



OPEN

Deregulation of extracellular matrix modeling with molecular prognostic markers revealed by transcriptome sequencing and validations in Oral Tongue squamous cell carcinoma

Soundara Viveka Thangaraj¹, Vidyarani Shyamsundar², Arvind Krishnamurthy³ & Vijayalakshmi Ramshankar¹✉

Oral Tongue Squamous Cell Carcinoma (OTSCC), a distinct sub-group of head and neck cancers, is characteristically aggressive in nature with a higher incidence of recurrence and metastasis. Recent advances in therapeutics have not improved patient survival. The phenomenon of occult node metastasis, even among the purportedly good prognosis group of early-stage and node-negative tongue tumors, leads to a high incidence of locoregional failure in OTSCC which needs to be addressed. In the current study, transcriptome analysis of OTSCC patients identified the key genes and deregulated pathways. A panel of 26 marker genes was shortlisted and validated using real-time PCR in a prospective cohort of 100 patients. The gene expression was correlated with clinicopathological features including occult node metastasis, survival, and therapeutic outcome. The up-regulation of a panel of 6 genes namely, matrix metalloproteinase 9 (MMP9), Laminin subunit Gamma 2 (LAMC2), Desmoglein 2 (DSG2), Plasminogen Activator Urokinase (PLAU), Forkhead Box M1 (FOXM1), and Myosin 1B (MYO1B) was associated with failure of treatment in the early stage (T1, T2). Up-regulation of Tenascin C (TNC) and Podoplanin (PDPN) was significantly correlated with occult node positivity. Immunohistochemical analysis of LAMC2, MMP9, and E-Cadherin (ECAD) confirmed these markers to be indicators of poor prognosis. We propose this panel of valuable prognostic markers can be clinically useful to identify poor prognosis and occult node metastasis in OTSCC patients.

Oral Tongue Squamous Cell carcinoma (OTSCC) represents a major portion of oral cavity cancers, especially in India. It is now evident that OTSCC needs to be studied as a separate entity and not be included in the broad scheme of oral cancers, as often reported—an approach that increases the power of the study but diminishes the credibility of the inference made from such studies. As per the data from National Cancer Registry Program (NCRP; www.ncdirindia.org), tongue cancer ranks 4th among the 10 leading sites of cancer among males in Chennai. There is a sharp increase in the incidence of tongue cancers not only in India¹ but also in young white women² and in Nordic countries³. Striking aspect of OTSCC epidemiology is the increase in incidence of OTSCC among young patients^{4,5} and women below 44 years of age^{6,7}.

OTSCC is an aggressive type of tumor with poor prognosis⁸ and the single most important aspect determining the staging, management and prognosis of patients with tongue cancers is the status of the cervical lymph nodes and occult node metastasis. However, the inaccuracy of clinical examination and imaging to reliably detect occult cervical lymph node micrometastasis has resulted in elective neck dissections becoming a standard of care for a vast majority of the OTSCC management, based on the evidences from randomised controlled trials⁹. The

¹Department of Preventive Oncology (Research), Cancer Institute (WIA), Adyar, Chennai 600 020, India. ²Center for Oral Cancer Prevention and Research (COCPAR), Sree Balaji Dental College, Chennai 600 100, India. ³Department of Surgical Oncology, Cancer Institute (WIA), Adyar, Chennai 600 020, India. ✉email: r.vijayalakshmi@cancerinstitutewia.org

| Up regulated genes | | | | Down regulated genes | | | |
|--|-------------|--------------------|----------|---|--------------|--------------------|----------|
| Gene name | Gene symbol | Fold change (Log2) | P value | Gene name | Gene symbol | Fold change (Log2) | P value |
| Matrix Metalloproteinase 12 | MMP12 | 8.61 | 1.44E-26 | Dopachrome tautomerase | DCT | -7.69 | 4.40E-08 |
| Matrix Metalloproteinase 9 | MMP9 | 8.29 | 7.64E-22 | Keratin 36 | KRT36 | -7.52 | 4.99E-16 |
| MAGE family member A3 | MAGEA3 | 8.85 | 3.08E-17 | Calmodulin like 6 | CALML6 | -7.51 | 5.35E-14 |
| MAGE family member A6 | MAGEA6 | 8.57 | 8.77E-16 | WNT inhibitory factor 1 | WIF1 | -6.61 | 6.14E-05 |
| Matrix Metalloproteinase 1 | MMP1 | 8.81 | 1.93E-15 | Myocilin | MYOC | -6.56 | 2.10E-07 |
| CXC motif chemokine ligand 1 | CXCL1 | 7.28 | 2.93E-15 | MT-CO1 pseudogene 3 | MTCO1P3 | -6.52 | 1.10E-05 |
| SPOC Domain containing 1 | SPOCD1 | 7.05 | 2.04E-13 | Tyrosinase | TYR | -6.32 | 2.87E-06 |
| Cyclic Nucleotide gated channel subunit Beta 1 | CNGB1 | 7.69 | 1.12E-11 | Paired Box 1 | PAX1 | -6.16 | 3.31E-07 |
| Inhibin subunit beta A | INHBA | 7.40 | 1.56E-11 | Cytochrome P450 Family 2 sub-family F member 3 | CYP4F35P | -6.15 | 9.69E-06 |
| Matrix Metalloproteinase 10 | MMP10 | 7.85 | 2.80E-11 | | RP11-641A6.8 | -6.15 | 1.37E-06 |
| Matrix Metalloproteinase 3 | MMP3 | 7.82 | 4.17E-11 | Leucine rich repeats and trans-membrane Domains 1 | LRTM1 | -6.13 | 1.74E-09 |
| C-C Motif chemokine Ligand 11 | CCL11 | 7.57 | 1.18E-10 | Keratin 72 | KRT72 | -6.03 | 6.38E-11 |
| Marginal Zone B and B1 cell specific protein | MZB1 | 7.01 | 8.81E-09 | Serine Peptidase inhibitor-9 | SPINK9 | -5.97 | 1.65E-05 |
| MAGE family member A4 | MAGEA4 | 6.97 | 1.60E-08 | Dehydrogenase/reductase 7C | DHRS7C | -5.87 | 5.43E-06 |
| F2R like thrombin or trypsin receptor 3 | F2RL3 | 6.80 | 2.85E-13 | Keratin 4 | KRT4 | -5.81 | 0.00E+00 |

Table 1. Top 15 upregulated and down regulated genes in OTSCC identified by transcriptome sequencing.

other clinicopathological factors predicting poor prognosis include increased depth of invasion^{10,11}, increasing pT, the presence of extra-capsular extension^{12,13} and endophytic appearance of tumor¹⁴. In spite of the multiple attempts made to classify molecular markers in OTSCC with promising candidates emerging¹⁵⁻¹⁹, they have not been translated to clinics and therapeutic decisions are still largely driven by the classical clinico-pathological factors like tumor stage, patient's age and performance status²⁰ which may be insufficient. Though numerous biomarkers have been presented as useful prognosticators for OTSCC, they have been unsatisfactory with no possible reliable conclusions. The alarmingly high rates of loco-regional recurrences seen in OTSCC warrants molecular markers that can predict patients at higher risk of morbidity and mortality and to design appropriate treatment approaches in patients.

Transcriptome sequencing (RNA-seq) as a platform of analysis not only has a better sensitivity and dynamic range than traditional microarrays, but also enables identification of novel transcripts and more exhaustive study of biological pathways²¹. In the current study, transcriptome sequencing (RNA seq) was done to identify the differentially expressed genes (DEG), in a set of OTSCC samples with validations by qPCR and Immunohistochemistry to derive salient markers for prognostication of OTSCC patients in a prospective study cohort.

Results

Transcriptome sequencing. Clinical characteristics of patients (n = 12) studied by transcriptome sequencing is shown in Supplementary Table S1. An average of 9.8 Gb of data was generated with more than 91% of bases passing Q30 Phred score. We identified a total of 3,705 (2,610 up-regulated, 1095 down-regulated) differentially expressed genes (DEG) (FDR < 0.05 and P value < 0.05) in tongue carcinoma samples. The quality of the results depends on the number of genes analyzed, which in turn is based on cut-offs and statistical significance. Hence for further functional analyses, we filtered only highly significant 2073 DEGs (1578 up-regulated, 495 down-regulated) with P value < 0.01 and Fold Change > 10. The heatmap representing the expression of top 100 DEG in OTSCC was derived from results obtained by transcriptome sequencing (Supplementary Figure S1).

