Received: 13 June 2020 | Revised: 15 June 2020 | Accepted: 16 June 2020

DOI: 10.1111/jop.13090

ORIGINAL ARTICLE

Journal of Oral Pathology & Medicine WILEY

Gene expression changes associated with malignant transformation of oral potentially malignant disorders

Hans P. Sathasivam[1,2] | John Casement[3] | Timothy Bates[4] | Philip Sloan[1,5] | Peter Thomson[6] | Max Robinson[1,5] | Ralf Kist[1,7]

[1]School of Dental Sciences, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

[2]Cancer Research Centre, Institute for Medical Research, National Institutes of Health, Setia Alam, Malaysia

[3]Bioinformatics Support Unit, Newcastle University, Newcastle upon Tyne, UK

[4]Queen Elizabeth Hospital, Birmingham, UK

[5]Newcastle upon Tyne Hospital NHS Foundation Trust, Newcastle upon Tyne, UK

[6]Oral and Maxillofacial Surgery, Faculty of Dentistry, The University of Hong Kong, Hong Kong SAR, Hong Kong

[7]Newcastle University Biosciences Institute, Newcastle University Centre for Cancer, Newcastle upon Tyne, UK

Correspondence
Ralf Kist, School of Dental Sciences, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, NE2 4BW, UK.
Email: ralf.kist@newcastle.ac.uk

Funding information
Newcastle upon Tyne Hospitals NHS Charity, Grant/Award Number: BH102216

Abstract

Background: A large number of oral squamous cell carcinomas (OSCCs) are believed to be preceded by oral potentially malignant disorders (OPMD) that have an increased likelihood of malignant transformation compared to clinically normal mucosa. This study was performed to identify differentially expressed genes between OPMDs that underwent malignant transformation (MT) and those that did not, termed "non-transforming" (NT) cases.

Methods: Total RNA was extracted from formalin-fixed paraffin-embedded tissue biopsies of 20 OPMD cases with known clinical outcomes (10 MT vs. 10 NT). Samples were assessed for quantity, quality and integrity of RNA prior to sequencing. Analysis for differential gene expression between MT and NT was performed using statistical packages in R. Genes were considered to be significantly differentially expressed if the False Discovery Rate corrected *P*-value was < 0.05.

Results: RNA yield was variable but RNA purity was good (A260/A280 > 1.90). Analysis of RNA-Sequencing outputs revealed 41 genes (34 protein-coding; 7 non-coding) that were significantly differentially expressed between MT and NT cases. The log2 fold change for the statistically significant differentially expressed genes ranged from −2.63 to 2.48, with 23 protein-coding genes being downregulated and 11 protein-coding genes being upregulated in MT cases compared to NT cases.

Conclusion: Several candidate genes that may play a role in malignant transformation of OPMD have been identified. Experiments to validate these candidates are underway. It is anticipated that this work will contribute to better understanding of the etiopathogenesis of OPMD and development of novel biomarkers.

KEYWORDS

malignant transformation, oral potentially malignant disorders, oral squamous cell carcinoma, RNA-Sequencing

1 | INTRODUCTION

Epidemiological studies estimate that more than 300,000 new cases and 145 400 deaths from oral cancers (inclusive of lip cancers) occur annually.[1] Approximately two-thirds of the new oral cancer cases occurred in men, and around 77% of oral cancer deaths were in less developed nations.[1] Most oral cancers are oral squamous cell carcinomas (OSCCs) and a proportion of OSCCs are believed to be preceded by clinical entities termed "oral potentially malignant disorders" (OPMDs).[2] OPMDs are defined as clinical disorders having an increased risk of developing OSCC in oral mucosa; either in recognisable lesions or clinically "normal" oral mucosa.[2]

There are several possible clinical outcomes for OPMD, the lesion remains unchanged, it increases in size, it regresses in size, it disappears completely or undergoes malignant transformation (MT). Several epidemiological studies conducted in different areas of the world have shown that most OPMD do not undergo MT although they may persist.[2,3] A contemporary systematic review and meta-analysis described a mean overall MT rate of 12.1% in oral epithelial dysplasia (OED) whilst a recent systematic review found a 7.9% prevalence rate of MT in OPMD.[3,4] Currently, there is no reliable method to determine the clinical outcome of patients with OPMDs. To compensate for the limitations in predicting malignant change, biomarkers have been sought based on an improved understanding of the underlying molecular pathogenesis of OSCC. Numerous individual biomarkers have been studied, but none have been validated for use in clinical practice.

