) with a high risk of undergoing malignant transformation.

METHODS: Patients with oral epithelial dysplasia at one hospital were selected as the 'training set' ($n = 56$) whilst those at another hospital were selected for the 'test set' ($n = 66$). RNA was extracted from formalin-fixed paraffin-embedded (FFPE) diagnostic biopsies and analysed using the NanoString nCounter platform. A targeted panel of 42 genes selected on their association with oral carcinogenesis was used to develop a prognostic gene signature. Following data normalisation, uni- and multivariable analysis, as well as prognostic modelling, were employed to develop and validate the gene signature.

RESULTS: A prognostic classifier composed of 11 genes was developed using the training set. The multivariable prognostic model was used to predict patient risk scores in the test set. The prognostic gene signature was an independent predictor of malignant transformation when assessed in the test set, with the high-risk group showing worse prognosis [Hazard ratio = 12.65, $p = 0.0003$].

CONCLUSIONS: This study demonstrates proof of principle that RNA extracted from FFPE diagnostic biopsies of OPMD, when analysed on the NanoString nCounter platform, can be used to generate a molecular classifier that stratifies the risk of malignant transformation with promising clinical utility.

British Journal of Cancer (2021) 125:413–421; <https://doi.org/10.1038/s41416-021-01411-z>

BACKGROUND

Oral squamous cell carcinoma (OSCC) has a high rate of morbidity and mortality worldwide.^{1–5} Around 30–50% of patients with OSCC die from the disease within 5 years and survival rates have not improved over many decades.^{2,5} Such adverse outcomes have mostly been attributed to late presentation of the disease, as early stage disease can be cured with effective treatment.^{1,2,6} Early detection of OSCC is feasible as they are usually preceded by clinically identifiable lesions termed 'oral potentially malignant disorders' (OPMD).^{1,2,7,8}

OPMD are defined as clinical disorders having an increased risk of developing OSCC compared to clinically 'normal' oral mucosa.^{1,7} The majority of OPMD do not transform to cancer, consequently the challenge is identifying those lesions that are most likely to undergo malignant transformation.^{9–12}

Clinical and histopathological features, though informative, are not very accurate in predicting the clinical behaviour of these lesions.¹³ Nevertheless, currently the presence and grade of oral epithelial dysplasia (OED) is considered to be the most useful indicator of malignant transformation in OPMD and provides the basis for patient stratification endorsed by the World Health

Organisation.¹ A systematic review and meta-analysis indicates that excision of oral dysplastic lesions reduces the risk of malignant transformation by ~3-fold.¹¹ Generally severe epithelial dysplasia or high-grade epithelial dysplasia is treated empirically by surgical excision;^{14–16} however, it is not clear how patient outcomes can be improved across all grades of dysplasia and those patients with non-dysplastic OPMD. Currently, it is unknown whether all OPMD should be excised or if only certain lesions benefit from a surgical intervention.

Numerous studies have assessed the prognostic ability of various biomarkers in OPMD; however, no molecular test has proved to be particularly useful in clinical practice.^{17–23} Discovering a molecular signature that is altered in OPMD and indicative of the progression to oral cancer could facilitate personalised management protocols for individual patients.

Contemporary gene expression profiling is being used to develop prognostic and predictive gene signatures in various cancers, including head and neck cancers.^{24,25} A study by Saintigny et al. (2011) proposed a gene expression-based prediction model for OPMD that showed superior prognostic accuracy when compared to models using clinico-pathological risk

¹School of Dental Sciences, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK; ²Cancer Research Centre, Institute for Medical Research, National Institutes of Health, Ministry of Health Malaysia, Setia Alam, Malaysia; ³Newcastle University Biosciences Institute, Newcastle University Centre for Cancer, Newcastle upon Tyne, UK; ⁴Department of Cellular Pathology, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK; ⁵Oral and Maxillofacial Surgery, Faculty of Dentistry, The University of Hong Kong, Hong Kong, Hong Kong SAR; ⁶Oral and Maxillofacial Surgery, Sunderland Royal Hospital, Sunderland, UK and ⁷The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK

Correspondence: Max Robinson (max.robinson@nhs.net)

These authors contributed equally: Hans Prakash Sathasivam, Ralf Kist, Syed Haider, Max Robinson

Received: 18 November 2020 Revised: 6 April 2021 Accepted: 16 April 2021

Published online: 10 May 2021

factors alone.²⁶ However, the patients in their study were enrolled in a clinical trial in which some patients received active intervention in the form of drugs that may have influenced clinical outcome and gene expression.²⁶ Furthermore, the findings of their study have yet to be validated.

Whilst formalin-fixed paraffin-embedded (FFPE) tissue is an invaluable resource linked to longitudinal disease-related outcome; it is often not possible to extract adequate amounts of high-quality nucleic acid for downstream analysis. A novel gene expression profiling system that relies on direct measurement of transcripts using colour-coded oligonucleotide probes producing molecular barcodes, the NanoString nCounter platform (NanoString Technologies, Seattle, USA), has been able to provide accurate gene expression data using RNA obtained from FFPE material.^{27,28} Recent studies have shown that mRNA expression analysis using the NanoString platform were equivalent to that achieved through quantitative real-time polymerase chain reaction (qPCR) and possibly superior to microarrays.^{27–31} Furthermore, the Prosigna™ breast cancer prognostic gene-signature assay is based on Nanostring technology and is approved by the US Food and Drug Administration and recommended by UK National Institute for Health and Care Excellence. The test is used to guide adjuvant chemotherapy decisions for women with oestrogen receptor positive, human epidermal growth factor receptor 2-negative and lymph node-negative early breast cancer.

Despite the global health burden and relatively poor prognosis associated with OSCC, a robust prognostic biomarker or prognostic model for predicting malignant transformation in OPMD has yet to be developed and validated. This study was undertaken to discover and then validate a transcriptomic-signature that identifies OPMD with a high risk of undergoing malignant transformation using FFPE-derived RNA analysed on the NanoString nCounter platform.

METHODS

Inclusion and exclusion criteria

Consecutive OPMD cases were identified from a database at Newcastle University. Cases with any one of the following characteristics were excluded: (i) patients with hereditary conditions that are linked to an increased risk of head and neck SCC (such as ataxia telangiectasia, xeroderma pigmentosum, Fanconi anaemia); (ii) history of head and neck cancer; (iii) history of radiotherapy to the head and neck region; (iv) patients that were diagnosed as having chronic hyperplastic candidosis/chronic candidosis.

OPMD were classified as having undergone malignant transformation (MT) when there was progression from an OPMD to OSCC after a period of 6 months or more from the time of initial diagnosis. Those patients with OPMD who were recorded as not having developed OSCC at their last known follow-up appointment were classified as non-transforming (NT) cases with the caveat that the patients were followed up for at least 12 months after diagnosis. All cases were assessed for high-risk human papillomavirus (HR-HPV), and positive cases were excluded from the study.

Patients

Patients were selected from a database containing patients from two different hospitals: (i) Newcastle upon Tyne Hospitals NHS Foundation Trust and; (ii) City Hospitals Sunderland NHS Foundation Trust. Patients from Newcastle Hospitals were selected as the 'training set' while patients from Sunderland Hospitals were selected for the 'test set'.