The top 15 over expressed and under expressed genes are listed in Table 1 and all the highly significant 2073 DEGs are listed in Supplementary Table S2. The top up-regulated genes included many genes from the Matrix metalloproteinase family (MMP—1, 3, 9, 10, 12 and 19) and Melanoma Antigen family (MAGE A—3,4 and 6). The top down regulated genes included many genes from the Keratin family (KRT—4, 13, 36, 72 and 78).

Gene set enrichment analysis on DEG obtained by transcriptome sequencing reveals deregulated extracellular matrix (ECM) in OTSCC. To understand the biological roles of the DEG from OTSCC, we performed a gene set enrichment analysis on the DEG derived by transcriptome sequencing. We found GO terms for biological process enriched for Inflammatory response, Extracellular Matrix organization, Cell adhesion, Collagen Catabolic Process, Immune response and Angiogenesis. The top most significant enriched GO term for Cellular components were related to extracellular matrix remodeling. The top most significant enriched GO term for Molecular Function were Metalloendopeptidase activity, Heparin binding, Extracellular matrix structural constituent, Collagen binding and Receptor activity. The chief deregulated KEGG pathways included Cytokine-cytokine receptor interaction, Focal adhesion, ECM-receptor interaction and PI3K-Akt

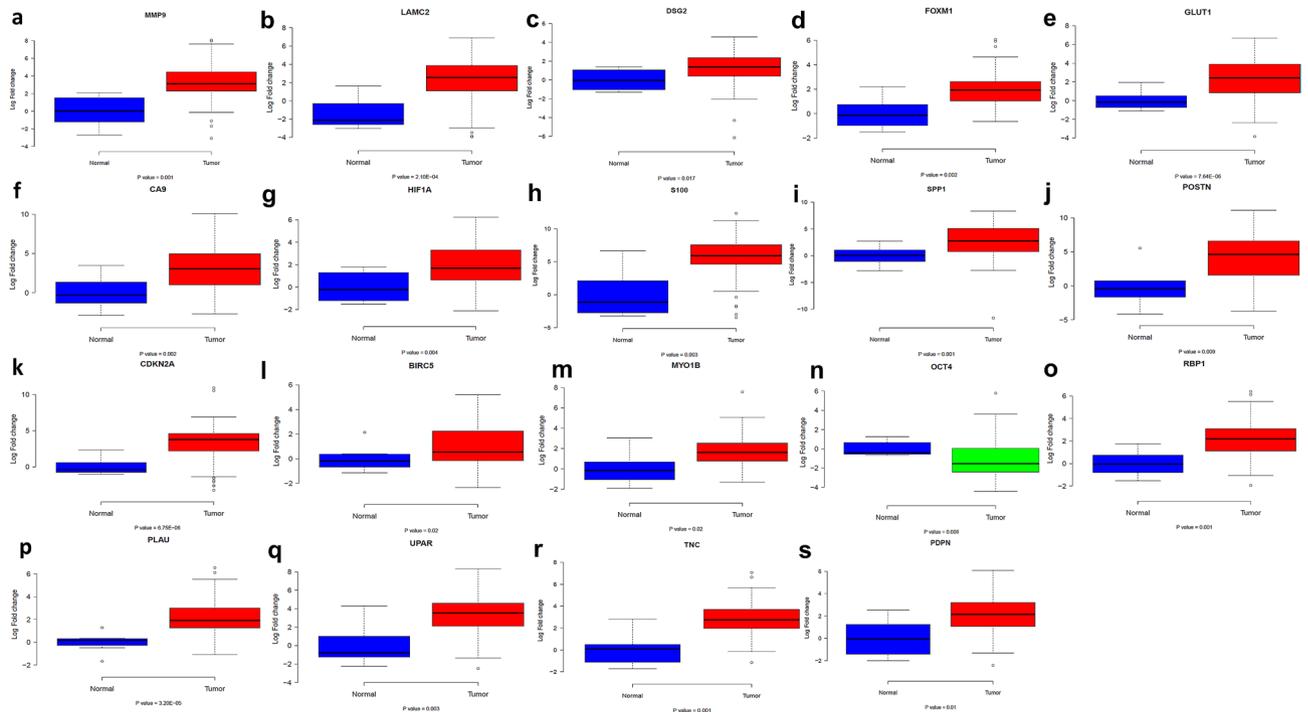


Figure 1. Box plot comparing the expression of genes in absolute normal tongue and OTSCC samples. Average \log_2 fold change from tumor and adjacent normal samples have been compared and the P value from an unpaired t test is indicated in the figure. (a) MMP9 ($P=0.001$) (b) LAMC2 ($P=2.10e-04$) (c) DSG2 ($P=0.017$) (d) FOXM1 ($P=0.002$) (e) GLUT1 ($P=7.64e-06$) (f) CA9 ($P=0.002$) (g) HIF1A ($P=0.004$) (h) S100 ($P=0.003$) (i) SPP1 ($P=0.001$) (j) POSTN ($P=0.009$) (k) CDKN2A ($P=6.75e-06$) (l) BIRC5 ($P=0.001$) (m) MYO1B ($P=0.02$) (n) OCT4 ($P=0.006$) (o) RBP1 ($P=0.001$) (p) PLAU ($P=3.20e-05$) (q) UPAR ($P=0.003$) (r) TNC ($P=0.001$) (s) PDPN ($P=0.01$).

signaling pathways. The GO enrichment and KEGG pathway for differentially expressed genes (DEG) is shown in Supplementary Figure S2.

Protein–protein interaction (PPI) network shows ECM and focal adhesion interaction pathways deregulated in OTSCC. The PPI network generated in STRING using the 1698 DEGs included 1392 nodes and 2516 edges with a PPI enrichment P value $< 1.0e-16$. The top 10 nodes are presented in Supplementary Table S3. The module analysis tool extracted 8 most important modules with MSCORE > 6 as sub-networks from the overall PPI network (Supplementary Table S4). The top hub genes, Amyloid Beta Precursor Protein (APP), G protein subunit beta 5 (GNB5) and G protein subunit Gamma Transducin 2 (GNGT2) were part of Module 1. Functional enrichment analysis of the top 2 modules identified the GO terms and KEGG pathways listed in Supplementary Table S5. Most significant module 1 was enriched for GO terms chemokine activity and G-protein coupled receptor binding pathway and for the KEGG pathways, Chemokine receptor signaling pathway and cytokine–cytokine interaction pathways. Module 2 showed enrichment of GO biological process terms Collagen catabolic process and Extracellular matrix organization while the KEGG pathways enriched consisted of ECM receptor interaction and Focal adhesion. Integrated sub network of the top most significant modules of OTSCC-PPI network is shown in Supplementary Figure S3.

Validation of differentially expressed proteins by qPCR. Biomarkers for a qPCR-based validation in prospective OTSCC patients ($n=100$) were chosen based on interrelationships assessed by Pearson correlation with respect to the genes involved in ECM remodeling and EMT. The markers chosen were LAMC2, TNC, PDPN, SPP1, PLAU, PLAUR, MMP9, FOXM1, VIM, CTNBN1, GLUT1, HIF1A, SOX2, OCT 4, VEGF, MYO1B, RBP1, SVN, CA9, DSG2, SURVIVIN, S100, TWIST2, CCND1, ECAD, POSTN. Figure 1 shows the box plot comparing the expression of genes in absolute normal tongue and OTSCC samples. We obtained 19 markers out of the 26 chosen to be significantly differentially expressed between tumor and adjacent, apparently normal tissue. We performed a qRT-PCR analysis of the 19 markers in tongue tumor and histologically normal tongue samples. There was a significant alteration in the mRNA expression of these markers between tongue tumor and normal tongue samples as depicted (Fig. 1). The biomarkers studied were based on having a significant association among the EMT markers—MMP9, SPP1, LAMC2, DSG2, S100 and TNC, hypoxia markers—HIF1 α , CA9 and GLUT1 and other markers including POSTN, CDKN2A, UPAR, PLAU, PDPN, RBP1, MYO1B, BIRC5, FOXM1 and OCT4 (Table 2).