By studying differential gene expression (DGE) between normal and abnormal tissue, in-depth understanding of the genetic pathways involved in carcinogenesis can be elucidated. Studies based on DGE have allowed researchers to dissect and examine the cancer transcriptome in a way that was not possible using conventional molecular biological methods. DGE has also contributed to the paradigm shift away from single biomarkers towards the use of gene expression signatures for diagnosis or prognosis.

The ability to identify patient sub-groups with similar molecular patterns in various tumour types have enabled researchers to define new molecular cancer sub-types enhancing better targeted therapy and patient care. A prime example is breast cancer where at least five molecular sub-types with prognostic correlation were discovered. The findings were then further refined and validated resulting in a predictive gene signature.[5] The lack of prognostic biomarkers in OPMD is a cogent reason to perform DGE-based studies to identify gene signatures for early diagnosis, therapy or prognosis in OPMD to inform targeted therapy. A recent meta-analysis performed by De Cecco et al. (2015) demonstrated the usefulness of DGE studies in stratifying HNSCC into six sub-types characterised by their respective clinico-pathological features and dysregulation of relevant signalling pathways.[6]

There are very few DGE studies on OPMD or oral epithelial dysplasia (OED).[7-11] As yet, only one truly investigated DGE between OPMD that transformed to OSCC and those that did not.[7] Saintigny et al (2011) proposed gene expression-based prediction models that showed superior prognostic accuracy when compared to models using clinico-pathologic risk factors.[7] As such, further studies in DGE between OPMD that undergo malignant transformation versus those that do not would provide much needed insight into the molecular mechanisms that translate into malignant transformation in OPMDs.

Whole transcriptome analysis is a major advancement in studying and understanding gene expression as it allows researchers to obtain a comprehensive view of the transcriptional profile at a given moment in time. A widely used method for profiling the whole transcriptome in a "snapshot" manner is RNA-Sequencing (RNA-Seq). As it captures the whole transcriptome, RNA-Seq is able to detect gene transcripts and is suitable for assessing genes that are differentially expressed between different disease states. In this study, we have used RNA-Seq as a discovery platform to identify transcripts of genes that may be involved in the malignant transformation of OPMD.

2 | MATERIALS AND METHODS

2.1 | Patients

OPMD cases for this study were selected from a previously studied cohort of OPMD patients.[12] A case was classified as having undergone MT when there was progression from an OPMD to oral squamous cell carcinoma (OSCC) after a period of six months or more from the time of the initial diagnosis of OPMD.

The following exclusion criteria were applied: i) Previous history of head and neck cancer; ii) Previous history of radiotherapy to the head and neck region; iii) Patients with hereditary/acquired conditions that are linked to an increased risk of head and neck SCC (such as ataxia telangiectasia, xeroderma pigmentosum, Fanconi anaemia etc); iv) Patients that were diagnosed as having chronic hyperplastic candidosis; v) Cases with incomplete/inconsistent records; vi) Cases with inadequate/damaged/unavailable FFPE tissue for analysis.

Demographic (age at diagnosis, sex) and clinico-pathological data (site, clinical diagnosis) were recorded for each patient. The clinical outcome and time to either malignant transformation or last follow-up was also recorded and calculated for the patients.

2.2 | Histopathological assessment

Archived haematoxylin and eosin (H&E)-stained sections of the cases identified from the OPMD database were retrieved and assessed to choose suitable formalin-fixed paraffin-embedded (FFPE) blocks from each case. Subsequently, selected FFPE blocks were retrieved from the Royal Victoria Infirmary Department of Cellular Pathology archives and 4 μm sections were prepared. H&E staining was performed on the DAKO CoverStainer (Agilent Technologies, USA).