Clinico-pathological data

Demographic and clinico-pathological features as well as outcome data were recorded for all cases. The following data points were

collected and entered into a Microsoft Excel spreadsheet: (i) age at first diagnosis of OPMD; (ii) sex; (iii) clinical diagnosis of lesion; (iv) clinical outcome of OPMD; (v) date of malignant transformation or last follow-up; (vi) World Health Organization (WHO) 2017 OED grading; (vii) binary OED grading.

OED grading was performed following a modified three-tier system adapted from the work published by Speight et al.³² The cases were graded using two different classification systems: (i) WHO 2017 (mild, moderate or severe);¹ binary (low-grade or high grade).^{1,33} All data were coded, link-anonymised and stored in password protected computer files.

RNA extraction

Ten-micrometres sections were cut from the FFPE blocks and placed in 2 ml microcentrifuge tubes after discarding the first two sections. Whole sections that included both epithelium and underlying connective tissue were used. The number of sections per sample was dependent on the size of the FFPE tissue; as a guide four sections were taken for small samples (<5 mm of epithelium), three for medium samples (5–10 mm) and two for larger samples (>10 mm). If the amount of RNA extracted was not sufficient, RNA extraction was repeated using an increased number of sections. RNA extraction was performed using the RNeasy® FFPE kit (QIAGEN, Manchester, UK) according to the manufacturer's protocol. FFPE blocks were sectioned immediately before RNA extraction. The concentration and the quality of the isolated RNA were measured using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, Swindon, UK). RNA was diluted to 150 ng/μL, aliquoted and stored in a –80 °C freezer prior to NanoString assay. RNA with a 260/280 ratio of 1.7–2.3 as well as a 260/230 ratio in the range of 1.8–2.3 were considered to be of acceptable quality for downstream assays.³⁴ RNA content for all samples was normalised to 30 ng/μL, and 150 ng of total RNA per sample was used for the assay.

NanoString nCounter customised panel

A list of target genes for the NanoString nCounter Customised Panel (42 genes; 38 target and 4 housekeeping genes) was compiled based on the results of previous experiments: a whole-transcriptome analysis with total RNA sequencing (RNA-Seq),³⁵ results of previous differential gene expression work performed by our group and review of published literature. The selection of candidate genes was discussed and finalised through consensus by the authors; the gene list is shown in Supplementary Table 1. Housekeeping/internal reference genes were selected on the basis of low variation and even coverage across samples.^{36–39}

NanoString nCounter hybridisation

The NanoString nCounter platform uses hybridisation of short length probes (35- to 50-base sequence) that are subsequently fixed to a biotin-coated cartridge, which is then digitally imaged and counted to quantify mRNA expression. In-depth details regarding NanoString technology can be obtained from Geiss et al.²⁷ NanoString assay was carried out at the Newcastle NanoString Unit, Newcastle University using the nCounter MAX/FLEX system (NanoString Technologies, Seattle, Washington, USA). Each assay comes with engineered External RNA Controls Consortium (ERCC) synthetic internal negative and positive control probes. The summarised laboratory workflow for the Customised CodeSet Panel gene expression assay according to the manufacturer's protocol is outlined in the Supplementary Methods.³⁴

Normalisation of data and development of prognostic gene signature

Nanostring profiling of codeset was pre-processed using R package NanoStringNorm version 1.2.1. Data were assessed for batch effects using R package FactoMineR version 1.39. Data were

normalised using grid search over parameter space as detailed previously,⁴⁰ resulting in the choice of parameters: 'geometric mean' of positive controls, 'mean' of negative controls and 'geometric mean' of top genes, and finally log2 transformed. Genes with zero counts in >50% samples were removed from subsequent analyses. This resulted in the removal of genes: *CDKN2A*, *MMP1*, *DSPP*, *CERS1* and *IBSP*. All visualisations were generated in R statistical environment version 3.6.1.

Statistical analysis and multivariable prognostic/survival modelling
Statistical analysis and prognostic model building were performed using IBM-SPSS for Windows (version 24, IBM-SPSS Inc., Chicago, Illinois, USA) and the R statistical environment version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria). Continuous data were always assessed for normality of distribution prior to choosing appropriate statistical tests. Parametric and non-parametric tests were used for initial analysis of demographic, clinical, pathological and molecular variables. For continuous data, descriptive results were appropriately expressed as either median with interquartile range (IQR) or mean with standard deviation (SD). For cross tabulations and Chi-squared tests, exact *p*-values were calculated where possible.

The Newcastle cohort was used as the training set, while the Sunderland cohort was used as a held-out test set. mRNA abundance data for genes was transformed to z-scores. A multivariable generalised linear model with *L1*-penalty was fitted in cross-validation (four-fold) settings to identify features predictive of patient outcome. This process was repeated 100 times to select optimal *lambda* minimising cross-validation error. The final model was used to predict risk scores in the test set and predicted risk scores were dichotomised into risk groups (using median risk score from the training set). These risk groups were tested for association with patient outcome using Cox proportional hazards model. Survival modelling was performed using R packages survival version 3.1–12 and glmnet version 4.0. All visualisations were generated in R statistical environment version 3.6.1. Survival model was adjusted for age, sex, site, OED grade and type of OPMD.

All statistical tests were two-sided and results were considered statistically significant at *p* < 0.05 value unless stated otherwise.

RESULTS

Of the cases that fulfilled the selection criteria, 134 cases were considered to have sufficient tissue for RNA extraction. The majority of cases (91%, 122 of 134) yielded RNA of suitable quality and quantity for the NanoString assay. All cases were successfully analysed using this assay, and the raw data generated passed the relevant quality control parameters. The training set (*n* = 56) was comprised of 20 cases of OPMD that underwent malignant transformation (MT) and 36 cases that were non-transforming (NT). The clinico-pathological features of the training set are shown in Table 1. The test set (*n* = 66) was made up of 23 MT and 43 NT cases. The clinico-pathological features of the test set are shown in Table 2. All the OPMD in the study had oral epithelial dysplasia. Kaplan–Meier time to event analyses (time to malignant transformation) for low- and high-grade epithelial dysplasia are shown in Supplementary Fig. 1 for both the training and the test sets. An accompanying swimmer plot of the timing of individual events and censor dates are presented in Supplementary Fig. 2.

Following pre-processing and normalisation of the NanoString gene expression data (Materials and Methods), univariable prognostic association of genes in the training and test sets was assessed. Of the 33 genes, eight were prognostic in the training set (Wald *p* < 0.05; Supplementary Table 2) and five were prognostic in the test set (Wald *p* < 0.05; Supplementary Table 3). Three genes (*NOTCH1*, *CD274* and *ITGB8*) were prognostic in both sets and also demonstrated consistency in

Table 1. Clinico-pathological features of training set (*n* = 56).

| Variable | Non-transforming (NT) <i>n</i> = 36 | Malignant transformation (MT) <i>n</i> = 20 | <i>p</i> -value |
|---------------------------------------|-------------------------------------|---|---------------------|
| Age at diagnosis (years) Mean (SD) | 58.5 (14.26) | 61.3 (12.58) | 0.462 ^a |
| Time to last follow-up/MT (months) | 128.5 | 28.5 | <0.001 ^b |
| Median (IQR) | (IQR = 110) | (IQR = 48) | |
| Sex | | | |
| Male | 22 (61.1%) | 13 (65.0%) | 1 ^c |
| Female | 14 (38.9%) | 7 (35.0%) | |
| OPMD | | | |
| Leukoplakia | 29 (80.6%) | 16 (80.0%) | 0.973 ^d |
| Erythroleukoplakia | 4 (11.1%) | 2 (10.0%) | |
| Erythroplakia | 3 (8.3%) | 2 (10.0%) | |
| Site | | | |
| Tongue | 12 (33.3%) | 6 (30.0%) | 0.341 ^d |
| Floor of mouth | 16 (44.4%) | 6 (30.0%) | |
| Others | 8 (22.3%) | 8 (40.0%) | |
| WHO | | | |
| Mild | 21 (58.3%) | 8 (40.0%) | 0.015 ^d |
| Moderate | 11 (30.6%) | 3 (15.0%) | |
| Severe | 4 (11.1%) | 9 (45.0%) | |
| Binary | | | |
| Low grade | 31 (86.1%) | 8 (40.0%) | 0.001 ^d |
| High grade | 5 (13.9%) | 12 (60.0%) | |

Bold values indicate statistical significance.