The interrelationships derived is shown as follows. LAMC2 expression was most strongly correlated to TNC ($r = +0.5$, P value = $3.16E-08$), PLAU ($r = +0.5$, P value = $4.22E-07$), PDPN ($r = +0.5$, P value = $2.44E-08$), SPP1

| Gene name | Fold change (Log 2) | P value |
|---------------|---------------------|----------|
| MMP9 | 3.34 | 0.001 |
| LAMC2 | 2.29 | 2.10E-04 |
| DSG2 | 1.27 | 0.017 |
| FOXM1 | 1.27 | 0.002 |
| S100 | 5.67 | 0.003 |
| GLUT1 | 2.35 | 7.64E-06 |
| HIF1 α | 1.90 | 0.004 |
| CA9 | 3.19 | 0.002 |
| SPP1 | 2.84 | 0.001 |
| POSTN | 4.27 | 0.009 |
| BIRC5 | 1.10 | 0.021 |
| CDKN2A | 3.35 | 6.75E-06 |
| OCT4 | -1.03 | 0.006 |
| RBP1 | 2.16 | 0.001 |
| PLAU | 2.21 | 3.20E-05 |
| UPAR | 3.22 | 0.003 |
| TNC | 2.86 | 0.001 |
| PDPN | 2.04 | 0.01 |
| MYO1B | 1.75 | 0.015 |

Table 2. Differentially expressed genes associated with OTSCC assessed by qPCR.

($r = +0.49$, P value = $1.24E-07$) and PLAU ($r = +0.47$, P value = $4.22E-07$) while SPP1 correlated strongly to UPAR ($r = +0.5$, P value = $3.43E-08$) and PLAU ($r = +0.47$, P value = $3.23E-07$). MMP9 expression was positively related to FOXM1 ($r = +0.49$, P value = $1.02E-07$) and TNC ($r = +0.46$, P value = $6.67E-07$). VIM expression was strongly associated with CTNNA1 ($r = +0.6$, P value = $8.33E-12$) while DSG2 was similarly correlated to TNC ($r = +0.6$, P value = $1.76E-10$) and S100 to GLUT1 ($r = +0.5$, P value = $6.41E-08$). TNC was positively related to PDPN ($r = +0.5$, P value = $1.16E-08$), UPAR ($r = +0.6$, P value = $2.31E-12$), HIF1A ($r = +0.6$, P value = $1.05E-09$), PLAU ($r = +0.5$, P value = $4.61E-08$) and MYO1B ($r = +0.5$, P value = $5.78E-09$); PDPN to PLAU ($r = +0.48$, P value = $2.19E-07$), UPAR ($r = +0.5$, P value = $5.06E-08$) and MYO1B ($r = +0.48$, P value = $1.51E-07$). The hypoxia marker HIF1A was correlated to the expression of SOX2 ($r = +0.56$, P value = $2.56E-10$), PLAU ($r = +0.58$, P value = $4.81E-11$), FOXM1 ($r = +0.58$, P value = $4.17E-11$), MYO1B ($r = +0.53$, P value = $4.37E-09$), SVN ($r = +0.47$, P value = $2.48E-07$) and GLUT1 ($r = +0.48$, P value = $1.96E-07$). TWIST2 was positively linked to SOX2 ($r = +0.5$, P value = $3.83E-08$), OCT4 ($r = +0.48$, P value = $1.81E-07$) and MYO1B ($r = +0.47$, P value = $3.22E-07$); SOX2 strongly linked to OCT4 ($r = +0.56$, P value = $3.2E-10$) and OCT4 to VEGF ($r = +0.54$, P value = $1.35E-09$). Similarly, CA9 correlated with RBP1 ($r = +0.53$, P value = $4.89E-09$), PLAU ($r = +0.48$, P value = $2.18E-07$) and GLUT1 ($r = +0.57$, P value = $1.44E-10$), OCT4 to VEGF ($r = +0.54$, P value = $1.35E-09$) and SVN ($r = +0.63$, P value = $2.98E-13$); RBP1 to PLAU ($r = +0.69$, P value = $2.13E-12$), FOXM1 ($r = +0.49$, P value = $8.74E-08$), SVN ($r = +0.51$, P value = $2.79E-08$) and GLUT1 ($r = +0.51$, P value = $3.03E-08$); PLAU to FOXM1 ($r = +0.65$, P value = $4.71E-14$), MYO1B ($r = +0.57$, P value = $1.58E-10$), SVN ($r = +0.62$, P value = $1.15E-12$) and GLUT1 ($r = +0.47$, P value = $3.12E-07$) and FOXM1 correlated to SVN ($r = +0.51$, P value = $2.5E-08$) and GLUT1 ($r = +0.59$, P value = $1.13E-11$). Moderately strong to weak significant correlations was observed among the biomarker expression in the tongue tumors and listed in Table 3.

The clinical characteristics of prospective cohort of OTSCC patients used for validation studies ($n = 100$) is shown in Supplementary Table S6. Tissue extracts from OTSCC tumor and normal was prepared as described in Methods. The expression of the different markers was assessed for correlation with pathological stage, outcome and lymph node status. One patient was lost to follow-up and was excluded from the survival analyses. The same pattern of expression could be visualized in the heatmap of the biomarker expression as shown in Fig. 2 showing the expression levels of the significant differentially expressed marker genes among the normal and OTSCC samples.

Panel of 6 molecular markers useful in predicting recurrence among patients in early staged OTSCC. In the early staged patients, treatment failed in 22.5% (9) cases and 77.5% (31) cases showed NED. Among the panel of 19 biomarkers studied, we identified a panel of 6 genes showing overexpression namely, MMP9 (P value = 0.02), LAMC2 (P value = 0.02), DSG2 (P value = 0.02), PLAU (P value = 0.02), FOXM1 (P value = 0.02) and MYO1B (P value = 0.02) to be associated with failure of treatment in the early stage patients (Table 4). No such panel could be derived for advanced stages of OTSCC.

Biomarkers depicting occult nodal and cervical nodal status. Overexpression of TNC (P value = 0.05) and PDPN (P value = 0.05) significantly indicated occult node positive status while the overexpres-

