Oral Oncology 96 (2019) 42–50

Contents lists available at ScienceDirect

Oral Oncology

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

Exome sequencing of oral leukoplakia and oral squamous cell carcinoma implicates DNA damage repair gene defects in malignant transformation

Camile S. Farah[a,b,c,*], Maryam Jessri[b,c], Nigel C. Bennett[c], Andrew J. Dalley[c], Kate D. Shearston[b], Simon A. Fox[b]

[a] Australian Centre for Oral Oncology Research & Education, Perth, WA 6009, Australia
[b] UWA Dental School, University of Western Australia, Perth, WA 6009, Australia
[c] UQ Centre for Clinical Research, The University of Queensland, Herston, QLD 4029, Australia

ARTICLE INFO

Keywords:
Oral leukoplakia
Oral dysplasia
Progression to cancer
Oral cancer
Malignant transformation
Whole exome sequencing
BRCA1
Fanconi anaemia
DNA damage repair
Double strand break

ABSTRACT

Objectives: To map the genomic pathways of patients with oral leukoplakia (OLK) which transformed to cancer (progressive) and those which did not (non-progressive), and to compare their exomic profiles.
Materials and methods: Whole exome sequencing was performed on 42 sequential samples from five progressive and eight non-progressive patients. Association of genomic variant frequencies with progression or lesion severity were analysed by non-parametric tests (Kruskal-Wallis and Mann-Whitney-Wilcoxon) and multivariate sparse partial least squares discriminant analysis (sPLS-DA). Enrichment analysis was used to characterise the effect of mutations upon biological pathways. Confirmatory studies used qPCR and immunohistochemistry.
Results: Using sPLS-DA, the variant frequency of a small number of genes could be used to classify the samples based on lesion severity or progressive status. Enrichment analysis showed that DNA damage repair gene related pathways were highly impacted in lesions which progressed to cancer. Multivariate analysis of a set of 148 DNA damage repair genes could be used to classify progressive lesions using mutation frequency. *BRCA1, BRCA*2 and other double strand break (DSB) repair Fanconi anaemia (FA)/BRCA pathway genes were prominent contributors to this classification.
Conclusion: Patients with progressive and non-progressive OLK can be differentiated using the frequency of exomic variants, particularly in DNA damage repair pathway genes. To our knowledge, this is the first report of FA/BRCA (DSB) pathway involvement in malignant transformation of OLK to oral squamous cell carcinoma (OSCC).

Introduction

Epithelial carcinogenesis is a multistep process involving environmental and genetic factors and can originate within metaplastic or dysplastic lesions [1,2]. Not all OLK progress to cancer; important clinically recognised indicators of malignant transformation being a higher degree of dysplasia [3], and a change in morphology towards a non-homogeneous presentation [4]. Despite success for other tumour types, translational research is yet to provide panels of biomarkers, genes or clinical observations that can predict OLK malignant transformation with certainty.

Early diagnosis and intervention are pivotal to the successful treatment of patients with OSCC. A breakthrough adjunct to the resources available in medical oncology, next generation sequencing (NGS) has already been used to explore the molecular aspects of head and neck squamous cell carcinoma (HNSCC) [5–9], overlooking perhaps, potentially predictive molecular changes present in OLK. The histopathology of OLK and the early stages of OSCC may have underlying genomic alterations with potential to become clonal; hence the optimal time for targeted treatment of trenchant molecular changes would ideally be before they contribute to clonal selection. Notwithstanding the prevailing research focus toward molecular changes that are concomitant with frank OSCC, there is great value to studying the earliest occurrences of oral epithelial dysplasia (OED) molecular changes that may be apparent in OLK.

Studies investigating the progression of potentially malignant lesions to cancer are indicative of a step-wise pattern for accumulation of mutations [10,11]. Agrawal *et al.*, sequenced the whole genome of 11 oesophageal adenocarcinomas and 2 adjacent Barrett's oesophagus lesions to observe mutations found in the cancerous tissues to be mostly

* Corresponding author at: Australian Centre for Oral Oncology Research & Education, Nedlands, WA 6009, Australia.
E-mail address: camile@oralmedpath.com.au (C.S. Farah).

https://doi.org/10.1016/j.oraloncology.2019.07.005
Received 9 April 2019; Received in revised form 28 May 2019; Accepted 5 July 2019
Available online 09 July 2019
1368-8375/

Table 1
Demographic and clinic-pathologic data of progressive and non-progressive samples.