All histopathological assessments were performed following a modified three-tier system adapted from the work published by Speight *et. al.* (2015) involving three oral and maxillofacial pathologists.[13]. The cases were graded using the three-tiered (mild, moderate or severe) World Health Organization (WHO) 2017 classification and binary grading systems.[2] The pathologists were blinded to clinical outcome of OPMD patients during the assessment and grading exercise.

2.3 | Total RNA extraction from formalin-fixed paraffin-embedded (FFPE) tissue

10 μm sections were cut from the FFPE blocks and collected in 2 ml microcentrifuge tubes. The number of sections per sample was dependent on the size of the tissue; 4 sections for small-sized samples, 3 - 4 sections for medium-sized samples, 2 - 3 sections for large-sized samples and 1 – 2 sections for very large-sized samples. RNA extraction and purification were performed using the QIAGEN RNeasy FFPE kit following the manufacturer's protocol (QIAGEN, Manchester, UK). Following RNA extraction, the concentration and the quality of the isolated RNA were measured using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, UK). The samples were then stored in a −80°C freezer prior to utilisation in downstream experiments.

2.4 | Whole transcriptome sequencing - RNA sequencing (RNA-Seq)

Whole transcriptome sequencing was performed using Illumina's Next Generation Sequencing RNA-Seq platform (Illumina, USA) following established protocols. RNA samples were assessed for quantity and integrity using the NanoDrop 8000 spectrophotometer V2.0 (Thermo Fisher Scientific, USA) and Agilent 2100 Bioanalyser (Agilent Technologies, Waldbronn, Germany). From each sample, 100 ng of total RNA was used to prepare RNA libraries using the KAPA Stranded RNA-Seq Kit with RiboErase (KAPA Biosystems, Massachusetts, USA). Prior to first strand cDNA synthesis, fragmentation was carried out using incubation conditions recommended by the manufacturer for degraded samples (65°C for 1 minute), and 14 cycles of PCR were performed for final library amplification. The libraries produced were quantified using the Qubit 2.0 spectrophotometer (Life Technologies, California, USA) and assessment of the average fragment size was performed using the Agilent 2200 TapeStation (Agilent Technologies, Waldbronn, Germany). The Illumina NextSeq®500 (Illumina Inc., Cambridge, UK) was used to generate 75 bp paired-end reads for each library. All RNA-Seq data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE156208.

2.5 | Statistical analysis

Statistical analysis was performed using IBM SPSS Statistics for Windows version 24 (IBM Corp., Armonk, N.Y., USA) and the R Environment for Statistical Computing version 3.2 (R Foundation for Statistical Computing, Vienna, Austria). A variety of methods were used to assess and analyse the data. 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) Statistical significance was defined at the 5% level. Confidence intervals (CI) at the 95% confidence level were reported where relevant.

2.6 | Bioinformatic analysis of RNA-Seq data

FastQ files generated from the sequencing runs were downloaded from the Illumina server using BaseMount, the command line interface for Illumina BaseSpace. Read quality of the FastQ files generated from the sequencing run was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and MultiQC (http://multiqc.info) was used to obtain summary statistics for quality control tests on the read quality. Reads were quantified against transcripts using "Kallisto".[14]

To obtain gene-level counts, a package from the R statistical programming language (R Foundation for Statistical Computing, Vienna, Austria), "tximport" was used. Gene annotation was obtained from Ensembl transcript IDs using the R package "biomaRt".[15]

The R package DESeq2 was used for normalisation and testing for differential gene expression by use of negative binomial generalised linear models.[16] Genes were considered to be significantly differentially expressed when the False Discovery Rate (FDR) using the Benjamini-Hochberg method corrected p-value was less than 0.05. A hypergeometric test was carried out to assess over-representation of gene ontology (GO) terms amongst genes found to be significantly differentially expressed. The R package "GOStats" was used to implement this test.[17]

2.7 | Ethics

This study was approved by the National Research Ethics Service Committee Northeast (Evaluation of the prognostic potential and functional significance of biomarkers in oral cancer; NRES Committee Northeast – Sunderland 11/NE/0118) and complies with UK legislation and guidelines.