| | LAM C2 | SP P1 | MM P9 | EC AD | V1 M | DS G2 | S1 00 | TN C | CTN NB1 | PD PN | UP AR | CCN D1 | HIF 1A | P1 6 | TW1 ST2 | SO X2 | C A9 | OC T4 | RB P1 | PL AU | VE GF | FOXMI | MYO 1B | SV N | POS TN | GLU T1 |
|---------|--------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|---------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|--------|
| LAM C2 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | |
| SPP1 | 0.49 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | |
| MMP 9 | 0.24 | 0.32 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | |
| ECA D | 0.40 | 0.42 | 0.03 | 1.00 | | | | | | | | | | | | | | | | | | | | | | |
| VIM | -0.10 | 0.26 | -0.15 | 0.26 | 1.00 | | | | | | | | | | | | | | | | | | | | | |
| DSG2 | 0.44 | 0.25 | 0.13 | 0.32 | 0.21 | 1.00 | | | | | | | | | | | | | | | | | | | | |
| S100 | 0.38 | 0.27 | 0.17 | 0.44 | -0.16 | 0.24 | 1.00 | | | | | | | | | | | | | | | | | | | |
| TNC | 0.50 | 0.44 | 0.46 | 0.27 | 0.14 | 0.57 | 0.39 | 1.00 | | | | | | | | | | | | | | | | | | |
| CTN NB1 | 0.16 | 0.37 | -0.15 | 0.41 | 0.60 | 0.26 | 0.08 | 0.30 | 1.00 | | | | | | | | | | | | | | | | | |
| PDPN | 0.51 | 0.43 | 0.41 | 0.10 | -0.02 | 0.43 | 0.23 | 0.52 | 0.10 | 1.00 | | | | | | | | | | | | | | | | |
| UPA R | 0.37 | 0.50 | 0.40 | 0.20 | -0.02 | 0.41 | 0.41 | 0.61 | 0.22 | 0.50 | 1.00 | | | | | | | | | | | | | | | |
| CCN D1 | 0.00 | 0.22 | 0.10 | 0.13 | 0.21 | 0.12 | 0.10 | 0.29 | 0.15 | 0.41 | 0.25 | 1.00 | | | | | | | | | | | | | | |
| HIF1 A | 0.30 | 0.44 | 0.43 | 0.42 | 0.11 | 0.32 | 0.39 | 0.55 | 0.25 | 0.40 | 0.36 | 0.46 | 1.00 | | | | | | | | | | | | | |
| P16 | 0.15 | 0.06 | 0.43 | 0.10 | 0.01 | 0.31 | 0.25 | 0.31 | -0.13 | 0.15 | 0.25 | 0.16 | 0.34 | 1.00 | | | | | | | | | | | | |
| TWIST2 | -0.14 | 0.16 | 0.23 | -0.02 | 0.16 | 0.08 | 0.15 | 0.15 | 0.02 | 0.22 | 0.09 | 0.35 | 0.42 | 0.26 | 1.00 | | | | | | | | | | | |
| SOX2 | -0.19 | -0.03 | 0.16 | 0.20 | 0.03 | 0.15 | 0.17 | 0.11 | -0.05 | -0.08 | -0.05 | 0.37 | 0.56 | 0.39 | 0.50 | 1.00 | | | | | | | | | | |
| CA9 | 0.36 | 0.19 | 0.35 | 0.17 | -0.26 | 0.16 | 0.43 | 0.29 | -0.14 | 0.28 | -0.26 | 0.10 | 0.43 | 0.34 | 0.29 | 0.21 | 1.00 | | | | | | | | | |
| OCT4 | -0.14 | 0.12 | 0.09 | 0.00 | 0.11 | -0.15 | 0.00 | -0.03 | 0.05 | -0.01 | -0.05 | 0.32 | 0.39 | 0.16 | 0.48 | 0.56 | 0.13 | 1.00 | | | | | | | | |
| RBP1 | 0.33 | 0.23 | 0.35 | 0.10 | -0.11 | 0.18 | 0.43 | 0.33 | -0.10 | 0.25 | 0.31 | 0.16 | 0.28 | 0.36 | 0.39 | 0.26 | 0.53 | 0.27 | 1.00 | | | | | | | |
| PLAU | 0.47 | 0.47 | 0.41 | 0.20 | -0.03 | 0.09 | 0.36 | 0.50 | 0.11 | 0.48 | 0.38 | 0.38 | 0.58 | 0.21 | 0.39 | 0.23 | 0.48 | 0.42 | 0.61 | 1.00 | | | | | | |
| VEGF | 0.11 | 0.08 | 0.02 | 0.03 | -0.04 | -0.10 | 0.14 | 0.03 | 0.10 | 0.08 | 0.09 | 0.42 | 0.34 | 0.13 | 0.31 | 0.36 | 0.27 | 0.54 | 0.31 | 0.42 | 1.00 | | | | | |
| FOXMI | 0.42 | 0.28 | 0.49 | 0.24 | -0.26 | 0.19 | 0.41 | 0.42 | -0.12 | 0.39 | 0.35 | 0.37 | 0.58 | 0.34 | 0.25 | 0.43 | 0.43 | 0.28 | 0.49 | 0.65 | 0.40 | 1.00 | | | | |
| MYO 1B | 0.37 | 0.41 | 0.23 | 0.15 | 0.06 | 0.28 | 0.41 | 0.53 | 0.24 | 0.48 | 0.44 | 0.19 | 0.53 | 0.18 | 0.47 | 0.22 | 0.43 | 0.28 | 0.44 | 0.57 | 0.33 | 0.42 | 1.00 | | | |
| SVN | 0.30 | 0.32 | 0.21 | 0.22 | 0.04 | 0.24 | 0.17 | 0.22 | -0.10 | 0.30 | 0.17 | 0.27 | 0.47 | 0.28 | 0.41 | 0.46 | 0.41 | 0.63 | 0.51 | 0.62 | 0.40 | 0.51 | 0.47 | 1.00 | | |
| POST N | 0.09 | 0.09 | 0.27 | 0.02 | 0.10 | 0.23 | 0.04 | 0.04 | -0.01 | 0.22 | 0.10 | 0.23 | 0.27 | 0.31 | 0.27 | 0.32 | 0.09 | 0.17 | 0.25 | 0.17 | 0.24 | 0.17 | 0.31 | 1.00 | | |
| GLU T1 | 0.40 | 0.24 | 0.30 | 0.42 | -0.14 | 0.36 | 0.50 | 0.36 | 0.00 | 0.26 | 0.30 | 0.22 | 0.48 | 0.31 | 0.30 | 0.32 | 0.57 | -0.01 | 0.51 | 0.47 | 0.18 | 0.59 | 0.41 | 0.37 | 1.00 | |

Values in bold indicate a P value <0.0001

Table 3. Inter-relationship between molecular markers in OTSCC as represented by Pearson’s correlation coefficient. Values in bold indicate a P value <0.0001.

sion of LAMC2 (P value = 0.02), VIM (P value = 0.001), TWIST2 (P value = 0.03) and SOX2 (P value = 0.02) were the significantly associated with the cervical lymph node involvement.

Molecular markers useful in differentiating the clinical stages in OTSCC. The relative expression of LAMC2 (P value = 0.02), VIM (P value = 0.02), HIF1A (P value = 0.01), TWIST2 (P value = 0.03) and SOX2 (P value = 0.02) were found to be significantly differentially expressed among the early and advanced stage tumors.

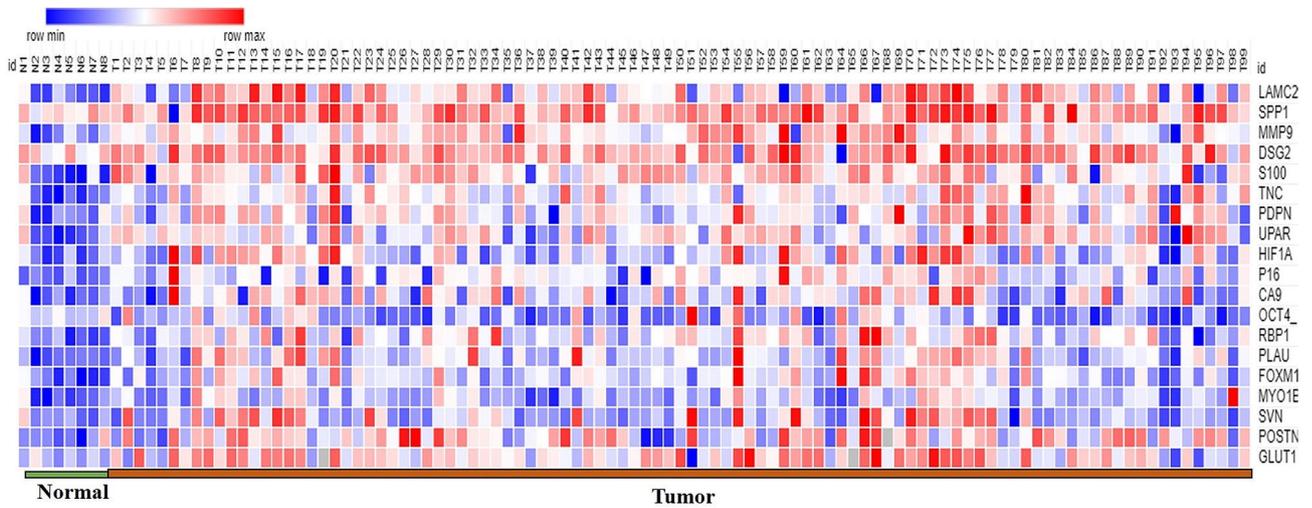


Figure 2. Heatmap representing the expression levels of the significant differentially expressed marker genes among the normal and OTSCC samples.

| | Gene name | P value | F value |
|--|-----------|---------|---------|
| Early stage versus advanced stage | LAMC2 | 0.016 | 14.32 |
| | VIM | 0.019 | 0.24 |
| | HIF1A | 0.012 | 2.39 |
| | TWIST2 | 0.029 | 2.65 |
| | SOX2 | 0.024 | 2.05 |
| Cervical node positive versus negative | LAMC2 | 0.017 | 4.18 |
| | VIM | 0.001 | 0.001 |
| | TWIST2 | 0.033 | 1.87 |
| | SOX2 | 0.024 | 0.64 |
| Occult node positive versus negative | TNC | 0.049 | 6.76 |
| | PDPN | 0.049 | 0.5 |
| Early staged OTCC prognostic panel | MMP9 | 0.016 | 0.1 |
| | LAMC2 | 0.022 | 10.7 |
| | DSG2 | 0.038 | 1.21 |
| | PLAU | 0.046 | 1.06 |
| | FOXM1 | 0.008 | 0.99 |
| | MYO1B | 0.027 | 0.89 |

Table 4. Molecular markers associated with stage, outcome and cervical node status.