SD Standard deviation, IQR Interquartile range.

^aIndependent samples *t*-test.

^bMann–Whitney U test.

^cFisher's Exact test.

^dPearson's Chi-square test.

Table 2. Clinico-pathological features of test set (*n* = 66).

| Variable | Non-transforming (NT) <i>n</i> = 43 | Malignant transformation (MT) <i>n</i> = 23 | <i>p</i> -value |
|---------------------------------------|-------------------------------------|---|---------------------|
| Age at diagnosis (years) Mean (SD) | 56.2 (12.31) | 62.2 (11.6) | 0.058 ^a |
| Time to last follow-up/MT (months) | 101 | 32 | <0.001 ^b |
| Median (IQR) | (IQR = 95) | (IQR = 47) | |
| Sex | | | |
| Male | 27 (62.8%) | 13 (56.5%) | 0.792 ^c |
| Female | 16 (37.2%) | 10 (43.5%) | |
| OPMD | | | |
| Leukoplakia | 37 (86.0%) | 17 (73.9%) | 0.053 ^d |
| Erythroleukoplakia | 6 (14.0%) | 3 (13.0%) | |
| Erythroplakia | 0 (0%) | 3 (13.0%) | |
| Site | | | |
| Tongue | 14 (32.6%) | 12 (52.2%) | 0.245 ^d |
| Floor of mouth | 14 (32.6%) | 4 (17.4%) | |
| Others | 15 (34.8%) | 7 (30.4%) | |
| WHO | | | |
| Mild | 36 (83.7%) | 9 (39.1%) | <0.001 ^d |
| Moderate | 5 (11.6%) | 6 (26.1%) | |
| Severe | 2 (4.7%) | 8 (34.8%) | |
| Binary | | | |
| Low grade | 39 (90.7%) | 12 (52.2%) | 0.001 ^d |
| High grade | 4 (9.3%) | 11 (47.8%) | |

Bold values indicate statistical significance.

SD Standard deviation, IQR Interquartile range.

^aIndependent samples *t*-test.

^bMann–Whitney U test.

^cFisher's Exact test.

^dPearson's Chi-square test.

Table 3. Characteristics of the genes in the prognostic signature, along with the estimated beta coefficients (weightage).

| HGNC gene symbol Ensembl Gene ID | Function | Role in cancer and oral carcinogenesis | Association with malignant transformation in current study (Weightage) | Key references |
|-------------------------------------|--|--|--|----------------|
| <i>CCNE1</i> ENSG00000105173 | Encodes cyclin E1, which is involved in the G1/S transition of the cell cycle. | Over-expression in breast and oesophageal cancer. | Reduced expression (−0.396966) | 51,52 |
| <i>TLX1</i> ENSG00000107807 | Encodes a nuclear transcription factor that belongs to the NK-linked or NK-like (NKL) subfamily of homeobox genes. | De-regulated expression in T-cell acute lymphoblastic leukaemia. | Reduced expression (−0.2506933) | 53 |
| <i>NOTCH1</i> ENSG00000148400 | Encodes a transmembrane receptor that regulates interactions between physically adjacent cells through binding of Notch family receptors to their cognate ligands. | De-regulated expression in several cancers. The role in oral cancer is variable: some studies suggest oncogenic and others tumour suppressive. | Reduced expression (−0.2417378) | 54–56 |
| <i>CTSL</i> ENSG00000135047 | Encodes for cathepsin L, a lysosomal cysteine proteinase that plays a major role in intracellular protein catabolism. | Over-expression in oral cancer. No association with OPMDs that undergo malignant transformation. | Reduced expression (−0.1207851) | 57,58 |
| <i>DUSP5</i> ENSG00000138166 | Encodes a phosphatase that negatively regulate members of the MAPK superfamily, which are associated with cellular proliferation and differentiation. | Putative role in oral carcinogenesis. | Reduced expression (−0.0590887) | 59–61 |
| <i>TP53</i> ENSG00000141510 | Encodes p53, which regulates cell cycle, apoptosis, senescence, DNA repair and cellular metabolism. | Inactivation is common in cancer and in oral carcinogenesis. | Reduced expression (−0.0177383) | 62,63 |
| <i>CHEK2</i> ENSG00000183765 | Encodes CHK2, a cell cycle checkpoint inhibitor that is activated in response to DNA damage. | De-regulated in breast and prostate cancer. | Reduced expression (−0.0119722) | 64–66 |
| <i>COL4A5</i> ENSG00000188153 | Encodes a subunit of type IV collagen a component of basement membranes. | Over-expression implicated in oral cancer vascular invasion. | Increased expression (0.14162271) | 67 |
| <i>ITGB8</i> ENSG00000105855 | Encodes a member of the integrin beta family of receptors and cell adhesion molecules. | De-regulated expression in hepatic, ovarian and prostate cancer. | Increased expression (0.3540842) | 68–70 |
| <i>CD274</i> ENSG00000120217 | Encodes programmed death-ligand 1 (PD-L1), an immune inhibitory receptor ligand that inhibits T-cell activation and cytokine production. | De-regulated expression in cancers, including oral cancer. Increased expression in OPMDs that undergo malignant transformation. | Increased expression (0.41203335) | 71,72 |
| <i>TP63</i> ENSG00000073282 | Encodes the p63 protein, a member of the p53 family of transcription factors. | Over-expression of ΔNp63 in OPMDs that undergo malignant transformation. | Increased expression (0.48742717) | 73–75 |

the direction of the estimated hazard ratio (Training set: *NOTCH1* HR = 0.26 and $p = 0.009$, *CD274* HR = 2.76 and $p = 0.032$, *ITGB8* HR = 3.04 and $p = 0.023$; Test set: *NOTCH1* HR = 0.19 and $p = 6.7 \times 10^{-4}$, *CD274* HR = 4.81 and $p = 0.001$, *ITGB8* HR = 5.55 and $p = 0.002$; Supplementary Fig. 3). Lower *NOTCH1* transcription and higher levels of *CD274* and *ITGB8* transcripts were associated with malignant progression.