Patient ID Sample ID	Age[a]	Gender	Smoking	Follow up	Biopsy date	Clinical presentation	Histopathological Diagnosis (grade of dysplasia)	Lesion site
Progressive								
P1	34	F	Yes	90 months				
PNP_BC01					06/11/2000	Homogeneous leukoplakia	Low risk dysplasia (mild)	Right ventrolateral tongue
PNP_BC02					02/03/2005	Homogeneous leukoplakia	Epithelial hyperplasia	Right lateral tongue
PNP_BC03					11/04/2008	Non-homogeneous leukoplakia	Well differentiated OSCC	Anterior right lateral tongue
PNP_BC04					11/04/2008	Non-homogeneous leukoplakia	Poorly differentiated OSCC	Posterior right lateral tongue
P2[b]	66	F	Yes	79 months				
PNP_BC05					16/08/2002	Non-homogeneous leukoplakia	Low risk dysplasia (mild)	Right lateral tongue
PNP_BC08					16/08/2002	Non-homogeneous leukoplakia	Epithelial hyperplasia	Right buccal mucosa
PNP_BC06					14/12/2004	Non-homogeneous leukoplakia	High risk dysplasia (moderate)	Anterior right lateral tongue
PNP_BC07					14/12/2004	Non-homogeneous leukoplakia	Low risk dysplasia (mild)	Posterior right lateral tongue
P3	55	F	Yes	18 months				
PNP_BC09					09/12/1993	Non-homogeneous leukoplakia	Low risk dysplasia (mild)	Floor of mouth
PNP_BC10					20/06/1995	Non-homogeneous leukoplakia	Well differentiated OSCC	Floor of mouth
P4	69	F	N/A[c]	28 months				
PNP_BC11					21/01/1994	Erythroleukoplakia	Low risk dysplasia (mild)	Left buccal mucosa
PNP_BC12					27/05/1994	Erythroleukoplakia	Well differentiated OSCC	Left buccal mucosa
P5	31	M	N/A[c]	16 months				
PNP_BC13					23/11/1994	Non-homogeneous leukoplakia	Low risk dysplasia (mild)	Anterior right lateral tongue
PNP_BC15					02/04/1996	Non-homogeneous leukoplakia	Well differentiated OSCC	Anterior right lateral tongue
Non-progressive								
NP1	57	M	No	156 months				
PNP_BC20					11/01/2006	Non-homogeneous leukoplakia	Epithelial hyperplasia	Right lateral tongue
PNP_BC21					11/01/2006	Non-homogeneous leukoplakia	High risk dysplasia (moderate)	Right lateral tongue
PNP_BC22					11/01/2006	Non-homogeneous leukoplakia	Low risk dysplasia (mild)	Right lateral tongue
PNP_BC23					24/05/2006	Non-homogeneous leukoplakia	Low risk dysplasia (mild)	Right lateral tongue
PNP_BC24					28/03/2007	Non-homogeneous leukoplakia	High risk dysplasia (severe)	Right lateral tongue
NP2	49	M	No	186 months				
PNP_BC25					04/07/2002	Erythroleukoplakia	Low risk dysplasia (mild)	Left lateral tongue
PNP_BC26					07/06/2006	Erythroleukoplakia	Low risk dysplasia (mild)	Left lateral tongue
PNP_BC27					13/12/2006	Erythroleukoplakia	Epithelial hyperplasia	Left lateral tongue
PNP_BC29					19/11/2008	Erythroleukoplakia	Low risk dysplasia (mild)	Left lateral tongue
NP3	80	F	Yes	126 months				
PNP_BC30					08/06/2007	Non-homogeneous leukoplakia	High risk dysplasia (severe)	Left lateral tongue
PNP_BC31					20/07/2007	Non-homogeneous leukoplakia	High risk dysplasia (moderate)	Left lateral tongue
NP4	57	M	Yes	186 months				
PNP_BC32					3/07/2002	Non-homogeneous leukoplakia	Low risk dysplasia (mild)	Right buccal mucosa
PNP_BC33					25/03/2004	Non-homogeneous leukoplakia	Low risk dysplasia (mild)	Right buccal mucosa
PNP_BC36					26/08/2004	Non-homogeneous leukoplakia	Epithelial hyperplasia	Right lower labial mucosa
PNP_BC34					15/09/2005	Non-homogeneous leukoplakia	Low risk dysplasia (mild)	Left ventral tongue
PNP_BC35					09/04/2008	Non-homogeneous leukoplakia	Low risk dysplasia (mild)	Left ventral tongue
NP5	65	M	Yes	138 months				
PNP_BC37					26/07/2006	Non-homogeneous leukoplakia	Epithelial hyperplasia	Right buccal mucosa
PNP_BC38					02/08/2006	Non-homogeneous leukoplakia	Low risk dysplasia (mild)	Right buccal mucosa
PNP_BC39					12/09/2008	Erythroleukoplakia	Low risk dysplasia (mild)	Right buccal mucosa
NP6	64	F	Yes	121 months				
PNP_BC40					21/11/2007	Non-homogeneous leukoplakia	High risk dysplasia (severe)	Left lateral tongue
PNP_BC45					21/11/2007	Homogeneous leukoplakia	Epithelial hyperplasia	Dorsal tip of tongue
PNP_BC41					04/04/2008	Homogeneous leukoplakia	Parakeratosis	Left lateral tongue
PNP_BC42					04/04/2008	Homogeneous leukoplakia	Parakeratosis	Left lateral tongue
PNP_BC43					10/09/2008	Non-homogeneous leukoplakia	Low risk dysplasia (mild)	Posterior left lateral tongue
PNP_BC46					10/09/2008	Homogeneous leukoplakia	Low risk dysplasia (mild)	Anterior right lateral tongue
NP7	75	M	Yes	220 months				

(*continued on next page*)

Table 1 (*continued*)

Patient ID Sample ID	Age[a]	Gender	Smoking	Follow up	Biopsy date	Clinical presentation	Histopathological Diagnosis (grade of dysplasia)	Lesion site
PNP_BC47					28/08/1990	Homogeneous leukoplakia	High risk dysplasia (moderate)	Gingival tuberosity
PNP_BC48					0703/1995	Homogeneous leukoplakia	Epithelial hyperplasia	Gingival tuberosity
NP8	56	F	Yes	190 months				
PNP_BC49					17/10/2002	Homogeneous leukoplakia	Epithelial hyperplasia	Floor of mouth

[a] Age at the time of first diagnosis.
[b] Patient developed OSCC in right lateral tongue on 28/01/2009 but sample was not available for sequencing.
[c] Smoking data not available.

present in the Barrett's oesophagus genome [12]. This is an important finding, as it once again suggests molecules predicting malignant progression to be theoretically present in OLK or even macroscopically and histologically normal appearing samples. These important findings support our hypothesis that molecular abnormalities that are predictive of malignant progression may be detectable in OLK or even in tissues that are macroscopically and histologically normal in appearance.

Using sequential patient biopsies of OLK that either progressed to cancer (progressive) or did not transform to cancer (non-progressive), we used whole exome sequencing (WES) to compare the number and nature of exomic variants between transformed and non-transformed samples. In so doing, we have established the molecular profile of transformed OLK and OED.

Materials and methods

Patient samples

This study was approved by the Human Research Ethics Committee of the University of Queensland (2007001478) and the Royal Brisbane & Women's Hospital (HREC/10/QRBW/336). Clinical diagnosis of OLK in 13 patients was confirmed by an oral medicine specialist and histopathological grading of dysplasia was determined on freshly prepared haematoxylin and eosin stained slides by an oral pathologist blinded to any patient details using WHO guidelines [13,14]. All lesions were diagnosed through incisional biopsy and patients were monitored regularly every 3–6 months by an oral medicine specialist. At the time of preparing this manuscript (31/12/2018) 5 patients had developed OSCC and 8 patients had not; hereafter these patients are referred to as progressive and non-progressive respectively. A total of 42 sequential samples from 5 progressive patients (2 hyperplasia, 7 OED and 5 OSCC) and 8 non-progressive patients (9 hyperplasia and 19 OED) were included in this study (Table 1). Upon a diagnosis of OSCC, lesions were fully resected. Immunohistochemical assessment of p16 was used for HPV testing. Slides with greater than 70% of p16-positive epithelial cells were considered positive for HPV [15,16]. None of the samples were determined HPV positive.

DNA extraction

DNA was isolated from formalin fixed paraffin embedded (FFPE) samples using Genomic DNA Isolation Kit (Agencourt® DNAdvance™, Beckman Coulter, Inc., CA, USA) according to the manufacturer's protocol. Briefly, 5 µm sections were cut to a depth of 1 mm and dewaxed then air-dried and resuspended in 400 µL lysis master mix according to the kit protocol. Genomic DNA was eluted in 50 µL nuclease free water and stored at −80 °C.