The differential expression of potential markers between early versus advanced staged OTSCC has been shown (Supplementary Figure S4).

Validation of protein expression by immunohistochemistry. Here, we present the results of protein expression of 3 biomarkers namely LAMC2 (Fig. 3), ECAD (Fig. 4) and MMP9 (Fig. 5) among the others which have been evaluated. The correlation of the IHC markers with various clinic-pathological parameters like age, sex, tobacco habits, alcohol habit, pathological stage, tumor grade, cervical node status, occult node status, perineural invasion status and type of growth was investigated.

Panel of poor prognostic molecular markers in OTSCC validated by IHC for treatment outcome, pattern of failure, cervical lymph node involvement and locally advanced disease. As shown by the qPCR-based mRNA expression studies, the immunohistochemical analysis also confirmed that LAMC2 ($\chi^2 = 9.45, P = 0.002$), MMP9 ($\chi^2 = 20.33, P < 0.001$) and ECAD expression at ITF ($\chi^2 = 18.05, P < 0.001$) had a significant correlation with the treatment outcome. LAMC2 ($\chi^2 = 12.97, P = 0.024$) and MMP9 expression ($\chi^2 = 20.5, P = 0.001$), also correlated with the pattern of failure. Altered ECAD expression at ITF and MMP9 overexpression denoted nodal positivity and poor outcome in OTSCC. The pattern of ECAD expression at ITF ($\chi^2 = 6.0, P = 0.01$) and over-expression of MMP9 ($\chi^2 = 7.09, P = 0.008$) correlated with the cervical node status.

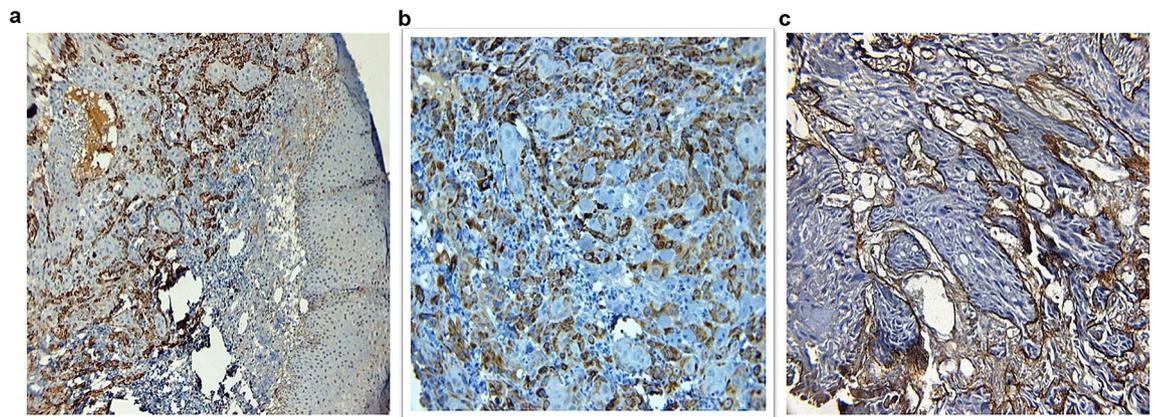


Figure 3. Immunorexpression of LAMC2: (a) WDSKC showing intense cytoplasmic positivity for LAMC2 10X. (b) WDSKC showing intense cytoplasmic positivity for LAMC2 20X. (c) MDSKC showing intense positivity for basement membrane, but cytoplasm in negative for LAMC2, 20X.

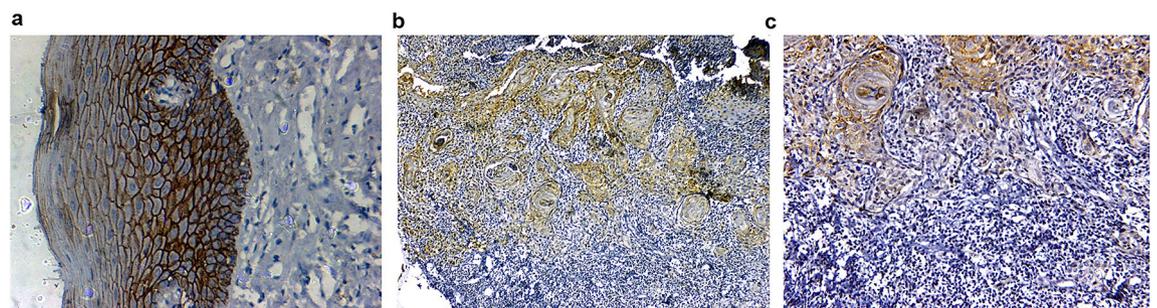


Figure 4. Immunorexpression of ECAD (a). Normal epithelium with intense E-cadherin positivity (b). Reduction of ECAD expression at ITF 10X (c). Reduction of ECAD expression at ITF 20X.

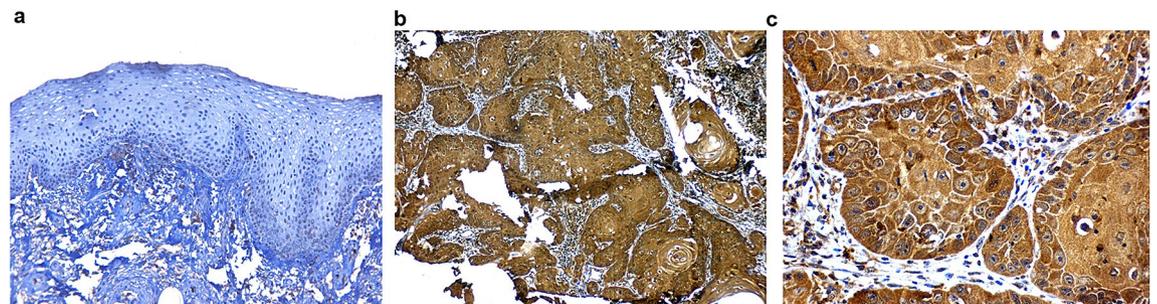


Figure 5. Immunorexpression of MMP9 (a). Normal mucosa showing minimum stain for MMP9 (b). WDSKC showing intense MMP9 stain 10X (c). WDSKC showing intense MMP9 stain 20X.

MMP9 also correlated with the locally advanced stage of tumor ($\chi^2 = 7.09$, $P = 0.008$) with higher MMP9 expression seen in 84.4% of all locally advanced tumors.

Clinico-pathological factors and molecular markers influencing tumor recurrence and as predictors of death analysed by univariate and multivariate analyses. The DFS median was 24 months with a range of 0 to 54 months while the median for OS was 30 months with a range of 4 to 74 months. Univariate and multivariate analyses were performed to identify the risk factors for recurrence of disease and death due to cancer using the Cox proportional hazard regression model (Table 5). Kaplan Meier survival analyses were also done to identify the significant predictors of DFS and OS.