Next, we used the training set to identify a prognostic gene signature associated with malignant transformation. A multivariable prognostic model (Cox model with *L1*-regularisation; 4-fold cross-validation) was created, which constituted 11 genes. The gene list together with the relevant weightage of each gene is shown in Table 3. This prognostic model was used to predict patient risk scores in the test set. The predicted risk scores were dichotomised into high- and low-risk groups (using median risk score derived from the training set). The risk groups demonstrated two clusters of patients in the test set when assessed against the mRNA abundance data of the underlying genes in the multivariable prognostic model (Fig. 1). These risk groups were further tested for association with patient outcome using Cox proportional hazards model adjusting for age at diagnosis, sex, site, type of OPMD and binary OED grade. The prognostic gene signature remained an independent predictor of malignant transformation when assessed in the test set, with high-risk group showing worse prognosis (hazard ratio (HR) = 12.65, $p = 0.0003$; Fig. 2a and Table 4). In the multivariable setting, in addition to the

gene-signature-derived risk scoring, binary OED grading was also statistically significant ($p = 0.017$). Predicted risk groups were also tested for association with malignant transformation using C-index, which also confirmed strong concordance between the predicted risk groups and survival times (Concordance index = 0.82, 0.75–0.88).

The predicted risk groups were verified for potential bias in the expression of the housekeeping genes (*GAPDH*, *SDHA*, *TBP*, *TUBB*), which showed stable expression levels across both groups except for a nominal difference in *TUBB* expression in the test set (log2 fold change = 0.23, $p = 0.01$, Wilcoxon rank-sum test) (Supplementary Fig. 4).

Although our predictor was trained and tested using the Nanostring nCounter platform, we tested it in an external cohort (GSE26549),²⁶ which was profiled using a microarray platform (Supplementary Methods). Our classifier accurately predicted the risk of oral cancer free survival in this independent cohort (HR = 2.38, $p = 0.041$) despite the differences arising from the RNA quantifying platform (Fig. 2b). Furthermore, we used the gene signature to explore the association with normal and malignant states in another microarray profiled dataset (GSE9844).⁴¹ We observed significantly elevated risk scores in tongue squamous cell carcinoma samples compared to normal oral mucosa samples confirming oncogenic roles of the signature genes exclusive to tumour samples ($p = 3.2 \times 10^{-5}$, Wilcoxon rank-sum test, Fig. 2c and Supplementary Methods).

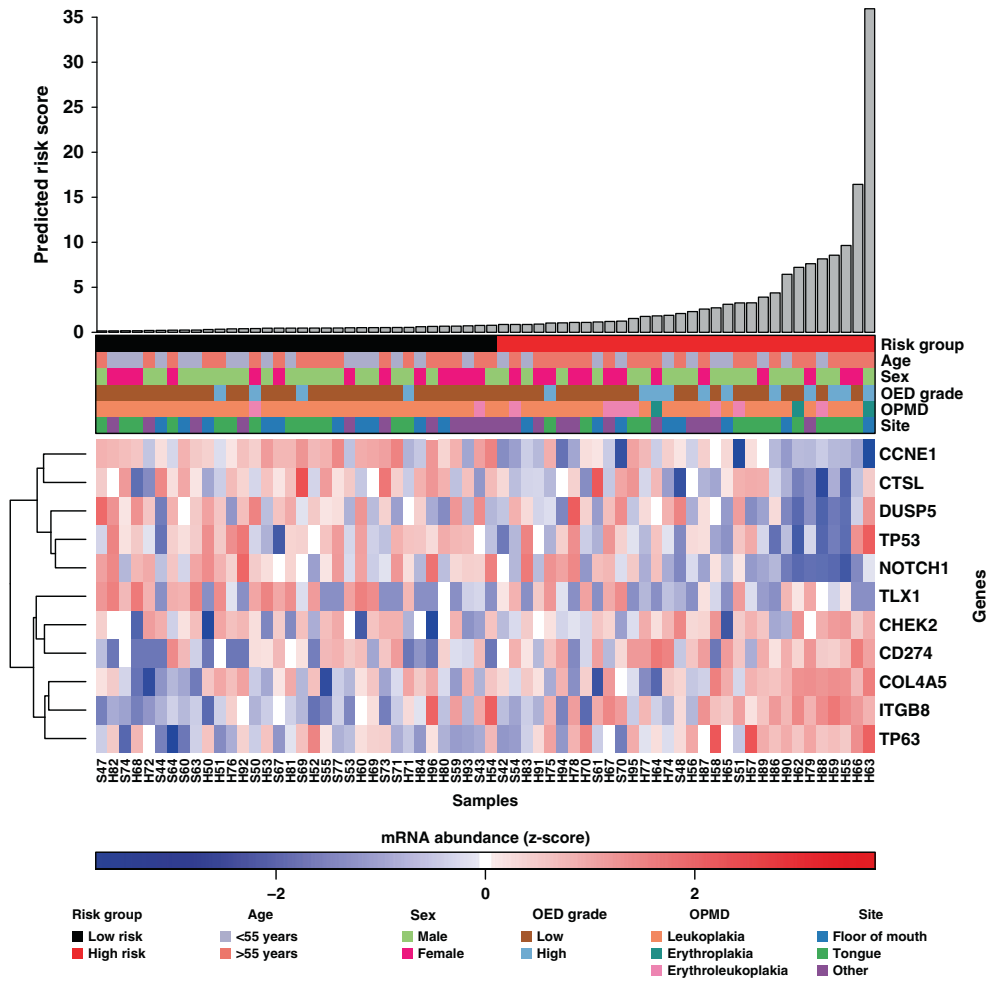


Fig. 1 Bar plot showing the predicted risk scores of the samples in the test set. The clinico-pathological covariates age at diagnosis, sex, OED grade, type of OPMD and site of the index lesion are shown in rows below the bar plot. A heatmap shows the mRNA abundance (z-score) of the genes from the prognostic gene signature for the test set samples.

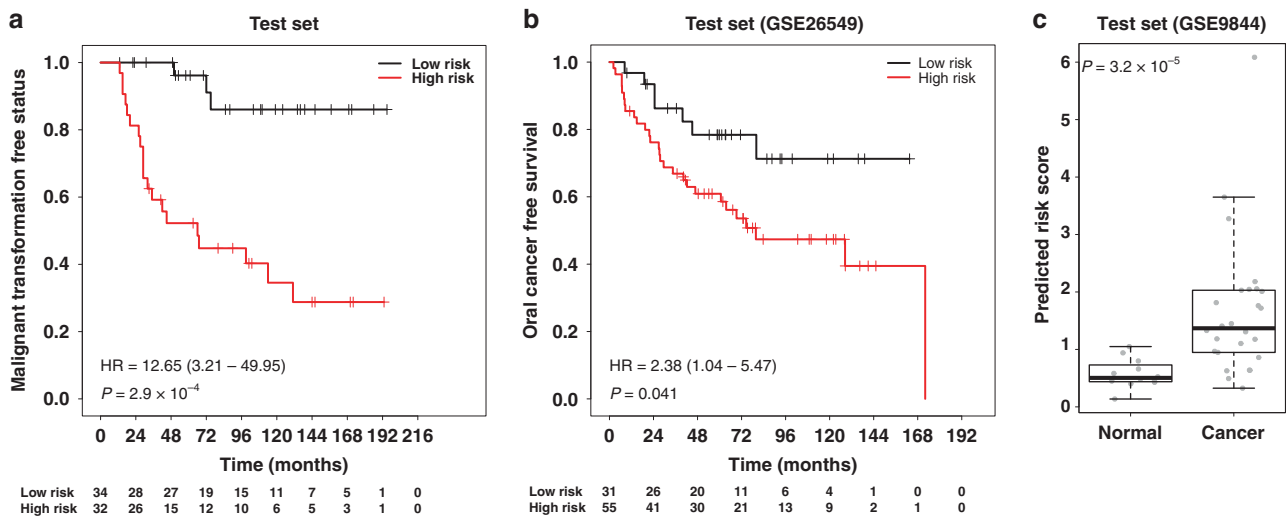


Fig. 2 Kaplan–Meier time to event analysis using Cox proportional hazards model comparing malignant transformation in the test set samples divided into low- or high-risk groups. Predicted risk scores were categorised into low- and high-risk groups using a threshold estimated as the median risk score of the training set (a). The gene expression-derived classifier was informative in an independent cohort (GSE26549 dataset)²⁶ (b) and was biologically relevant as the predicted risk scores were significantly higher in tongue squamous cell carcinoma samples compared with normal oral mucosa samples (GSE9844 dataset;⁴¹ Wilcoxon rank-sum test) (c).