3 | RESULTS

3.1 | Clinical parameters

Twenty cases (10 malignant transforming vs 10 non-transforming cases) with complete clinical data and applicable RNA for the experiment were selected for total RNA sequencing. The

demographic, clinical and histopathological characteristics of the cases are shown in Table 1. None of the clinical or pathological features were significantly correlated with clinical outcome. The cohort was composed predominantly of males and patients aged more than 50 years at diagnosis. The median time to MT was 17 months (IQR: 42.75 months).

3.2 | Differential gene expression (DGE)

All samples passed the quality control assessments to proceed for downstream analysis. Reads were assessed using FastQC and overall quality was high. RNA yield was variable but RNA purity was good (A260/A280 > 1.90). As expected, the RNA was highly degraded (RIN 1.4 - 2.6). Bioinformatic analysis of RNA-Seq outputs revealed 41 genes that were significantly differentially expressed between MT and NT cases (Table 2). The log2 fold change for the statistically significant differentially expressed genes ranged from −2.63 to 2.48, with 27 genes being downregulated and 14 genes being upregulated in MT cases compared to NT cases (Table 2). When the statistically significant gene list from our study was compared to the 2182 genes associated with oral cancer risk from the study by Saintigny et al (2011)[7], there were only 4 genes that overlapped: *CYP19A1*, *HIST1H2AJ*, *CCDC129 and MUC16*. However, these four genes were not in the gene-signature based predictive models developed by Saintigny et al (2011).[7]

3.3 | GO enrichment analysis

To discover the functions of the differentially expressed genes, we performed GO enrichment analysis (Table S1). Twenty of the most significant GO biological process (GOBP) terms associated with the identified significantly differentially expressed genes are listed in Table 3. Of these, three GOBP terms were noted to have a high degree of association with oral carcinogenesis: Regulation of response to wounding (Genes: *IER3*, *CD46* and *FAM46A)*, regulation of response to DNA damage stimulus (Genes: *IER3*, *SPIDR* and *MUC1)* and regulation of Notch signalling pathway (Genes: *DLX2 and CD46)*.

4 | DISCUSSION

Using RNA-Seq technology, our study has identified several novel differentially expressed genes that are associated with malignant transformation of OPMDs. The one previous study that similarly assessed DGE in OPMD used microarray technology.[7] Although we identified four of the same genes (*CYP19A1*, *HIST1H2AJ*, *CCDC129 and MUC16*), most of the statistically significant genes from our study do not overlap with the gene list (2182 in total) of the study by Saintigny et al (2011).[7] This could possibly be due to the following reasons: i) patient heterogeneity; ii) tissue heterogeneity; iii) molecular heterogeneity of OPMDs; iv) treatment heterogeneity; v) different gene expression analysis platforms employed; vi) differences in bioinformatic analysis

TABLE 1 Characteristics of OPMD cases according to clinical outcome (n = 20)

Characteristic	Non-transforming n = 10	Malignant transforming n = 10	*P*-value
Age [Mean (± SD)]	55.7 (±14.86)	60.0 (±12.41)	.491[a]
Sex			
Male	9	7	.582[b]
Female	1	3	
Site of OPMD			
Tongue	4	3	1.000[b]
Other sites	6	7	
OPMD			
Leukoplakia	9	8	1.000[b]
Erythroleukoplakia	1	2	
OED grading (WHO 2017)			
Mild	3	3	.635[c]
Moderate	1	3	
Severe	6	4	
Binary OED grading			
Low-grade	3	3	1.000[b]
High-grade	7	7	

Abbreviation: SD, Standard deviation.

[a]Independent t test.

[b]Fisher's Exact test.

[c]Pearson's Chi-square test.