Among all the patients, the classical prognostic factors like habits—tobacco chewing ($P = 0.01$), alcohol ($P = 0.04$), clinical factors like node status ($P = 0.003$), stage ($P = 0.01$) and occult node status ($P = 0.005$) were found to be significant predictors of disease recurrence in OTSCC. Tobacco chewing, positive nodal status and advanced stage of tumor increased the hazard of tumor recurrence by more than two-fold while a positive occult node increased the risk of recurrence by 4.27-fold. The same clinico-pathological parameters were also significant predictors of death. Tobacco chewing ($P = 0.02$) and alcohol ($P = 0.01$) were associated with an increased

| Variables | Disease-free survival | | | Overall survival | | |
|--------------------|--------------------------|--------------|------------------|--------------------------|--------------|------------------|
| | Hazard ratio (95% C.I.) | P value | Events | Hazard ratio (95% C.I.) | P value | Events |
| Age | 0.91 (0.45–1.84) | 0.8 | 9 (47.4) | 1.07 (0.47–2.43) | 0.87 | 7 (36.8) |
| Gender | 0.73 (0.35–1.51) | 0.4 | 10 (43.5) | 0.94 (0.43–2.05) | 0.87 | 9 (39.1) |
| Tobacco chewing | 2.29 (1.2–4.34) | 0.01 | 35 (61.4) | 2.45 (1.16–5.21) | 0.02 | 30 (53.6) |
| Tobacco smoking | 1.39 (0.77–2.5) | 0.28 | 19 (63.3) | 1.74 (0.91–3.35) | 0.09 | 17(56.7) |
| Alcohol | 1.85 (1.04–3.29) | 0.04 | 24 (61.5) | 2.36 (1.23–4.55) | 0.01 | 22 (57.9) |
| Habits | 1.46 (1–2.13) | 0.05 | 22 (55) | 1.78 (1.12–2.84) | 0.01 | 24 (58.5) |
| Node status | 2.6 (1.39–4.88) | 0.003 | 35 (67.3) | 2.67 (1.29–5.52) | 0.008 | 29 (58) |
| Stage | 2.16 (1.18–3.96) | 0.01 | 34 (66.7) | 2.45 (1.21–4.96) | 0.01 | 29 (59.2) |
| Grade | 1.18 (0.77–1.82) | 0.45 | 10 (66.7) | 1.68 (1.02–2.78) | 0.04 | 10 (66.7) |
| PNI | 0.99 (0.44–2.22) | 0.99 | 7 (46.7) | 0.85 (0.33–2.17) | 0.73 | 5 (38.5) |
| Occult node status | 4.27 (1.53–11.91) | 0.005 | 8 (53.3) | 5.19 (1.46–18.46) | 0.01 | 6 (46.2) |
| LAMC2 | 2.69 (1.38–5.22) | 0.004 | 33 (66) | 2.91 (1.36–6.21) | 0.006 | 27 (54) |
| ECAD at ITF | 3.19 (1.64–6.23) | 0.001 | 31 (75.6) | 3.11 (1.48–6.51) | 0.003 | 24 (58.5) |
| MMP9 | 5.39 (1.88–15.48) | 0.002 | 31 (83.8) | 3.09 (1.07–8.9) | 0.04 | 25 (67.6) |

Table 5. Univariate survival analysis for disease-free and overall survival in all patients. Factors denoting increased hazard of death and recurrence in OTSCC; Numbers in brackets denote percentages. The numbers in bold denote values that are of statistical significance.

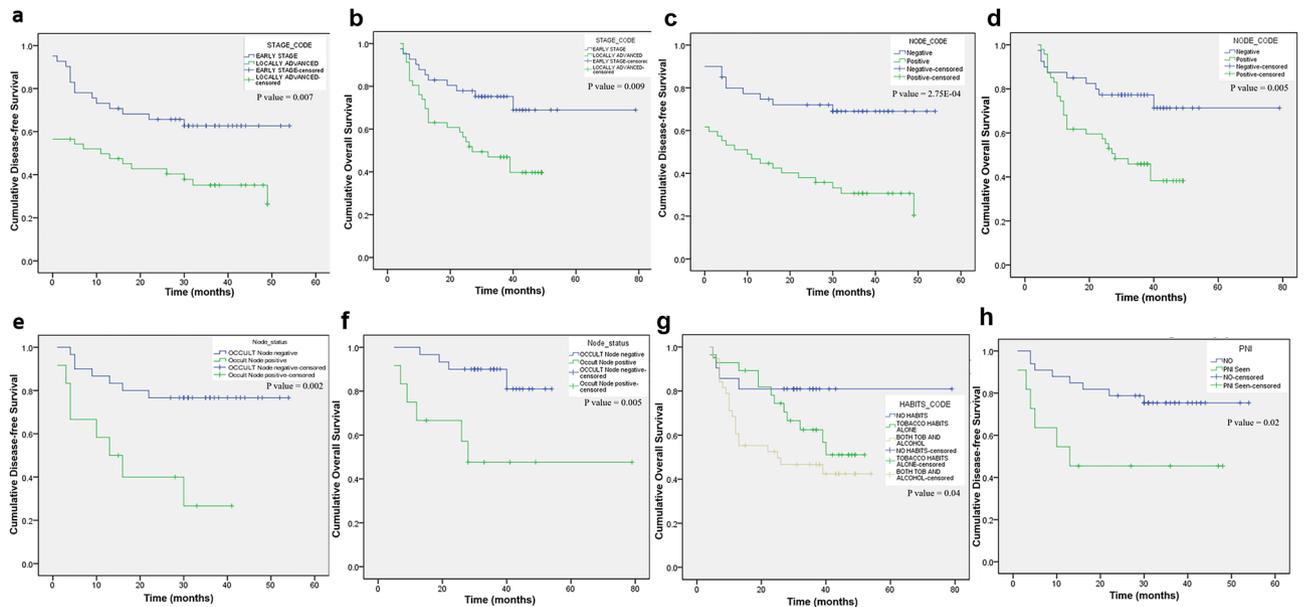


Figure 6. Kaplan–Meier survival analysis of clinicopathological variables in OTSCC patients (a). Disease free survival by clinical stage (b). Overall survival by clinical stage (c). Disease free survival by neck node status (d). Overall survival by neck node status (e). Disease free survival by occult node status (f). Overall survival by occult node status (g). Overall survival by habits status (h). Disease free survival by Perineural invasion in early stage OTSCC treated by surgery. P values correspond to the log-rank test comparing the survival curves.

hazard of death as seen by the HR of more than 2. Similarly, positive cervical node status ($P=0.008$), advanced clinical stage ($P=0.01$) and grade ($P=0.04$) of the tumor were also significant predictors of death. Occult node positivity increased the risk of death by 5.19-fold and was the most significant clinic-pathological parameter among all factors. Kaplan–Meier survival analysis also confirmed the parameters clinical stage, node, presence of perineural invasion and occult node status to be predictors of disease recurrence and death (Fig. 6) with the presence of tobacco habits identified as an additional predictor of death.

Among the molecular markers, high LAMC2 expression (Fig. 7a,b) was associated with an increased risk of recurrence ($P=0.004$) and an increased risk of death ($P=0.006$). Loss of membrane positivity of E-cadherin was associated with an increased hazard of disease-recurrence as seen by the increased HR of 3.19 for DFS ($P=0.001$), when compared to no E-cadherin expression. It was also a significant predictor of death ($P=0.003$) with a HR of 3.11. High MMP9 expression was associated with an increased risk of both disease recurrence ($P=0.002$) and death ($P=0.004$). These results were also mirrored by the Kaplan–Meier survival analysis of factors predicting DFS and OS (Fig. 7c,d) (Supplementary Table S9). Among the molecular markers, ECAD expression pattern