Table 4. Multivariable Cox proportional hazards model (test set).

| Variable | HR | 95% CI lower | 95% CI upper | p-value |
|---|---------|--------------|--------------|---------------|
| Age at diagnosis | 1.0002 | 0.9485 | 1.055 | 0.9927 |
| Sex (male, reference = female) | 1.0227 | 0.3647 | 2.868 | 0.9659 |
| OPMD (erythroplakia, reference = leukoplakia) | 1.1140 | 0.2691 | 4.612 | 0.8816 |
| OPMD (erythroleukoplakia, reference = leukoplakia) | 0.4618 | 0.1149 | 1.856 | 0.2763 |
| Site (tongue, reference = floor of mouth) | 1.8956 | 0.4307 | 8.342 | 0.3976 |
| Site (other, reference = floor of mouth) | 0.6305 | 0.1445 | 2.750 | 0.5394 |
| OED grade (high grade, reference = low grade) | 3.4661 | 1.2498 | 9.613 | 0.0169 |
| Gene-signature based risk group (high risk, reference = low-risk) | 12.6548 | 3.2058 | 49.954 | 0.0003 |

Bold values indicate statistical significance.
HR hazard ratio, CI confidence interval.

DISCUSSION

Currently, risk-stratification of OPMD patients in clinical practice is usually based on a combination of clinical and histopathological features.^{1,23,42} However, the prognostic utility of these features has been found to be lacking and inconsistent.^{1,11,15,43} In this study, when considering clinico-pathological parameters, only OED grading was found to be statistically significant in the training set. When the clinico-pathological variables were fitted together using a Cox proportional hazards model, only the binary OED grading of cases was found to be statistically significant. This suggests that of all the clinico-pathological parameters, OED grading is the most useful prognostic indicator for malignant transformation in OPMD and supports the use of the binary grading system in clinical practice. This is consistent with the findings of most studies that have indicated that OED grading is currently the 'gold-standard' for prognosticating clinical outcome in OPMD cases.^{1,23} A confounding factor in the accurate risk assessment of the patients in this study was the lack of data on smoking habits. Smoking status is typically presented in broad categories such as current smoker, ex-smoker, never smoker; however, there are very few studies that provide detailed life-time exposure in pack-years, furthermore there is evidence that patients tend to under-report their smoking habits leading to inaccurate risk estimates.⁴⁴

Archived FFPE tissues are an invaluable resource that can be successfully used for molecular-based assays despite the degradation of nucleic acids that accompanies fixation and embedding of samples in paraffin wax.^{27,28,45-48} Our study provides evidence of the clinical utility of the NanoString nCounter platform in providing robust gene expression outputs using RNA from FFPE tissue.^{27,28,30,31,49,50} Although relatively new, the NanoString nCounter assay has been shown by several studies to be sensitive and reproducible, with sensitivity and accuracy levels that are better than microarrays and comparable to real-time quantitative PCR (qPCR).^{28,30,31} A recent study by Veldman-Jones et al (2015) that evaluated the robustness of the nCounter platform in analysing clinical samples showed that the platform has high sensitivity of target detection and good reproducibility even with low RNA amounts, making it suitable for developing clinical tests.³⁰ There are two main advantages of NanoString technology compared to conventional gene expression analysis methods such as qPCR and microarrays. In the nCounter platform, transcript levels are measured from non-amplified total RNA, unlike other platforms, thus reducing errors/biases that may be introduced through increased sample manipulation and enzymatic reactions.^{27,28} Another advantage of NanoString is that it can be multiplexed to measure up to 800 target genes in one sample, unlike qPCR-based methods that are usually only able to measure the

expression of a few genes at a time.^{27,28,30,31} These features were key to developing Prosigna™, which is a licenced prognostic test for breast cancer.

The gene signature developed in our study using the NanoString assay shows good potential in prognosticating clinical outcome. Our findings are analogous to the findings reported by Saintigny et al. (2011) where the authors showed that gene expression-based methods were superior to clinical and histological variables in determining clinical outcome in OPMD patients.²⁶ In their study, Saintigny et al. (2011) compared microarray-derived gene expression-based models against a model that contained only age, histology (dysplasia vs hyperplasia) and two biomarkers (Δ Np63 and podoplanin).²⁶ The two models containing microarray data showed much better performance compared to the model without any microarray data. Their final model, which combined the microarray data with clinical and pathological covariates, showed a slight improvement compared to the model with only microarray data. However, only nine transcripts were similar between the two microarray-based models, highlighting the rather unstable methodology employed in constructing their prognostic model. Aside from that, other major differences between their study and the current study are the type of tissue, the platform utilised to obtain the gene expression data and the statistical methodology used to arrive at the final gene expression profile.²⁶ Nevertheless, our gene classifier accurately predicted the risk of oral cancer free survival in the Saintigny dataset.²⁶ We also discovered that our gene signature was significantly different in matched normal oral mucosa samples and tongue squamous cell carcinoma.⁴¹ Together, these data suggest that the gene expression-derived classifier reported in this study is potentially generalisable and is likely to be underpinned by biologically relevant changes in oral carcinogenesis. Several novel genes (*TLX1*, *CCNE1*, *ITGB8* and *COL4A5*) with no known prior associations with oral carcinogenesis contributed to the gene signature that was developed. The characteristics of all the genes in the classifier are summarised in Table 3.⁵¹⁻⁷⁵

One major issue with prognostic/predictive models is clinical validation. For example, the molecularly driven prognostic model for malignant transformation of oral leukoplakia developed by Saintigny et al. (2011), though initially promising, has not been translated into clinical practice.²⁶ To promote translation into clinical practice, new prognostic/predictive models should be validated by an independent research team using independent patient cohorts.⁷⁶ Lack of independence between the training and test/validation cohorts can lead to an over-estimation of the prognostic ability of such models. Another barrier for successful validation of a prognostic gene signature is the presence of inter- and intra-tumour heterogeneity in OSCC, as well as heterogeneity in OPMD.

Even though the current study has demonstrated the value of a molecularly driven prognostic model over traditional risk-stratification methods for OPMD patients, molecular-based methods are not without their drawbacks. A major limitation of the current study is the sample size and the almost equal number of MT and NT cases that is not truly representative of the population where MT is variable and ranges between 0.13 and 36.4% depending on the cohort.⁹ However, this study was designed to be a proof-of-principle study to explore the possibility of using FFPE-derived material for development of a gene-signature prognostic of clinical outcome in OPMD patients. As such, we acknowledge that our study is only the first step in the development of a definitive gene expression-based prognostic model for OPMD. We also recognise that Nanostring is an expensive 'research use only' assay, nevertheless, it is conceivable that development of a clinical test would reduce costs by economy of scale. Prosigna™ a Nanostring-based breast cancer test, is proof that the technology can be translated into a cost-effective clinical test.