Library construction and sequencing

Exome capture and library preparation were performed according to the kit manufacturer's recommendations with minor modifications (SureSelect^{XT} for SOLiD 5500, vA0, Agilent Technologies, CA, USA). Input genomic DNA (range 250 ng–3 µg) was fragmented (Covaris® Inc., MA, USA), size-selected (range 150–300 bp) using AMPure XP beads (Beckman Coulter Inc, CA, USA), end-repaired, A-tailed and ligated with specific adapters. The libraries were amplified and purified then quantified with 2100 Bioanalyzer (Agilent). After purification, 500 ng of the library was hybridized to SureSelect^{XT} Human All Exon V4 (Agilent) overnight and then amplified for 8 cycles according to the manufacturer's protocol. Quality control of the target enriched library was performed using the 2100 Bioanalyzer (Agilent). Standard sample preparation for sequencing on the SOLiD 5500XL platform E120 scale were used (Life Technologies, CA, USA) and samples were sequenced on the 5500 Genetic Analysis System (Applied Biosystems, CA, USA).

Bioinformatic analysis

Analysis of raw sequencing datafiles (fastq) was performed using a customised pipeline. Sequence reads were aligned to human genome build hg19 using the SHRiMP tool (v2.2.3) [17]. Reads were further filtered, sorted, indexed and converted to BAM format using the Picard software toolkit (http://broadinstitute.github.io/picard/). Genetic variants (single nucleotide and indels) were identified using the GATK software [18,19] and the hg19 reference human genome. Bias from PCR duplicates was limited by flagging likely duplicates using the Picard MarkDuplicates tool. GATK was used to identify and realign reads around known indels [20], and base quality score recalibration. Variant calling was performed using the GATK HaplotypeCaller tool. A confidence score of at least 30 was required for variant calling. Quality filtering of variants used reference data from dbSNP, the HapMap project, the 1000 Genomes project and Illumina array based genotype data. The SNP variants were filtered using the VariantAnnotation R package in order to select for those that corresponded to protein coding regions, met the quality criteria (GATK PASS flag) and were predicted to result in nonsense and non-synonymous amino acid substitutions. These variants were used for subsequent statistical analysis.

Statistical analysis

All statistical tests were performed using R package (www.**r-project**.org), and MetaCore platform (https://portal.genego.com/) was used for pathway analyses. Due to the small number of samples and in order to increase statistical power, no distinction was made in relation to the grade of dysplasia or the grade of differentiation of OSCC. The numbers of variants in each gene were compared in different groups of samples using non-parametric tests namely Kruskal-Wallis and Mann-Whitney-Wilcoxon test. Separation of samples was tested using a multilevel sPLS-DA using mixOmics R package [21] which considered repeated measurements. All p-values in this study were corrected for multiple comparisons using Bonferroni correction method. The curated list of 148 DNA repair pathway related genes was derived from MetaCore.

RT-qPCR

RNA was extracted from a different set of biopsies of OSCC tumour, dysplastic tissue and the normal margin, following which sample preparation, RNA extraction and RNA quality control were performed as we have previously described [22]. Gene specific PCR primers were designed using the Primer 3 software (Supplementary Table S1). qPCR was performed using the SsoAdvanced SYBR Green SuperMix (Bio-Rad) according to the manufacturers' instructions and run on a CFX Connect Real-Time PCR Detection System (Bio-Rad). Data analysis was performed using the CFX Manager software 3.1 (Bio-Rad). Data normalisation was performed using *GAPD, HPRT1* and *UBC* as reference genes following stability assessment using the CFX Manager software. Results were presented as deltaCq (ΔCq), calculated by subtraction of the Cq value for the target gene from the mean Cq value of the reference genes. Statistical analysis of qPCR data was performed using Prism 6.07 (Graphpad Software, CA, USA).

BRCA1 and BRCA2 immunohistochemistry staining

Confirmation of protein expression was undertaken on 33 new FFPE specimens, eight of which were negative for dysplasia (herein considered normal), four hyperplasia, four mild OED, five moderate to severe OED and 12 OSCC; with four being highly differentiated OSCC, and the remaining eight moderate/poorly differentiated OSCC. Of these patients, 12 were female, 12 male and one gender unknown, with a mean age of 57 years (standard deviation 15 years). Diagnosis was confirmed retrospectively by an oral pathologist according to the WHO classification system [13,14]. Immunohistochemistry was performed as described by us [23], and immunoreactivity assessed using a semi-quantitative method as previously described [24].

Results

Whole exome sequencing of 42 FFPE samples which were collected from 13 patients over a period of at least 10 years was performed successfully. A mean sequence coverage of targeted exomic regions of 270-fold with 91% of loci covered at > 50 folds was achieved. On average 167,809 variants per sample were identified, of which 132,894 (~79%) were Single Nucleotide Polymorphisms (SNP), 14,685 (~8%) were insertion/deletions which caused 421 (0.2% frameshift) alterations per sample. From the 10,167 genes included in this study, the number of SNPs in 833 genes showed no variance across all samples and consequently analysis was performed on the number of variants in remaining 9334 genes. We found Zinc Finger 717 protein (*ZNF717*) to have the highest number of variants amongst our samples.

The SNP and indel data for each sample was filtered using databases of known human genetic variation in order to remove known germline variants since there was no matched normal control. SNPs were further selected to include only those variants within protein coding regions which were predicted to result in nonsense or non-synonymous amino acid substitutions. The resulting variant profiles for each sample were then used for statistical analysis to identify genes which exhibit significant differences in mutation frequency between the different clinical groups. Initial analysis was performed using a univariate method (Mann-Whitney test with Bonferroni correction for multiple testing) to examine individual gene level differences between groups. This approach did not identify any genes having significantly different variant frequencies between samples with varying grades of histopathological abnormality within either the progressive group (ie normal vs dysplastic and OSCC) or non-progressive group (ie normal vs dysplastic). We then examined the first diagnosed dysplastic lesions in each group in order to identify if any genes demonstrated variant frequencies associated with progression to cancer. The univariate statistical approach did not discover any statistically significant genes between the progressive and non-progressive dysplastic lesions.

We next applied an alternative statistical approach based upon multivariate dimension reduction techniques by using multi-level sparse partial least square regression discriminant analysis (sPLS-DA) to identify gene signatures with differing mutation frequencies between the clinical groups [21]. The application of sPLS-DA was able to successfully provide discrimination across the clinical groups. This multivariate method was able to discriminate between lesions on the basis of sample severity within both the progressive (Supplementary Fig. S1A) and non-progressive (Supplementary Fig. S1B). More significantly, this multivariate analysis was able to distinctly separate the first dysplastic OLK based upon their progressive status using a signature of 15 genes (*FAAH, OR2A5, FAT1, DAGLB, PTPRZ1, OR2L2, ACOXL, C17orf78, CCDC149, DCLRE1B, FRMD4A, PRCP, DCST1, GPR128, PIM3*) (Fig. 1). For most of these genes the frequency of variants was higher in progressive samples than non-progressive lesions with fatty acid amide hydrolase (*FAAH*) and the G-protein-coupled receptor, olfactory receptor 2A5 (*OR2A5*) contributing the most to the separation on component 1 as seen in Fig. 1. There were four signature genes (*OR2L2, DCST1, GPR128* and *PIM3*) which had a lower variant frequency across the dataset in the progressive samples. These results demonstrated that it was possible to derive potentially classifying gene signatures able to discriminate, in this sample cohort, dysplastic OLK which progressed to OSCC from those which did not.