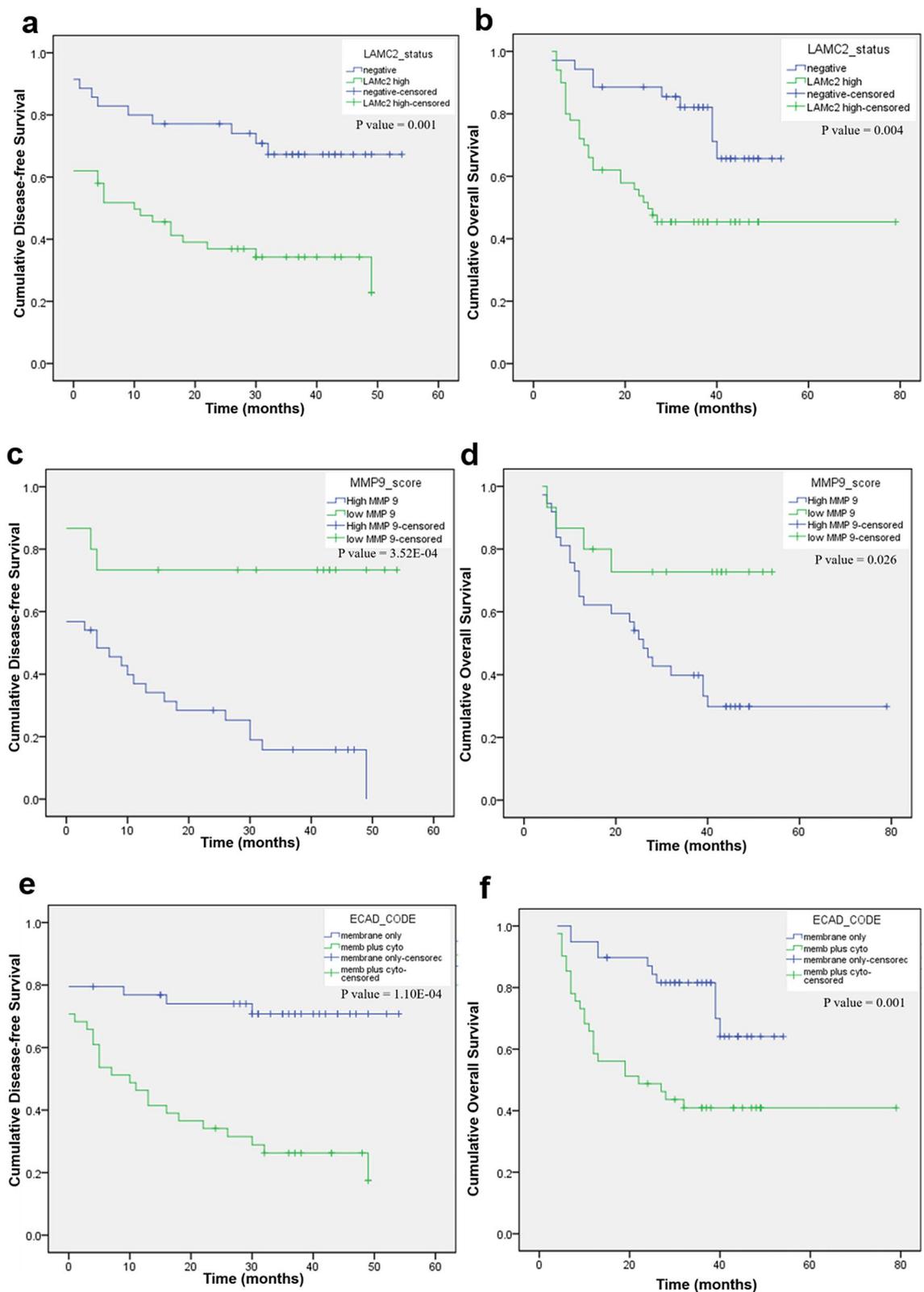


Figure 7. Kaplan–Meier survival analysis of molecular markers in OTSCC patients (a). Disease free survival by LAMC2 immunoexpression (b). Overall survival by LAMC2 immunoexpression (c). Disease free survival by MMP9 immunoexpression (d). Overall survival by MMP9 immunoexpression (e). Disease free survival by ECAD immunoexpression (f). Overall survival ECAD immunoexpression. *P* values correspond to the log-rank test comparing the survival curves.

at the ITF was the most significant predictor of both disease recurrence ($P=0.002$) and death ($P=0.03$) with an increased HR of 25.39 and 10.81 respectively. These results were confirmed by the Kaplan–Meier survival analysis (Fig. 7e,f).

Discussion

OTSCC is an aggressive cancer with high invasion and metastatic properties. Lymph node metastasis is the single most important prognostic indicator, however, in early stages, the lymph node metastasis is occult without clinical signs. Though, elective neck dissection is recommended for early stages, majority of the patients may not benefit substantially due to the surgery related morbidities. Biomarkers are therefore important in OTSCC and more importantly in early stages to prevent mortality and in identifying the high-risk patients. There is a need for reliable tests to detect aggressive early stage OTSCC to provide adequate treatment for these patients.

The findings of our current transcriptomic sequencing study were in complete agreement to our published meta-analysis study on OTSCC²². This study shows that role of tumor microenvironment in OTSCC with a number of extracellular matrix (ECM) components playing a crucial role in patient prognosis. Based on the data obtained from RNA-seq, the top up regulated genes included members of Matrix metalloproteinases and Melanoma associated gene families. The Gene Ontology terms enriched among the DEGs included inflammatory response, ECM organization, cell adhesion, collagen catabolic process and metalloendopeptidase activity. Studies have shown that biomarkers of stromal microenvironment might have a greater impact on prognosis compared to biomarkers related to tumor cells²³. We found ECM receptor interaction and Focal adhesion as the most significant deregulated pathways in OTSCC. However, biomarkers related to tumor microenvironment have not been widely studied except activin A²⁴. Degradation of the ECM that expedites the movement of tumor cells to surrounding tissue, is one of the important hallmarks of cancer progression and is facilitated by collagenases. Focal adhesion kinase is a tyrosine kinase that mediates signaling by the integrin family of cell surface receptors for ECM, leading to tumor cell migration and progression²⁵. The module analysis additionally identified Neuroactive ligand–receptor interaction pathway, which is implicated in OTSCC²⁶ and other types of cancers, like breast²⁷ and bladder²⁸ to be dysregulated in OTSCC. Collectively, these results complement each other and reinforce the pivotal role played by ECM remodeling and stromal microenvironment in OTSCC which in turn clarifies the aggressive nature of the disease and its propensity to metastasize.

The hub genes in a network are expected to be key drivers influencing the normal functioning of a cell, owing to the central role played by them in the network with multiple interactions²⁹. APP was identified as the most significant hub gene in both transcriptomic and meta-analysis studies²³. APP, a type-I integral membrane protein, has shown higher expression levels in OSCC and is known to be associated with poor prognosis³⁰. Secretory proteins assume a lot of importance in the context of a high-effect local disease like oral cancer owing to its potency to be detected in saliva. With more than 500 of the DEGs in the current study being secreted proteins, patient-based saliva proteomics could be an engaging approach to biomarker discovery in oral cancer³¹. SPOCK1 was found to be up-regulated in OTSCC but its mode of action and its role in OSCC has not been explored and merits an exhaustive analysis. SPOCK1 has the potential to act as a prognostic marker and the added advantage of it being a secretory protein, offers the possibility of non-invasive detection.

While surgery is the accepted first line of treatment in OTSCC, there exists several views over the clinical management of node negative patients with suggestions ranging from observation to elective neck dissection^{32–35}. However there has been no single, decisive prognostic/predictive factor for a positive occult node³² and clinical staging is insufficient and often underestimates the extent of occult node positivity³⁶. These shortcomings have called for an exigent need to identify molecular markers to predict occult tumor burden and our study has attempted to answer it.

Our cohort comprised of stage I to stage IV cancers and was representative of the entire spectrum of OTSCC. Out of the 26 markers evaluated, 19 of the tested biomarkers were able to significantly differentiate tumor from normal samples. We report 12 genes namely, LAMC2, VIM, HIF1A, TWIST2, SOX2, TNC, PDPN, MMP9, DSG2, PLAU, FOXM1 and MYO1B found to be relevant clinical markers that can be easily tested in OTSCC tumors to indicate prognosis, cervical and occult node positivity and outcome. We have derived a panel of molecular markers significantly predicting recurrence of the early stage OTSCC based on the gene expression levels of 6 genes namely, MMP9, LAMC2, DSG2, PLAU, FOXM1 and MYO1B.

Several studies including our previous study have shown that members of MMP family are involved in OTSCC with very significant upregulation^{23,37–39}. Additional studies by Zhou et al., showed up-regulation of the MMP9 gene to be associated with advanced OTSCC and having a predictive value for the identification of lymph node metastasis⁴⁰. These evidences along with our findings emphasize the importance of evaluating MMP9 in OTSCC.

LAMC2 was identified as a relevant gene and an important indicator of poor prognosis and cancer progression in OTSCC^{41,42} with its overexpression indicating an advanced stage and positive cervical node status. Our results are in confirmation of earlier studies identifying LAMC2 as a prognostic factor in early stage and advanced stage OTSCC tumors⁴³. Interestingly, in our cohort, the tumor area facing the stroma had a higher expression of LAMC2. Most of the extracellular spaces of the tumor areas showed accumulation of LAMC2 and expression of LAMC2 at tumor stroma interfaces indicates invasion and cancer progression⁴⁴.

Other relevant genes were DSG2, and PLAU. DSG2 is an adhesion molecule and a known epithelial EMT marker overexpressed and is also known for bad prognosis marker in hepatocellular carcinoma⁴⁵ and gastric carcinoma⁴⁶. Our observation of PLAU as a marker of relapse in OTSCC is in line with prior study in OSCC demonstrating the role of PLAU as a strong independent prognostic factor of DFS⁴⁷. Baker et al., have previously reported the over-expression of PLAU and its correlation to aggression of tumors in OSCC⁴⁸.