Although our study has successfully shown that the prognostic model developed is superior to conventional risk-stratification methods in a test set, the patients were obtained in a retrospective manner and the number of samples was small. Future studies require external validation in a sufficiently powered, prospective cohort study, recruiting consecutive patients with OPMD or as an observational component in a clinical trial. Ideally, such studies should be large enough to allow for data to be analysed by dysplasia grade, since this would provide valuable insight into the strengths and limitations of the molecular classifier against the current gold standard for risk assessment.

CONCLUSIONS

We have shown proof of principle that RNA extracted from FFPE tissue, when analysed on the NanoString nCounter platform, can be used to model a gene expression signature that accurately predicts the risk of oral potentially malignant disorders undergoing malignant transformation. The molecular classifier was developed on a training set and validated on a test set, but still requires external validation in an appropriately powered cohort study before it can be used in clinical practice.

ACKNOWLEDGEMENTS

We would like to acknowledge Ms. Anastasia Resteu from the Newcastle NanoString Unit, Newcastle University, for assisting with the NanoString assay.

AUTHOR CONTRIBUTIONS

H.P.S. conceived and designed the study, performed experiments, analysed data, carried out statistical analysis and wrote major parts of the manuscript. R.K. participated in study design, provided expert advice, performed experiments, interpreted results and edited the manuscript. P.S. participated in study design, provided expert advice, performed dysplasia grading, contributed to retrieval of tissue, interpreted results and edited the manuscript. P.T. and M.N. contributed to case selection, retrieval of follow-up data, quality assurance of data, interpreted results and edited the manuscript. J.A. and S.H. contributed to study design, analysed data, carried out statistical analysis, quality assurance of results, interpreted results and wrote major parts of the manuscript. M.R. conceived and designed the study, provided expert advice, performed dysplasia grading, contributed to retrieval of tissue, analysed data, interpreted results and wrote major parts of the manuscript. All authors read and approved the final manuscript.

ADDITIONAL INFORMATION

Ethics approval and consent to participate The study was performed with approval from a Health Research Authority (UK) Research Ethics Committee (North East—Tyne and Wear South Research Ethics Committee; NRES Committee North

East—Sunderland 11/NE/0118) and complies with UK legislation and guidelines. Patients were not recruited to the study and therefore individual patient consent was not sought. Link-anonymised patient tissue samples, surplus to diagnostic requirements, were analysed in accordance with the terms of the ethical approval. The study was performed in accordance with the Declaration of Helsinki.

Consent for publication Not applicable.

Data availability The datasets used and/or analysed during the current study are available at <https://doi.org/10.5281/zenodo.4643470>.

Competing interests The authors declare no competing interests.

Funding information H.P. Sathasivam was supported by scholarship from Ministry of Health Malaysia. S.H. is supported by Breast Cancer Now as part of Programme Funding to The Breast Cancer Now Toby Robins Research Centre.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41416-021-01411-z>.

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

REFERENCES

1. El-Naggar, A. K., Chan, J. K. C., Grandis, J. R., Takata, T., Slootweg, P. J., editors *WHO Classification of Head and Neck Tumours* (International Agency for Research on Cancer, 2017).
2. Barnes, L., Eveson, J. W., Reichart, P., Sidransky, D., editors *Pathology and Genetics of Head and Neck Tumours* (World Health Organization, IARC, 2005).
3. Moore, S., Johnson, N., Pierce, A. & Wilson, D. The epidemiology of lip cancer: a review of global incidence and aetiology. *Oral Dis.* **5**, 185–195 (1999).
4. Moore, S. R., Johnson, N. W., Pierce, A. M. & Wilson, D. F. The epidemiology of mouth cancer: a review of global incidence. *Oral Dis.* **6**, 65–74 (2000).
5. Warnakulasuriya, S. Global epidemiology of oral and oropharyngeal cancer. *Oral Oncol.* **45**, 309–316 (2009).
6. Goodson, M. L. & Thomson, P. J. Management of oral carcinoma: benefits of early precancerous intervention. *Br. J. Oral Maxillofac. Surg.* **49**, 88–91 (2011).
7. van der Waal, I. Potentially malignant disorders of the oral and oropharyngeal mucosa; terminology, classification and present concepts of management. *Oral Oncol.* **45**, 317–323 (2009).
8. Warnakulasuriya, S., Johnson, N. W. & van der Waal, I. Nomenclature and classification of potentially malignant disorders of the oral mucosa. *J. Oral Pathol. Med.* **36**, 575–580 (2007).
9. Warnakulasuriya, S. & Ariyawardana, A. Malignant transformation of oral leukoplakia: a systematic review of observational studies. *J. Oral Pathol. Med.* **45**, 155–166 (2016).
10. Shariff, J. A. & Zavras, A. I. Malignant transformation rate in patients presenting oral epithelial dysplasia: systematic review and meta-analysis. *J. Oral Dis.* **2015**, 1–10 (2015).
11. Mehanna, H. M., Rattay, T., Smith, J. & McConkey, C. C. Treatment and follow-up of oral dysplasia—a systematic review and meta-analysis. *Head Neck* **31**, 1600–1609 (2009).
12. Thomson, P. J., Goodson, M. L. & Smith, D. R. Profiling cancer risk in oral potentially malignant disorders—a patient cohort study. *J. Oral Pathol. Med.* **46**, 888–895 (2017).
13. Napier, S. S. & Speight, P. M. Natural history of potentially malignant oral lesions and conditions: an overview of the literature. *J. Oral Pathol. Med.* **37**, 1–10 (2008).
14. Field, E. A., McCarthy, C. E., Ho, M. W., Rajlawat, B. P., Holt, D., Rogers, S. N. et al. The management of oral epithelial dysplasia: the Liverpool algorithm. *Oral Oncol.* **51**, 883–887 (2015).
15. Villa, A. & Woo, S. B. Leukoplakia—a diagnostic and management algorithm. *J. Oral Maxillofac. Surg.* **75**, 723–734 (2017).
16. Awadallah, M., Idle, M., Patel, K. & Kademani, D. Management update of potentially premalignant oral epithelial lesions. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **125**, 628–636 (2018).
17. Reibel, J. Prognosis of oral pre-malignant lesions: significance of clinical, histopathological, and molecular biological characteristics. *Crit. Rev. Oral Biol. Med.* **14**, 47–62 (2003).