Pathway analysis

In addition to the above gene-based approaches to examine genetic differences between the clinical groups, we also employed a systems

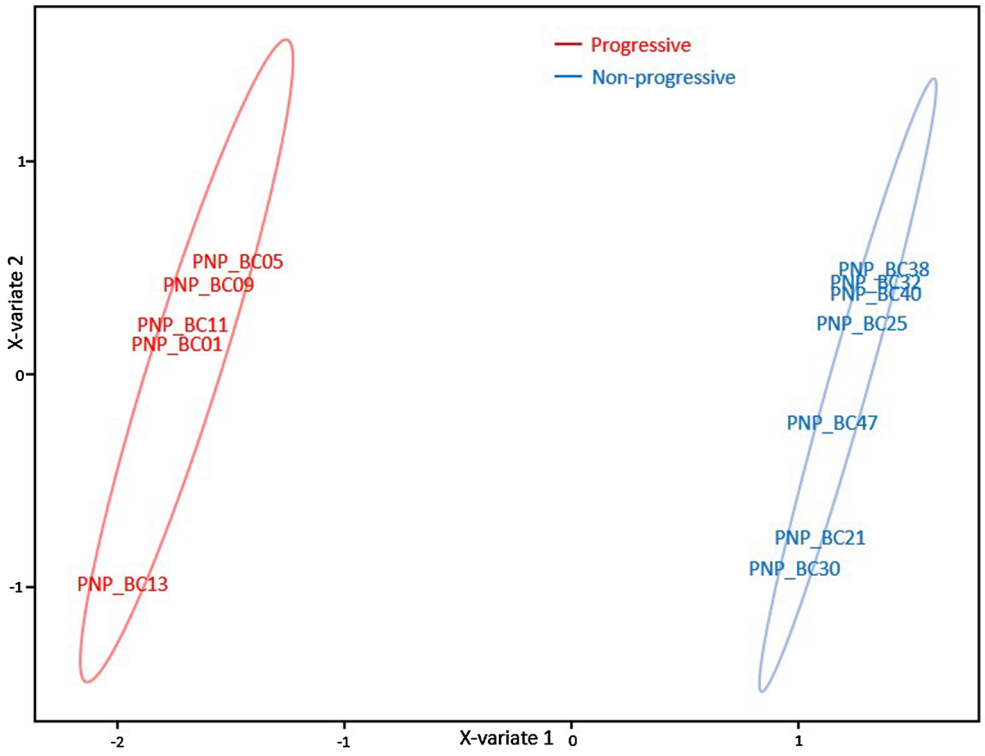

Fig. 1. Score plot of two-component sPLS-DA for the first dysplastic lesions diagnosed in progressive and non-progressive patients. The groups were clearly separated on the basis of the first component alone (15 genes). The ellipses correspond to 95% confidence. Patient NP8 was not included in this analysis since the initial diagnosis was epithelial hyperplasia.

biology based strategy by characterising the differences in the biological pathways most significantly affected by mutation in each sample. These were analysed on an individual patient basis and then compared between the groups to explore whether there were distinctions in the impact of mutation upon biological signalling between the clinical groups. Pathway analysis for all samples from each patient was performed simultaneously. The top 10 most significantly affected pathways in each progressive patient are presented in Fig. 2. A common finding for all progressive patients was the reporting of DNA damage repair pathways in the top 10 significantly enriched pathways. For example, "Role of BRCA1 and BRCA2 in DNA repair" pathway was enriched in 3/5 patients (P3, P4 and P5), general schema of breast cancer pathway (including *BRCA1* and *BRCA2*) was enriched in two patients (P2 and P3), ovarian cancer pathways was enriched in two patients (P3 and P4), while in patient (P1), the mismatch repair (MMR) pathway was significantly affected by mutation. Patient P3 showed the highest number of enriched pathways for exomic variants in DNA damage repair mechanisms presented as 4/10 of the most highly affected pathways being related to DNA damage repair. Immunological defence mechanism and keratin filaments pathways were the other most commonly enriched biological pathways in the progressive patient samples.

In contrast, in non-progressive patients, only 2/8 patients had DNA damage repair related pathways amongst the most highly enriched (Supplementary Fig. S2). Keratin filaments (8/8), DAP12 receptors regulation of NK cells pathways (6/8) and mucin expression in CF airways (5/8) were amongst the most highly impacted pathways in non-progressive patients. Interestingly, despite the absence of OSCC in the non-progressive cohort, the *NOTCH1* mediated *NF-kB* activation (3/8) and epithelial-to-mesenchymal transition (EMT) (3/8) pathways were amongst the most highly affected.

DNA damage repair in progression of cancer

Having identified a potential biological role for DNA repair mechanisms through analysis of genetic perturbation of signalling pathways, we next explored our dataset by focusing upon a group of genes encoding proteins with key roles in the DNA damage repair machinery. We curated a list of 148 DNA damage repair genes (Supplementary Table S2), extracted the variant data for these genes in the first dysplastic lesions from our dataset and examined their classification performance for progressive vs non-progressive using sPLS-DA. In this analysis, progression was used as the outcome and the number of variants in the 148 genes as predictors, yielding 2 PLS components. The first component represented 0.8993 of the original data, and the second component added 0.0788 to provide a final 0.9781 representation of data. This equates to the ability of a blind examiner determining the progressive status of lesions with knowledge of the number of non-synonymous variants in the 148 DNA repair genes with 97.81% accuracy. *BRCA1* and *BRCA2* with 0.26 and 0.24 respectively, were the most contributing DNA repair genes to classification of first dysplastic samples from progressive and non-progressive patients. The most highly contributing genes (> 0.15) to this classification are listed in Table 2. Fig. 3 shows a graphical representation of variant frequency (non-synonymous, non-sense and frameshift) in a selected panel of key MMR and DSB repair genes across the entire sample dataset. To provide additional supporting evidence we examined gene and protein expression of selected DNA damage repair genes in an independent cohort of OSCC, dysplasia and normal samples. Quantitative assessment of DNA damage repair gene expression (*BRCA1, BRCA2, MLH1, MSH2, MSH6, PMS2*) by qPCR showed that expression of *BRCA1/2* and *MMR* genes decreased with increasing lesion severity from normal to OSCC in an independent group of samples (Fig. 4). This was further confirmed at protein level for BRCA1 and BRCA2 by immunohistochemistry (Fig. 5), which clearly shows reduced expression of both in OSCC compared to normal tissue and hyperplasia, and reduced expression of BRCA2 in both OED and OSCC compared to normal tissue and hyperplasia.

