

**Objectives:** Oral squamous cell carcinoma (OSCC) is associated with a high degree of morbidity and mortality. OSCCs are often preceded by oral potentially malignant disorders (OPMD) which have a higher propensity to undergo malignant transformation (MT) compared to clinically normal oral mucosa. Currently there is no reliable method to determine which OPMD cases will undergo MT. This study was performed to construct a prognostic classifier for patients with OPMD by integrating clinical, histopathological and molecular factors and to discover a gene expression signature that characterises OPMD with a high risk of undergoing MT.

**Findings:** Statistical analysis of an OPMD patient cohort (23 MT vs. 25 with no MT) showed that site of initial OPMD ( $p = 0.043$ ), binary oral epithelial dysplasia (OED) grading ( $p = 0.009$ ) and loss of heterozygosity at 3p/9p/17p ( $p = 0.026$ ) were statistically significant. Other demographic factors, clinical features and the WHO 3-tiered OED grading system were not statistically significant. Gene expression experiments revealed several genes that were differentially expressed between OPMD that underwent MT and those that did not [false discovery rate of  $< 0.05$ ]. Statistical model building was performed, and the outputs were used to construct a prognostic classifier.

**Conclusions:** Our findings show that a classifier combining histopathological and molecular factors outperforms conventional methods for prognosticating clinical outcome in patients with OPMD. We have also shown that formalin-fixed paraffin-embedded tissue can be used to generate a molecular classification with clinical utility.