TABLE 2 Significant differentially expressed genes associated with malignant transformation of OPMD

Ensembl gene ID	HGNC symbol	Gene biotype	Log2 fold change	FDR [BH *P*-value]
ENSG00000196805	*SPRR2B*	Protein coding	2.48	.015
ENSG00000283029	NA	Non-coding	2.44	.015
ENSG00000115844	*DLX2*	Protein coding	2.30	.015
ENSG00000229035	*SPRR2C*	Unprocessed pseudogene	2.28	.015
ENSG00000223802	*CERS1*	Protein coding	2.19	.044
ENSG00000166165	*CKB*	Protein coding	2.11	.015
ENSG00000137869	*CYP19A1*	Protein coding	2.10	.046
ENSG00000235852	NA	Antisense	2.00	.019
ENSG00000186648	*LRRC16B*	Protein coding	1.79	.030
ENSG00000276368	*HIST1H2AJ*	Protein coding	1.78	.037
ENSG00000123416	*TUBA1B*	Protein coding	1.72	.025
ENSG00000137331	*IER3*	Protein coding	1.61	.046
ENSG00000066248	*NGEF*	Protein coding	1.58	.031
ENSG00000127824	*TUBA4A*	Protein coding	1.48	.015
ENSG00000162836	*ACP6*	Protein coding	−1.05	.031
ENSG00000164808	*SPIDR*	Protein coding	−1.32	.025
ENSG00000117335	*CD46*	Protein coding	−1.43	.031
ENSG00000111670	*GNPTAB*	Protein coding	−1.44	.037
ENSG00000135338	*LCA5*	Protein coding	−1.45	.035
ENSG00000166432	*ZMAT1*	Protein coding	−1.47	.020
ENSG00000181804	*SLC9A9*	Protein coding	−1.67	.026
ENSG00000204789	*ZNF204P*	Processed pseudogene	−1.78	.037
ENSG00000165186	*PTCHD1*	Protein coding	−1.79	.036
ENSG00000112773	*FAM46A*	Protein coding	−1.80	.037
ENSG00000139292	*LGR5*	Protein coding	−1.82	.046
ENSG00000185499	*MUC1*	Protein coding	−1.83	.026
ENSG00000214290	*COLCA2*	Protein coding	−1.83	.016
ENSG00000267395	*DM1-AS*	Antisense	−1.86	.033
ENSG00000196724	*ZNF418*	Protein coding	−1.91	.019
ENSG00000177707	*NECTIN3*	Protein coding	−2.01	.026
ENSG00000115648	*MLPH*	Protein coding	−2.03	.019
ENSG00000279387	NA	NA	−2.12	.019
ENSG00000180347	*CCDC129*	Protein coding	−2.19	.020
ENSG00000235902	NA	Antisense	−2.21	.024
ENSG00000115112	*TFCP2L1*	Protein coding	−2.22	.015
ENSG00000116039	*ATP6V1B1*	Protein coding	−2.31	.025
ENSG00000177685	*CRACR2B*	Protein coding	−2.33	.015
ENSG00000134398	*ERN2*	Protein coding	−2.41	.015
ENSG00000167165	*UGT1A6*	Protein coding	−2.44	.015
ENSG00000107807	*TLX1*	Protein coding	−2.46	.016
ENSG00000181143	*MUC16*	Protein coding	−2.63	.015

Abbreviations; FDR, False discovery rate; BH, Benjamini-Hochberg. False Discovery Rate was calculated using Benjamini-Hochberg method and significance set at the 5% level ($p < 0.05$). HGNC, Human Genome Organisation Gene Nomenclature Committee. NA, not available. Genes that overlap with the gene-list from the study by Saintigny et al (2011) are underlined.[8]

TABLE 3 Twenty most significant GOBP terms associated with malignant transformation of OPMD

GOBP ID	*P*-value	Count	Term
GO:2001311	.002	1	lysobisphosphatidic acid metabolic process
GO:0010677	.003	2	negative regulation of cellular carbohydrate metabolic process
GO:1903034	.003	4	regulation of response to wounding
GO:2001020	.003	3	regulation of response to DNA damage stimulus
GO:0045912	.004	2	negative regulation of carbohydrate metabolic process
GO:0051084	.004	2	"de novo" posttranslational protein folding
GO:0016256	.004	1	N-glycan processing to lysosome
GO:0021893	.004	1	cerebral cortex GABAergic interneuron fate commitment
GO:0006458	.004	2	"de novo" protein folding
GO:0016266	.005	2	O-glycan processing
GO:0010760	.006	1	negative regulation of macrophage chemotaxis
GO:0021882	.006	1	regulation of transcription from RNA polymerase II promoter involved in forebrain neuron fate commitment
GO:0072757	.006	1	cellular response to camptothecin
GO:0006885	.006	2	regulation of pH
GO:0008593	.007	2	regulation of Notch signalling pathway
GO:0021898	.008	1	commitment of multipotent stem cells to neuronal lineage in forebrain
GO:0043382	.008	1	positive regulation of memory T cell differentiation
GO:0072710	.008	1	response to hydroxyurea
GO:0072711	.008	1	cellular response to hydroxyurea
GO:1901563	.008	1	response to camptothecin