Discussion

We performed exome sequencing of multiple biopsies from 13 patients with progressive and non-progressive OLK in order to map genomic pathways and identify classifying variant profiles. We successfully applied multivariate statistics to show that the number of non-synonymous variants classified the samples based upon lesion severity

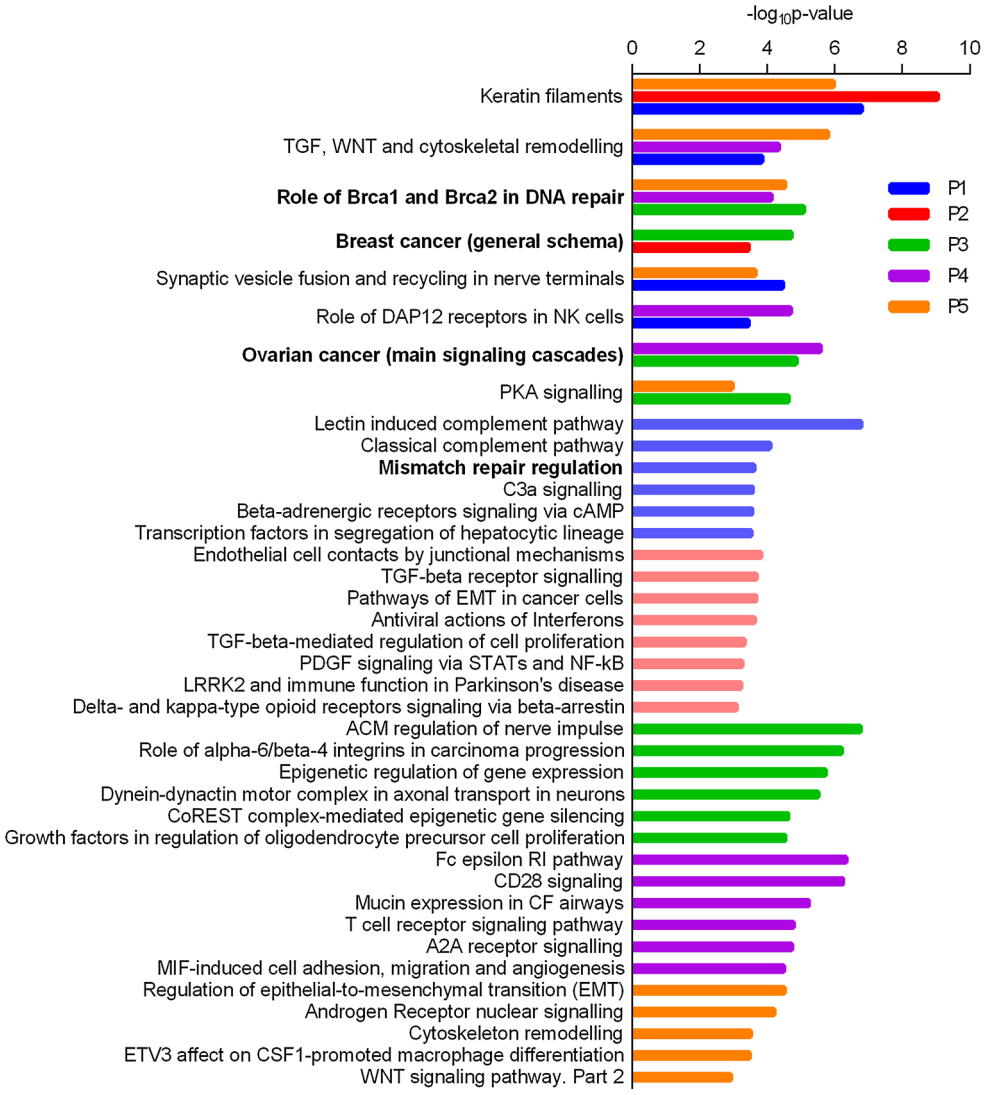

Fig. 2. Pathway analysis of gene variants in the progressive cohort by patient. Enriched pathways were determined using the MetaCore platform. Pathways for each patient are ranked according to frequency of appearance in the cohort.

Table 2
Most contributing genes to the sPLS-DA stratification of the first dysplastic lesions using variants in 148 DNA damage repair genes with progression as the outcome.

Gene symbol	Component 1[a]	Component 2[a]
BRCA1	0.26	0.02
BRCA2	0.24	0.05
FANCA	0.21	0.02
DCLRE1B	0.20	0.00
NEIL3	0.18	0.04
RPA4	0.17	0.02
MLH1	0.16	−0.05
CHEK2	0.16	−0.13

[a] Each gene contributes with a different weighting to the classification performance of the model.

or progression to malignancy. sPLS-DA builds models that fit the response variable (lesion severity or progression) with a minimal set of predictor variables (genes). The key finding was that using sPLS-DA we could distinguish lesions which progressed to cancer based upon the variant profile of a small number of genes in the first dysplastic lesions. In most analyses performed, samples from progressive patients showed a higher number of SNPs compared to their non-progressive counterparts. This is in line with previous evidence suggesting genetic instability to be an early event during progression [25,26].

We used longitudinal data of sequential samples from the same patients for pathway analysis, identifying the biological mechanisms most impacted by genetic change. While there were common pathways identified across the two cohorts, namely keratin filaments and DAP12 receptors in NK cells, there were also some notable distinctions. In particular pathways with significant DNA damage repair components were more prominent in progressive patients. As a consequence we curated a set of 148 genes from DNA damage repair pathways and examined the variant frequencies in first dysplastic lesions of progressive and non-progressive patients in a sPLS-DA analysis. We found that a subset of the genes could be used to classify progressive and non-progressive OLK in this cohort using sPLS-DA. Amongst the most discriminatory features were genes involved in double strand break (DSB)