18. Pitiyage, G., Tilakaratne, W. M., Tavassoli, M. & Warnakulasuriya, S. Molecular markers in oral epithelial dysplasia: review. *J. Oral Pathol. Med.* **38**, 737–752 (2009).
19. Nankivell, P. & Mehanna, H. Oral dysplasia: biomarkers, treatment, and follow-up. *Curr. Oncol. Rep.* **13**, 145–152 (2011).
20. Smith, J., Rattay, T., McConkey, C., Helliwell, T. & Mehanna, H. Biomarkers in dysplasia of the oral cavity: a systematic review. *Oral Oncol.* **45**, 647–653 (2009).
21. Lingen, M. W., Pinto, A., Mendes, R. A., Franchini, R., Czerninski, R., Tilakaratne, W. M. et al. Genetics/epigenetics of oral premalignancy: current status and future research. *Oral Dis.* **17**, 7–22 (2011).
22. Nikitakis, N. G., Pentenero, M., Georgaki, M., Poh, C. F., Peterson, D. E., Edwards, P. et al. Molecular markers associated with development and progression of potentially premalignant oral epithelial lesions: Current knowledge and future implications. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **125**, 650–669 (2018).
23. Speight, P. M., Khurram, S. A. & Kujan, O. Oral potentially malignant disorders: risk of progression to malignancy. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **125**, 612–627 (2018).
24. De Cecco, L., Nicolau, M., Giannoccaro, M., Daidone, M. G., Bossi, P., Locati, L. et al. Head and neck cancer subtypes with biological and clinical relevance: meta-analysis of gene-expression data. *Oncotarget* **6**, 9627–9642 (2015).
25. Tonella, L., Giannoccaro, M., Alfieri, S., Canevari, S. & De Cecco, L. Gene expression signatures for head and neck cancer patient stratification: are results ready for clinical application? *Curr. Treat. Options Oncol.* **18**, 32 (2017).
26. Saintigny, P., Zhang, L., Fan, Y. H., El-Naggar, A. K., Papadimitrakopoulou, V. A., Feng, L. et al. Gene expression profiling predicts the development of oral cancer. *Cancer Prev. Res. (Philos.)* **4**, 218–229 (2011).
27. Geiss, G. K., Bumgarner, R. E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D. L. et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26**, 317–325 (2008).
28. Reis, P. P., Waldron, L., Goswami, R. S., Xu, W., Xuan, Y., Perez-Ordóñez, B. et al. mRNA transcript quantification in archival samples using multiplexed, color-coded probes. *BMC Biotechnol.* **11**, 46 (2011).
29. Balko, J. M., Cook, R. S., Vaught, D. B., Kuba, M. G., Miller, T. W., Bholra, N. E. et al. Profiling of residual breast cancers after neoadjuvant chemotherapy identifies DUSP4 deficiency as a mechanism of drug resistance. *Nat. Med.* **18**, 1052–1059 (2012).
30. Veldman-Jones, M. H., Brant, R., Rooney, C., Geh, C., Emery, H., Harbron, C. G. et al. Evaluating robustness and sensitivity of the NanoString technologies nCounter platform to enable multiplexed gene expression analysis of clinical samples. *Cancer Res.* **75**, 2587–2593 (2015).
31. Veldman-Jones, M. H., Lai, Z., Wappett, M., Harbron, C. G., Barrett, J. C., Harrington, E. A. et al. Reproducible, quantitative, and flexible molecular subtyping of clinical DLBCL samples using the NanoString nCounter system. *Clin. Cancer Res.* **21**, 2367–2378 (2015).
32. Speight, P. M., Abram, T. J., Floriano, P. N., James, R., Vick, J., Thornhill, M. H. et al. Interobserver agreement in dysplasia grading: toward an enhanced gold standard for clinical pathology trials. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **120**, 474–482 e2 (2015).
33. Kujan, O., Oliver, R. J., Khattab, A., Roberts, S. A., Thakker, N. & Sloan, P. Evaluation of a new binary system of grading oral epithelial dysplasia for prediction of malignant transformation. *Oral Oncol.* **42**, 987–993 (2006).
34. NanoString. *nCounter XT Assay User Manual* (NanoString Technologies Inc., 2016).
35. Sathasivam, H. P., Casement, J., Bates, T., Sloan, P., Thomson, P., Robinson, M., et al. Gene expression changes associated with malignant transformation of oral potentially malignant disorders. *J. Oral Pathol. Med.* <https://doi.org/10.1111/jop.13090> (2020).
36. Lohavanichbutr, P., Méndez, E., Holsinger, F. C., Rue, T. C., Zhang, Y., Houck, J. et al. A 13-gene signature prognostic of HPV-negative OSCC: discovery and external validation. *Clin. Cancer Res.* **19**, 1197 (2013).
37. Lallemand, B., Evrard, A., Combescure, C., Chapuis, H., Chambon, G., Raynal, C. et al. Reference gene selection for head and neck squamous cell carcinoma gene expression studies. *BMC Mol. Biol.* **10**, 78 (2009).
38. Taihi, I., Nassif, A., Berbar, T., Isaac, J., Berald, A., Gogly, B. et al. Validation of housekeeping genes to study human gingival stem cells and their in vitro osteogenic differentiation using real-time RT-qPCR. *Stem Cells Int.* **2016**, 6261490 (2016).
39. Rentoft, M., Hultin, S., Coates, P. J., Laurell, G. & Nylander, K. Tubulin α -6 chain is a stably expressed reference gene in archival samples of normal oral tissue and oral squamous cell carcinoma. *Exp. Ther. Med.* **1**, 419–423 (2010).
40. Haider, S., Yao, C. Q., Sabine, V. S., Grzadzowski, M., Stimper, V., Starmans, M. H. W. et al. Pathway-based subnetworks enable cross-disease biomarker discovery. *Nat. Commun.* **9**, 4746 (2018).
41. Ye, H., Yu, T., Temam, S., Ziober, B. L., Wang, J., Schwartz, J. L. et al. Transcriptomic dissection of tongue squamous cell carcinoma. *BMC Genomics* **9**, 69 (2008).
42. van der Waal, I. Potentially malignant disorders of the oral and oropharyngeal mucosa; present concepts of management. *Oral Oncol.* **46**, 423–425 (2010).
43. Dost, F., Le Cao, K., Ford, P. J., Ades, C. & Farah, C. S. Malignant transformation of oral epithelial dysplasia: a real-world evaluation of histopathologic grading. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **117**, 343–352 (2014).
44. Hald, J., Overgaard, J. & Grau, C. Evaluation of objective measures of smoking status—a prospective clinical study in a group of head and neck cancer patients treated with radiotherapy. *Acta Oncol.* **42**, 154–159 (2003).
45. von Ahlfen, S., Missel, A., Bendrat, K. & Schlumpberger, M. Determinants of RNA quality from FFPE samples. *PLoS ONE* **2**, e1261 (2007).
46. Eikrem, O., Beisland, C., Hjelte, K., Flatberg, A., Scherer, A., Landolt, L. et al. Transcriptome sequencing (RNAseq) enables utilization of formalin-fixed, paraffin-embedded biopsies with clear cell renal cell carcinoma for exploration of disease biology and biomarker development. *PLoS ONE* **11**, e0149743 (2016).
47. Mittempergher, L., de Ronde, J. J., Nieuwland, M., Kerkhoven, R. M., Simon, I., Rutgers, E. J. et al. Gene expression profiles from formalin fixed paraffin embedded breast cancer tissue are largely comparable to fresh frozen matched tissue. *PLoS ONE* **6**, e17163 (2011).
48. Wimmer, I., Troscher, A. R., Brunner, F., Rubino, S. J., Bien, C. G., Weiner, H. L. et al. Systematic evaluation of RNA quality, microarray data reliability and pathway analysis in fresh, fresh frozen and formalin-fixed paraffin-embedded tissue samples. *Sci. Rep.* **8**, 6351 (2018).
49. Scott, D. W., Chan, F. C., Hong, F., Rogic, S., Tan, K. L., Meissner, B. et al. Gene expression-based model using formalin-fixed paraffin-embedded biopsies predicts overall survival in advanced-stage classical Hodgkin lymphoma. *J. Clin. Oncol.* **31**, 692–700 (2013).
50. Saba, N. F., Wilson, M., Doho, G., DaSilva, J., Benjamin Isett, R., Newman, S. et al. Mutation and transcriptional profiling of formalin-fixed paraffin embedded specimens as companion methods to immunohistochemistry for determining therapeutic targets in oropharyngeal squamous cell carcinoma (OPSCC): a pilot of proof of principle. *Head Neck Pathol.* **9**, 223–235 (2015).
51. Scaltriti, M., Eichhorn, P. J., Cortes, J., Prudkin, L., Aura, C., Jimenez, J. et al. Cyclin E amplification/overexpression is a mechanism of trastuzumab resistance in HER2+ breast cancer patients. *Proc. Natl Acad. Sci. USA* **108**, 3761–3766 (2011).
52. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576–582 (2015).
53. Hatano, M., Roberts, C. W., Minden, M., Crist, W. M. & Korsmeyer, S. J. Deregulation of a homeobox gene, HOX11, by the t(10;14) in T cell leukemia. *Science* **253**, 79–82 (1991).
54. Yap, L. F., Lee, D., Khairuddin, A., Pairan, M. F., Puspita, B., Siar, C. H. et al. The opposing roles of NOTCH signalling in head and neck cancer: a mini review. *Oral Dis.* **21**, 850–857 (2015).
55. Ranganathan, P., Weaver, K. L. & Capobianco, A. J. Notch signalling in solid tumours: a little bit of everything but not all the time. *Nat. Rev. Cancer* **11**, 338–351 (2011).
56. Agrawal, N., Frederick, M. J., Pickering, C. R., Bettegowda, C., Chang, K., Li, R. J. et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* **333**, 1154–1157 (2011).
57. Alevizos, I., Mahadevappa, M., Zhang, X., Ohyama, H., Kohno, Y., Posner, M. et al. Oral cancer in vivo gene expression profiling assisted by laser capture microdissection and microarray analysis. *Oncogene* **20**, 6196–6204 (2001).
58. Macabeo-Ong, M., Shiboski, C. H., Silverman, S., Ginzinger, D. G., Dekker, N., Wong, D. T. et al. Quantitative analysis of cathepsin L mRNA and protein expression during oral cancer progression. *Oral Oncol.* **39**, 638–647 (2003).
59. Kang, C.-J., Chen, Y.-J., Liao, C.-T., Wang, H.-M., Chang, J. T., Lin, C.-Y. et al. Transcriptome profiling and network pathway analysis of genes associated with invasive phenotype in oral cancer. *Cancer Lett.* **284**, 131–140 (2009).
60. Mallery, S. R., Zwick, J. C., Pei, P., Tong, M., Larsen, P. E., Shumway, B. S. et al. Topical application of a bioadhesive black raspberry gel modulates gene expression and reduces cyclooxygenase 2 protein in human premalignant oral lesions. *Cancer Res.* **68**, 4945 (2008).
61. Kuriakose, M. A., Chen, W. T., He, Z. M., Sikora, A. G., Zhang, P., Zhang, Z. Y. et al. Selection and validation of differentially expressed genes in head and neck cancer. *Cell. Mol. Life Sci.* **61**, 1372–1383 (2004).
62. Leemans, C. R., Braakhuis, B. J. & Brakenhoff, R. H. The molecular biology of head and neck cancer. *Nat. Rev. Cancer* **11**, 9–22 (2011).
63. Brennan, J. A., Boyle, J. O., Koch, W. M., Goodman, S. N., Hruban, R. H., Eby, Y. J. et al. Association between cigarette smoking and mutation of the p53 gene in squamous-cell carcinoma of the head and neck. *N. Engl. J. Med.* **332**, 712–717 (1995).
64. Apostolou, P. & Papatotiriou, I. Current perspectives on CHEK2 mutations in breast cancer. *Breast Cancer (Dove Med. Press)* **9**, 331–335 (2017).