Abbreviations: GOBP, Gene Ontology Biological Process; ID, Identifier.

methods/pipeline; vii) sample size. Furthermore, the patient cohort in the Saintigny et al (2011) study was enrolled in a chemo-preventive trial for treatment of leukoplakia which may have influenced the outcome of the OPMD as well as the gene expression profile.[7]

A recent study by Conway et al (2015) also employed RNA-Seq to assess DGE in "normal", OED and OSCC tissues; however, all three tissue states ("normal", OED and OSCC) were obtained from the same excision specimen.[8] Due to the well-recognised theory of field change in OPMD patients, it is understood that histologically "normal" tissue may not be molecularly "normal" and free from molecular change which introduces a confounder to the results obtained by Conway et al (2015). This confounding problem of normal epithelial tissue affects the majority of published gene expression studies involving OPMD and OSCC. Such studies may only provide an approximation of the molecular events that take place during malignant transformation of OPMD.

The relatively small number of significantly differentially expressed genes identified in our study highlights the high degree of similarity between cases that undergo MT and those that do not. This finding is consistent with the overall clinico-pathological features of OPMDs whereby it is difficult to accurately predict the clinical outcome of a patient with OPMD. Instead of focusing too much on individual genes, more emphasis should be placed on the pathways and biological processes involved.

Three of the GOBP terms found from the enrichment analysis; "regulation of response to DNA damage stimulus", "regulation of response to wounding" and "regulation of Notch signalling pathway", have been shown to be associated with carcinogenesis and have some degree of association with one another.[18-21] The relationship between DNA damage response (DDR) and carcinogenesis is one that is well established, and in recent years, there has been interest in the association between regulation of DDR and the regulatory effect of the Notch signalling pathway on DDR.[22,23] The association between regulation of wounding, cancer and the Notch signalling pathway is also one that is being studied with renewed interest in recent years, consistent with the hypothesis that cancer is an "over-healing wound".[18]

Although promising, the role of the Notch signalling pathway in head and neck cancers is quite complex as it has been shown to be able to exert both tumour suppressive and oncogenic effects.[19,21] Mutations in Notch pathway genes has been previously described in head and neck squamous cell carcinoma with much work being centred around *NOTCH1*.[19,24] Recent findings are more supportive of Notch as a tumour suppressor especially in head and neck squamous cell carcinoma as loss of Notch signalling has been shown to affect regulation of cell fate decisions in stem cells and stromal remodelling.[19,24-26] However, the role of Notch signalling in oral carcinogenesis specifically with regard to clinical outcome of OPMDs has yet to be fully elucidated.

The two significant genes from our study related to regulation of Notch signalling pathway were *DLX2* and *CD46*. Our results showed that *DLX2*, a homeobox gene that is involved in embryonic development, was overexpressed in cases that underwent MT compared to non-transforming cases. Increased expression of *DLX2* has been shown recently to be overexpressed in breast and ovarian cancers as well as advanced stages of gastric adenocarcinoma suggesting a potential role in carcinogenesis.[27-29] The study by Lee *et. al.* (2011) suggests that *DLX2* may be involved in tumour progression via metabolic-stress induced necrosis.[28] *DLX2* has also been implicated in transforming the role of transforming growth factor β (TGFβ) from a tumour suppressor to a tumour promoter by increasing the expression of the mitogenic transcription factor c-Myc, directly suppressing TGFβ receptor II and reducing expression of cell-cycle inhibitor $p21^{CIP1}$.[29] The role of DLX2 in oral carcinogenesis, however, is currently unknown.