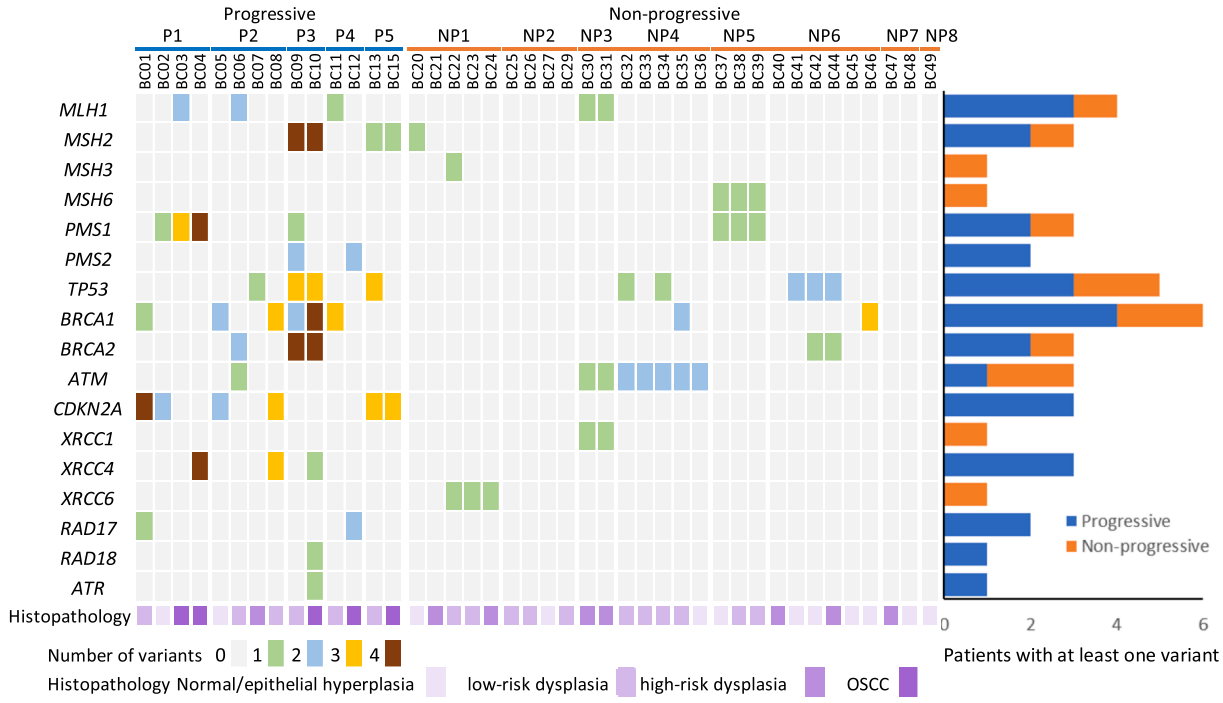

Fig. 3. Mutational profiles of the progressive and non-progressive cohort in key MMR and DSB pathway genes. *Top*, sample grouping according to patient and cohort. *Middle*, variant frequency for each gene in all patient lesion types. *Right*, bar chart of the number of patients with at least on variant per gene. *Bottom*, histopathological types of the samples for each case. Samples are displayed as columns.

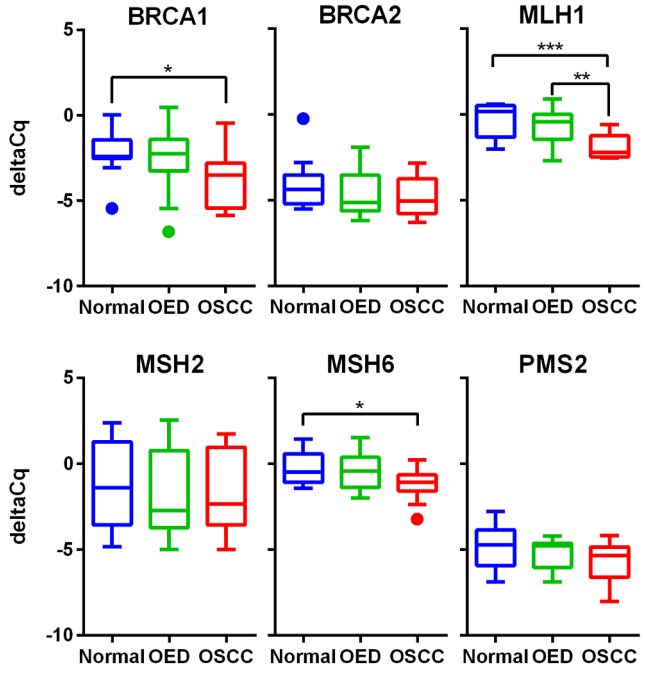

Fig. 4. Expression of selected DNA damage repair genes (*BRCA1, BRCA2, MLH1, MSH2, MSH6, PMS2*) from OSCC (n = 12), OED (n = 13) and normal (n = 11) tissue assayed by RT-qPCR using gene specific primers. Results are displayed as Tukey box and whiskers plots of Cq following normalization to reference genes (*UBC, GAPD, HPRT1*). Test for significance was by *t*-test with Welch's correction. P values are represented graphically as follows: $p < 0.001$ (***), $p < 0.01$ (**) and $p < 0.05$ (*).

repair pathways including prominently *BRCA1* and *BRCA2* as well as *FANCA* and *DCLRE1B*.

Numerous studies have shown mutations in DSB repair to increase individual risk for cancer and unrepaired DSB lesions accumulate with age [27,28]. We have shown differences in efficiency of DSB repair pathways in cell lines derived from different stages of oral tumorigenesis with neoplastic cell lines having the most defective DSB repair system [29]. Furthermore, disruption of FA/BRCA pathway has been implicated in malignant transformation [30]. The role of inherited mutations of *BRCA1/2* in risk for breast and ovarian cancer is well recognised [31]. However, there is also compelling evidence that mutations in FA/BRCA pathway genes contribute significantly to susceptibility to other cancers including head and neck SCC (HNSCC) [32–36].

In HNSCC, comparative genomic analysis has demonstrated that increased *FANCA* copy number is associated with poor prognosis [37]. Despite their well-studied role in carcinogenesis, at this stage, a detailed mechanism of contribution of these genes to progression of OLK/OED lesions is unknown. To our knowledge this is the first report of FA/BRCA pathway involvement in malignant transformation of OLK to OSCC. Apart from FA/BRCA pathway, genes from the MMR repair pathway were the major contributors to the classification of progressive lesions found in our sPLS-DA analysis of DNA repair genes. There is existing evidence from our laboratory and others identifying a potential role for defective MMR repair in OSCC [38–40]. In particular, this pathway may be a potential therapeutic target since MLH1, PMS2, MSH2 and MSH6 are downregulated in OED and OSCC [38,39], and MSH6 expression is lost in the stratum basale of carcinoma in situ [38].