65. Cybulski, C., Górski, B., Huzarski, T., Masojć, B., Mierzejewski, M., Debniak, T. et al. CHEK2 is a multiorgan cancer susceptibility gene. *Am. J. Hum. Genet.* **75**, 1131–1135 (2004).
66. Cybulski, C., Huzarski, T., Górski, B., Masojć, B., Mierzejewski, M., Debniak, T. et al. A novel founder CHEK2 mutation is associated with increased prostate cancer risk. *Cancer Res.* **64**, 2677–2679 (2004).
67. Chen, Y.-L., Tsai, W.-H., Ko, Y.-C., Lai, T.-Y., Cheng, A.-J., Shiah, S.-G. et al. Discoidin domain receptor-1 (DDR1) is involved in angiolymphatic invasion in oral cancer. *Cancers (Basel)* **12**, 841 (2020).
68. Wang, W. W., Wang, Y. B., Wang, D. Q., Lin, Z. & Sun, R. J. Integrin beta-8 (ITGB8) silencing reverses gefitinib resistance of human hepatic cancer HepG2/G cell line. *Int. J. Clin. Exp. Med.* **8**, 3063–3071 (2015).
69. Cui, Y., Wu, F., Tian, D., Wang, T., Lu, T., Huang, X. et al. miR-199a-3p enhances cisplatin sensitivity of ovarian cancer cells by targeting ITGB8. *Oncol. Rep.* **39**, 1649–1657 (2018).
70. Mertens-Walker, I., Ferdinani, B. C., Maharaj, M. S., Rockstroh, A., Nelson, C. C., Herington, A. C. et al. The tumour-promoting receptor tyrosine kinase, EphB4, regulates expression of integrin-beta8 in prostate cancer cells. *BMC Cancer* **15**, 164 (2015).
71. Lenouvel, D., González-Moles, M. Á., Talbaoui, A., Ramos-García, P., González-Ruiz, L., Ruiz-Ávila, I. et al. An update of knowledge on PD-L1 in head and neck cancers: physiologic, prognostic and therapeutic perspectives. *Oral Dis.* **26**, 511–526 (2020).
72. Dave, K., Ali, A. & Magalhaes, M. Increased expression of PD-1 and PD-L1 in oral lesions progressing to oral squamous cell carcinoma: a pilot study. *Sci. Rep.* **10**, 9705 (2020).
73. Saintigny, P., El-Naggar, A. K., Papadimitrakopoulou, V., Ren, H., Fan, Y. H., Feng, L. et al. DeltaNp63 overexpression, alone and in combination with other biomarkers, predicts the development of oral cancer in patients with leukoplakia. *Clin. Cancer Res.* **15**, 6284–6291 (2009).
74. Matsubara, R., Kawano, S., Kiyosue, T., Goto, Y., Hirano, M., Jinno, T. et al. Increased DeltaNp63 expression is predictive of malignant transformation in oral epithelial dysplasia and poor prognosis in oral squamous cell carcinoma. *Int. J. Oncol.* **39**, 1391–1399 (2011).
75. Chen, Y. K., Hsue, S. S. & Lin, L. M. Expression of p63 protein and mRNA in oral epithelial dysplasia. *J. Oral Pathol. Med.* **34**, 232–239 (2005).
76. Chibon, F. Cancer gene expression signatures—the rise and fall? *Eur. J. Cancer* **49**, 2000–2009 (2013).