Decreased expression of *CD46* that encodes for a complement regulatory protein (a membrane co-factor protein) was detected in cases that underwent malignant transformation. CD46 is also known as complement restriction factor as it facilitates inactivation of C3b and C4b of the complement system. Interestingly, other studies have shown that CD46 together with other complement restriction factors such as CD55 and CD59 are expressed at higher levels in head & neck cancer tissue compared to non-tumour tissue proposing that these proteins may play a role in tumour evasion of the complement system.[30] The decreased expression of CD46 observed in our study is different to that seen in OSCCs suggesting that *CD46* is dynamically expressed during oral carcinogenesis with possible temporal differences in expression before, during and after malignant transformation.

Archived formalin-fixed paraffin-embedded (FFPE) tissues are an invaluable resource that can be successfully used for molecular-based assays despite the degradation that often accompanies fixation and embedding of tissues in paraffin wax. Our study adds to the increasing body of work on utilisation of FFPE material for gene expression studies.

One of the limitations of our study is the relatively small number of cases included compared to the study by Saintigny et al (2011) that had an 86-patient cohort.[7] This was due to strict quality control resulting in exclusion of poor quality RNA samples. Another limitation is that gene expression studies only allow a snapshot of the transcriptomic profile at a given point in time, and as such is a very simplistic and static representation of a dynamic temporal process. Furthermore, an OPMD that was categorised as being a non-transforming case may eventually undergo MT. However, RNA-Seq analysis for this study was to serve only as an initial broad overview of the transcriptomic differences between OPMD cases that undergo MT and those that do not.

In summary, our study has identified candidate genetic pathways that may play a role in malignant transformation of OPMD. Experiments to validate these pathways and relevant genes are currently underway, and it is anticipated that this work will contribute to better understanding of the pathogenesis of OPMD and the development of novel prognostic biomarkers.

ACKNOWLEDGEMENTS

We thank Eva Wozniak and Charles Mein at the Genome Centre, Queen Mary University of London for performing RNA-Seq experiments. This study was funded by the Newcastle upon Tyne Hospitals NHS Charity.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

AUTHOR CONTRIBUTION

Hans Prakash Sathasivam: Data curation; Formal analysis; Investigation; Methodology; Writing-original draft; Writing-review & editing. John Casement: Formal analysis; Investigation; Writing-review & editing. Timothy Bates: Data curation; Methodology; Resources; Writing-review & editing. Philip Sloan: Data curation; Investigation; Writing-review & editing. Peter James Thomson: Data curation; Methodology; Resources; Writing-review & editing. Max Robinson: Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing-review & editing. Ralf Kist: Conceptualization; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing-review & editing.

ORCID

Hans P. Sathasivam https://orcid.org/0000-0002-0510-9772
Peter Thomson https://orcid.org/0000-0002-2007-7975
Max Robinson https://orcid.org/0000-0003-4491-6865
Ralf Kist https://orcid.org/0000-0003-3729-5693

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
2. El-Naggar AK, Chan JKC, Grandis JR, Takata T, Slootweg PJ, eds. WHO Classification of Head and Neck Tumours. *WHO Classification of Tumours* (4th edn), vol. No. 9. Lyon, France: International Agency for Research on Cancer; 2017.
3. Iocca O, Sollecito TP, Alawi F et al Potentially malignant disorders of the oral cavity and oral dysplasia: A systematic review and meta-analysis of malignant transformation rate by subtype. *Head Neck*. 2020;42:539–555.