Our study supports the call for a concerted research effort towards the generation of a Pre-Cancer Genome Atlas (PCGA), and the comprehensive genetic profiling of potentially malignant lesions performed over time, with associated histological and clinical outcomes [41]. The study of genomic alterations within dysplastic lesions has great potential to identify the principal factors that initiate oral carcinogenesis and drive it forward. Additionally, characterization of molecular alterations in potentially malignant lesions would hasten the development of biomarkers for early detection and risk stratification as well as suggest chemo-preventive interventions to reverse or delay the development of cancer [41]. To this end, drugs targeting DNA damage pathways taking advantage of clinical synthetic lethality have already shown therapeutic benefit; for example, the Poly (ADP-ribose) polymerase (PARP)

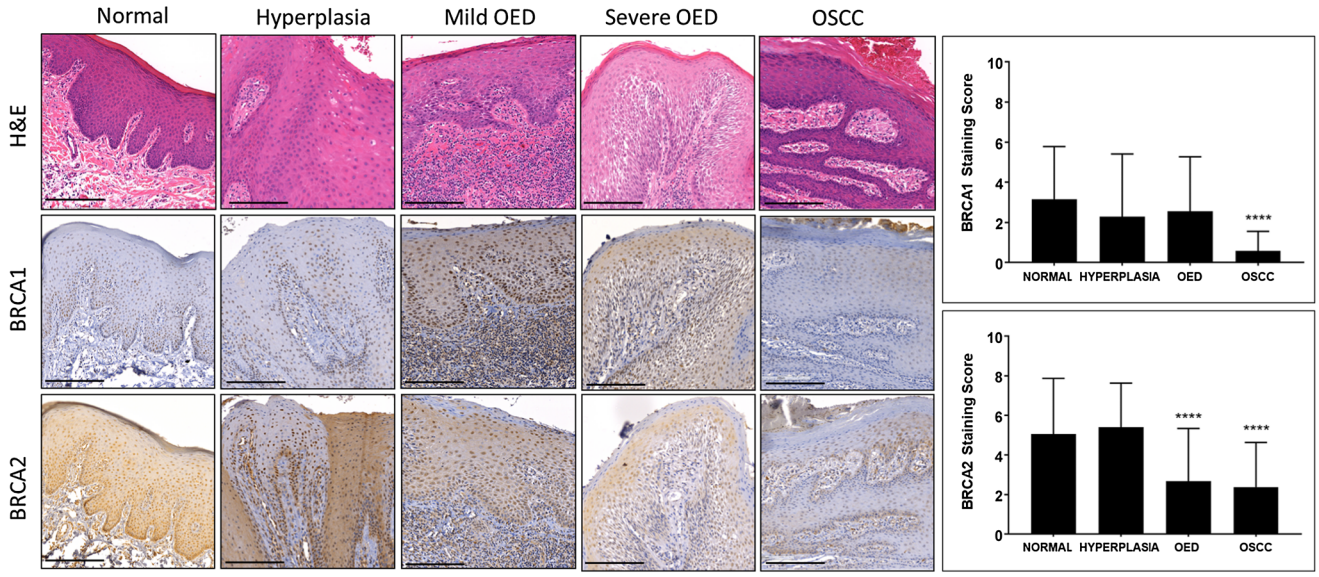

Fig. 5. FFPE sections were immunostained for BRCA1 (1:600) or BRCA2 (1:500) and assessed semi-quantitatively. Representative staining in normal, hyperplasia, mild OED, severe OED and OSCC. Scale bars indicate 200 μm. Bar graphs represent staining score ± SD. Differences between groups were assessed using 1-way ANOVA and Tukey's multiple comparison test. $p < 0.0001$ (****). One way ANOVA with Tukey's post-test, OSCC displays significantly reduced BRCA1 expression ($p < 0.0001$) compared with normal, hyperplasia and OED; while OED and OSCC display significantly reduced BRCA2 expression ($p < 0.0001$) compared with normal and hyperplasia cases.

inhibitor olaparib has shown benefit in BRCA-mutant ovarian, breast and prostate cancer [27].

Clearly, more work is required to understand the influence of identified variants and genes and their contribution to oral carcinogenesis. In this study we did not have normal matched controls or fresh frozen whole blood available for WES, but used stringent bioinformatic methods to remove false positive variations and overcome this. Lack of fresh tissues for sequencing is another limitation, but are difficult to obtain for a study design such as ours. The small sample size compromises the general applicability of our findings, but simultaneously draws the need for larger prospective multi-centre collaborative studies with robust validation and controls, and long term prospective follow-up [42]. Functional studies including the assessment of the tumour microenvironment should be undertaken, possibly with utility of single cell sequencing approaches. Given the relatively low prevalence of OLK, the scarcity of samples, and the extended time taken for malignant transformation to occur, the limitations outlined above are acceptable considering the insights gained from the current study.

Conclusion

In this WES study of progressive and non-progressive OLK, we found alterations in a number of genes that identify different grades of pathological changes in oral epithelium, and substantiated the role of accumulation of mutations in progression of dysplastic OLK to cancer. Importantly, we have revealed patients with progressive and non-progressive OLK can be differentiated using the frequency of exomic variants, particularly in DNA damage repair pathway genes. To our knowledge, this is the first report of FA/BRCA pathway involvement in malignant transformation of OLK to OSCC.

Declaration of Competing Interest

None declared.

Acknowledgment

The authors wish to thank Sarah Wagner, Yvette Emmanuel, Sandra Stein, Glen Francis (MACH-R Laboratory at the Princess Alexandra Hospital), Emma Campbell (Life Technologies), Yee Chee Yang and Russell McInnes (Agilent Technologies), Richard Allcock (University of Western Australia Lotterywest State Biomedical Facility: Genomics), Stephen Rudd, Anne Bernard (Queensland Facility for Advanced Bioinformatics), and Omar Kujan (UWA Dental School) for assistance with this project.

Funding

This work was funded by a grant from the Queensland Government Smart Futures Co-Investment Fund, supported by Cancer Australia, in collaboration with Life Technologies and Agilent Technologies awarded to CSF.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.oraloncology.2019.07.005.

References

[1] Weaver JMJ, Ross-Innes CS, Shannon N, Lynch AG, Forshew T, Barbera M, et al. Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. Nat Genet 2014.
[2] Yokota J. Tumor progression and metastasis. Carcinogenesis 2000;21:497–503.
[3] Speight PM, Torres-Rendon A. Oral epithelial dysplasia may progress to squamous cell carcinoma. Pathol Case Rev 2011;16:141–5.
[4] Dost F, Le Cao K, Ford PJ, Ades C, Farah CS. Malignant transformation of oral epithelial dysplasia: a real-world evaluation of histopathologic grading. Oral Surg Oral Med Oral Pathol Oral Radiol 2014;117:343–52.
[5] Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, et al. The mutational landscape of head and neck squamous cell carcinoma. Science (New York, NY). 2011;333:1157–60.
[6] Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science (New York, NY). 2011;333:1154–7.
[7] Gaykalova DA, Mambo E, Choudhary A, Houghton J, Buddavarapu K, Sanford T, et al. Novel insight into mutational landscape of head and neck squamous cell carcinoma. PLoS One 2014;9:e93102.
[8] Hedberg ML, Goh G, Chiosea SI, Bauman JE, Freilino ML, Zeng Y, et al. Genetic landscape of metastatic and recurrent head and neck squamous cell carcinoma. J Clin Investig 2016;126:1606.

[9] Jessri M, Farah CS. Next generation sequencing and its application in deciphering head and neck cancer. Oral Oncol 2014;50:247–53.
[10] Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. Cell 2012;149:994–1007.
[11] Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. N Engl J Med 1988;319:525–32.
[12] Agrawal N, Jiao Y, Bettegowda C, Hutess SM, Wang Y, David S, et al. Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. Cancer Discovery 2012;2:899–905.
[13] Barnes L, Eveson JW, Reichart P, editors. SD. World Health Organization Classification of Tumours. Pathology and Genetics of Head and Neck Tumours. 9 ed.Lyon: IARC Press; 2005.
[14] WHO. classification of head and neck tumours. Lyon: International Agency for Research on Cancer; 2017.
[15] Jordan RC, Lingen MW, Perez-Ordonez B, He X, Pickard R, Koluder M, et al. Validation of methods for oropharyngeal cancer HPV status determination in US cooperative group trials. Am J Surg Pathol 2012;36:945–54.
[16] Singhi AD, Westra WH. Comparison of human papillomavirus in situ hybridization and p16 immunohistochemistry in the detection of human papillomavirus-associated head and neck cancer based on a prospective clinical experience. Cancer 2010;116:2166–73.
[17] Rumble SM, Lacroute P, Dalca AV, Fiume M, Sidow A, Brudno M. SHRiMP: Accurate mapping of short color-space reads. PLoS Comput Biol 2009;5.
[18] Depristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491–501.
[19] Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr Protocols Bioinformat 2013.
[20] Mills RE, Pittard WS, Mullaney JM, Farooq U, Creasy TH, Mahurkar AA, et al. Natural genetic variation caused by small insertions and deletions in the human genome. Genome Res 2011;21:830–9.
[21] Rohart F, Gautier B, Singh A, Le Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol 2017;13:e1005752.
[22] Farah CS, Dalley AJ, Nguyen P, Batstone M, Kordbacheh F, Perry-Keene J, et al. Improved surgical margin definition by narrow band imaging for resection of oral squamous cell carcinoma: A prospective gene expression profiling study. Head Neck 2016;38:832–9.
[23] Dalley AJ, Abdul Majeed AA, Pitty LP, Major AG, Farah CS. LGR5 expression in oral epithelial dysplasia and oral squamous cell carcinoma. Oral Surg Oral Med Oral Pathol Oral Radiol 2015;119(436–40):e1.
[24] Chaw SY, Abdul Majeed A, Dalley AJ, Chan A, Stein S, Farah CS. Epithelial to mesenchymal transition (EMT) biomarkers–E-cadherin, beta-catenin, APC and Vimentin–in oral squamous cell carcinogenesis and transformation. Oral Oncol 2012;48:997–1006.
[25] Palmieri G, Ascierto PA, Cossu A, Colombino M, Casula M, Botti G, et al. Assessment of genetic instability in melanocytic skin lesions through microsatellite analysis of benign naevi, dysplastic naevi, and primary melanomas and their metastases. Melanoma Res 2003;13:167–70.
[26] Willenbucher RF, Aust DE, Chang CG, Zelman SJ, Ferrell LD, Moore 2nd DH, et al. Genomic instability is an early event during the progression pathway of ulcerative-colitis-related neoplasia. Am J Pathol 1999;154:1825–30.
[27] Minchom A, Aversa C, Lopez J. Dancing with the DNA damage response: next-generation anti-cancer therapeutic strategies. Therap Adv Med Oncol 2018;10. 1758835918786658.
[28] Jackson SP, Bartek J. The DNA-damage response in human biology and disease. Nature 2009;461:1071–8.
[29] Jessri M, Dalley AJ, Farah CS. Deficient double-strand break repair in oral squamous cell carcinoma cell lines. J Oral Pathol Med : Off Publ Int Associat Oral Pathol Am Acad Oral Pathol 2017;46:695–702.
[30] Nalepa G, Enzor R, Sun Z, Marchal C, Park SJ, Yang Y, et al. Fanconi anemia signaling network regulates the spindle assembly checkpoint. J Clin Investig 2013;123:3839–47.
[31] Friedenson B. BRCA1 and BRCA2 pathways and the risk of cancers other than breast or ovarian. MedGenMed Medscape General Med 2005;7.
[32] Wreesmann VB, Estilo C, Eisele DW, Singh B, Wang SJ. Downregulation of Fanconi anemia genes in sporadic head and neck squamous cell carcinoma. ORL J Oto-Rhino-Laryngol Relat Specialties 2007;69:218–25.
[33] Tremblay S, Pintor Dos Reis P, Bradley G, Galloni NN, Perez-Ordonez B, Freeman J, et al. Young patients with oral squamous cell carcinoma: study of the involvement of GSTP1 and deregulation of the Fanconi anemia genes. Arch Otolaryngol–Head Neck Surg 2006;132:958–66.
[34] Kietthubthew S, Sriplung H, Au WW, Ishida T. Polymorphism in DNA repair genes and oral squamous cell carcinoma in Thailand. Int J Hyg Environ Health 2006;209:21–9.
[35] Gal TJ. Erratum: DNA repair gene polymorphisms and risk of second primary neoplasms and mortality in oral cancer patients (Laryngoscope (December 2005) 115, 12, (2221–2231)). Laryngoscope 2006;116:507.
[36] Sparano A, Quesnelle KM, Kumar MS, Wang Y, Sylvester AJ, Feldman M, et al. Genome-wide profiling of oral squamous cell carcinoma by array-based comparative genomic hybridization. Laryngoscope 2006;116:735–41.
[37] Bauer VL, Braselmann H, Henke M, Mattern D, Walch A, Unger K, et al. Chromosomal changes characterize head and neck cancer with poor prognosis. J Mol Med 2008;86:1353–65.
[38] Jessri M, Dalley AJ, Farah CS. hMSH6: a potential diagnostic marker for oral carcinoma in situ. J Clin Pathol 2015;68:86–90.
[39] Jessri M, Dalley AJ, Farah CS. MutSalpha and MutLalpha immunoexpression analysis in diagnostic grading of oral epithelial dysplasia and squamous cell carcinoma. Oral Surg Oral Med Oral Pathol Oral Radiol 2015;119:74–82.
[40] Amaral-Silva GK, Martins MD, Pontes HA, Fregnani ER, Lopes MA, Fonseca FP, et al. Mismatch repair system proteins in oral benign and malignant lesions. J Oral Pathol Med : Off Publ Int Associat Oral Pathol Am Acad Oral Pathol 2017;46:241–5.
[41] Campbell JD, Mazzilli SA, Reid ME, Dhillon SS, Platero S, Beane J, et al. The case for a pre-cancer genome Atlas (PCGA). Cancer Prevent Res (Philadelphia, Pa). 2016;9:119–24.
[42] Bennett NC, Farah CS. Next-generation sequencing in clinical oncology: next steps towards clinical validation. Cancers 2014;6:2296–312.