4. Mehanna HM, Rattay T, Smith J, McConkey CC. Treatment and follow-up of oral dysplasia - a systematic review and meta-analysis. *Head Neck*. 2009;31:1600–1609.
5. Parker JS, Mullins M, Cheang MC et al Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27:1160–1167.
6. De Cecco L, Nicolau M, Giannoccaro M et al Head and neck cancer subtypes with biological and clinical relevance: Meta-analysis of gene-expression data. *Oncotarget*. 2015;6:9627–9642.
7. Saintigny P, Zhang L, Fan YH et al Gene expression profiling predicts the development of oral cancer. *Cancer prevention research (Philadelphia, Pa)*. 2011;4:218–229.
8. Conway C, Graham JL, Chengot P et al Elucidating drivers of oral epithelial dysplasia formation and malignant transformation to cancer using RNAseq. *Oncotarget*. 2015;6:40186–40201.
9. Zhu G, He Y, Yang S, Chen B, Zhou M, Xu XJ. Identification of Gene and MicroRNA Signatures for Oral Cancer Developed from Oral Leukoplakia. *Biomed Res Int*. 2015;2015:841956.
10. Chen C, Mendez E, Houck J et al Gene expression profiling identifies genes predictive of oral squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev*. 2008;17:2152–2162.
11. Farah CS, Fox SA. Dysplastic oral leukoplakia is molecularly distinct from leukoplakia without dysplasia. *Oral Dis*. 2019;25:1715–1723.
12. Bates T, Kennedy M, Diajil A et al Changes in Epidermal Growth Factor Receptor Gene Copy Number during Oral Carcinogenesis. *Cancer Epidemiol Biomarkers Prev*. 2016;25:927–935.
13. Speight PM, Abram TJ, Floriano PN et al Interobserver agreement in dysplasia grading: toward an enhanced gold standard for clinical pathology trials. *Oral Surg Oral Med Oral Pathol. Oral Radiol*. 2015;120:474–482 e472.
14. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34:525–527.
15. Durinck S, Moreau Y, Kasprzyk A et al BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*. 2005;21:3439–3440.
16. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
17. Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. *Bioinformatics*. 2007;23:257–258.
18. Sundaram GM, Quah S, Sampath P. Cancer: the dark side of wound healing. *The FEBS Journal*. 2018;285:4516–4534.
19. Agrawal N, Frederick MJ, Pickering CR et al Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science*. 2011;333:1154–1157.
20. Chigurupati S, Arumugam TV, Son TG et al Involvement of Notch Signaling in Wound Healing. *PLOS ONE*. 2007;2:e1167.
21. Porcheri C, Meisel CT, Mitsiadis T. Multifactorial Contribution of Notch Signaling in Head and Neck Squamous Cell Carcinoma. *Int J Mol Sci*. 2019;20:1520.
22. Vermezovic J, Adamowicz M, Santarpia L et al Notch is a direct negative regulator of the DNA-damage response. *Nature Structural & Molecular Biology*. 2015;22:417–424.
23. Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. *Mol Cell*. 2010;40:179–204.
24. Stransky N, Egloff AM, Tward AD et al The mutational landscape of head and neck squamous cell carcinoma. *Science*. 2011;333:1157–1160.
25. Wilson A, Radtke F. Multiple functions of Notch signaling in self-renewing organs and cancer. *FEBS Lett*. 2006;580:2860–2868.
26. Liu J, Sato C, Cerletti M, Wagers A. Notch signaling in the regulation of stem cell self-renewal and differentiation. *Curr Top Dev Biol*. 2010;92:367–409.
27. Tang P, Huang H, Chang J, Zhao GF, Lu ML, Wang Y. Increased expression of DLX2 correlates with advanced stage of gastric adenocarcinoma. *World J Gastroenterol*. 2013;19:2697–2703.
28. Lee SY, Jeon HM, Kim CH et al Homeobox gene Dlx-2 is implicated in metabolic stress-induced necrosis. *Molecular Cancer*. 2011;10:113.
29. Yilmaz M, Maass D, Tiwari N et al Transcription factor Dlx2 protects from TGFβ-induced cell-cycle arrest and apoptosis. *EMBO J*. 2011;30:4489–4499.
30. Kesselring R, Thiel A, Pries R et al The complement receptors CD46, CD55 and CD59 are regulated by the tumour microenvironment of head and neck cancer to facilitate escape of complement attack. *Eur J Cancer*. 2014;50:2152–2161.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Sathasivam HP, Casement J, Bates T, et al. Gene expression changes associated with malignant transformation of oral potentially malignant disorders. *J Oral Pathol Med*. 2021;50:60–67. https://doi.org/10.1111/jop.13090