Several studies have implicated FOXM1 to play a major role in chemo-resistance in gastric cancer^{49,50}, glioblastoma⁵¹, non-small-cell lung cancer⁵² and most recently, also in oral cancer where it is proven to predict

poor patient survival⁵³. This is similar to our findings where FOXM1 overexpression is associated with relapse in OTSCC patients. MYO1B, a motor protein involved in cellular motility⁵⁴, is known to be overexpressed in HNSCC⁵⁵ and is found to play a pivotal role in lymph node metastasis by way of augmenting cancer cell motility⁵⁶. SOX2 is a transcription factor and stem cell marker, known to be expressed in several cancers, playing a major role in cell proliferation, metastasis⁵⁷, tumor drug resistance⁵⁸ and is hence a valuable therapeutic target⁵⁹. SOX2 maintains the stemness of CSC, ultimately resulting in cancer relapse and resistance to treatment⁶⁰. Liu et al. demonstrated the involvement of SOX2 in promoting tumor aggressiveness and EMT in OTSCC⁶¹. In this scenario, our finding of SOX2 to be a marker of failure in OTSCC cases treated by radiotherapy is significant.

We identified TNC and PDPN to be significant markers of occult node metastasis. TNC, a protein of the extracellular matrix with important functions in angiogenesis and tumor cell proliferation⁶², has been established as a diagnostic marker and potential cancer-associated fibroblast marker in breast ductal cancer⁶³ and cervical cancer⁶⁴ and a prognostic marker in early stage OTSCC⁶⁵. The functions of TNC and its role as a cancer stem cell (CSC) marker⁶⁶ makes it an ideal marker of occult node positivity as found in our study. Another important marker of occult node positivity and poor outcome identified in radiation-treated locally advanced OTSCC was PDPN. PDPN is a trans-membrane receptor glycoprotein that increases clonal cell capacity, EMT, invasion and metastasis⁶⁷, playing a pivotal role in several cancers. Preclinical studies have identified PDPN as a therapeutic target to combat several cancers including oral cancer^{68–71} and nasopharyngeal cancer⁷². These findings corroborate our identification of PDPN as an occult node and poor prognosis predictor.

Our cohort had about 26.9% of occult positive cases that falls within the range of previous studies reported³⁴. In another study on OTSCCs from our centre, the incidence of occult metastasis was 32.69%⁷³. Currently, sentinel Lymph node biopsy has been suggested as a minimally invasive technique for nodal staging to reduce the needless morbidity of nearly 70–75% of pathologically node negative patients undergoing elective neck dissection. Since, sentinel node is generally believed to be the first lymph node or lymph nodes group receiving lymphatic drainage from the primary tumor, if this node is metastasis negative, the non-sentinel nodes in the neighboring regional basins are also deemed to be negative of metastases. Currently, there are no molecular markers that can identify metastasis in lymphatic basins, and we suggest that TNC and PDPN expression can be evaluated in a larger series of cases to identify patients at risk of occult micrometastasis.

Survival analyses of the clinicopathological factors predicting recurrence and death identified clinical stage, cervical node, occult node and tobacco habits as the classic prognostic factors in tongue cancer^{74,75}. Our validations based on protein expression predicting risk of disease relapse and death in OTSCC patients confirmed the expression of LAMC2, MMP9 and ECAD at ITF as the important prognostic indicators. Thus, among our panel of markers, MMP9, LAMC2, VIM, DSG2 and TNC are well-known EMT markers, emphasizing the importance of EMT in OTSCC.

In the exclusive cohort of early stage patients treated by surgery, cervical node and peri-neural invasion significantly predicted failure in early stage OTSCC as reported earlier^{76,77}. Loss of membrane ECAD at ITF predicted both risk of recurrence and death. Among the locally advanced cases treated by radiation, LAMC2, MMP9 and POSTN were the predictors of disease relapse while node, LAMC2 and POSTN were the prognostic factors predicting death.

Murthy et al. reported node to be an important prognostic factor in oral cancer patients treated by radiotherapy and went on to state that among the different oral subsites, OTSCC patients treated by radiotherapy perform poorly⁷⁸. Our locally advanced tumors were treated by radiation and POSTN turned out to be an important prognostic factor in these tumors. High levels of POSTN are found to be associated with aggressive tumor behaviour, tumor progression, resistance to treatment and hence, bad prognosis^{79–81}. Although Choi et al. have identified POSTN to be a diagnostic marker in OSCC by tissue microarray⁸² and its role in OSCC tumorigenesis is well established^{83,84}, its prognostic role in OTSCC has not been reported earlier.

Although many earlier studies have reported diagnostic biomarkers based on mRNA expression in OSCC⁸⁵, very few have derived meaningful mRNA-based prognostic markers⁸⁶, more so in OTSCC. Using a panel of markers has superior discriminatory power over using individual candidate genes and hence, our panel of biomarkers with further validations in a larger study, may have better utility in the clinical setting to indicate the risk of recurrence in OTSCC.

Conclusion

OTSCC transcriptome sequencing upholds prior findings on the pivotal role of ECM degradation in OTSCC tumorigenesis and progression. Our cohort study addressed the most pressing issue of OTSCC management, namely, occult node metastasis prediction with 2 molecular markers and identified a panel of 6 molecular markers that can be used to differentiate the early stage tumors at higher risk of recurrence. Identification of novel biomarkers for early detection, prediction of response, and for use as treatment targets is of utmost importance to increase survival in OTSCC and the current study has fulfilled that. With further validation in a larger cohort, preferably, a multi-centric one, the panel of prognostic markers ascertained in our study can be used routinely to make important clinical decisions with respect to treatment modalities.

Methods

Patient materials. The Ethical approval number of the study was CIWIA/Protocol 1/HNCOG 2014, namely Cancer Institute Womens India Association/ Protocol 1 /Head and Neck Oncology Group 2014). All research involving human participants had been approved by the authors' Institutional Review Board (IRB) and all clinical investigations had been conducted according to the principles expressed in the Declaration of Helsinki. A written informed consent was obtained from all the participants and the content of the informed consent was approved by the Institutional Research Board. Prospective primary tongue cancer samples (n = 100), both punch

biopsy and tissue samples from surgical specimens with corresponding apparently uninvolved adjacent tissue were collected from the patients presenting with OTSCC between the years 2014 and 2016 to the Head and Neck Oncology Clinic, Cancer Institute (WIA), Chennai, India. Collected samples were immediately immersed in RNAlater and stored in -80°C till further use. The patients were treated as per the NCCN guidelines. Variables recorded and evaluated for the study included age, sex, site, size of the tumor, clinical stage, histological grade, tobacco habits, alcohol habits, comorbidities like Diabetes mellitus and Hypertension, node status, perineural invasion and occult node status.

Transcription sequencing. Transcriptome libraries for sequencing were constructed using $0.5\ \mu\text{g}$ of total RNA as per manufacturer's instructions (TruSeq RNA Sample Preparation Kit, Illumina) and was analysed as described previously⁸⁷. Briefly, the libraries were multiplexed and sequenced on HiSeq 2500 to obtain on average ~ 108 million single end 95 bp reads per sample. RNA-seq reads were mapped to the human genome version NCBI GRCh37 using GSNAP. Only uniquely mapped reads were considered for downstream analysis. Gene models were based on RefSeq transcripts, NCBI and ENSEMBL gene annotations. Differential gene expression was performed with the R/Bioconductor package DESeq2. For clustering, count data were transformed by variance stabilization and genes were centered to have mean zero. Clustering was performed using 1-Pearson correlation as distance metric and average linkage. Heat maps were generated using the R/Bioconductor package NMF. Fold change was calculated based on comparing the average of tumor samples with the average of normal samples. The relative abundance metric parameter FPKM (Fragments Per Kilobase of exon per Million reads sequenced) was used to represent the value of gene expression. To detect and filter DEGs, we applied the Student's *t* test (*P* values), and the FDR correction (*q*-values).

Functional analyses. Functional interpretation of the DEGs was performed with Database for Annotation, Visualization and Integrated Discovery (DAVID, Version 6.8)⁸⁸, a web-based tool for Gene Ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analyses⁸⁹. A Benjamini-corrected *P* value less than 0.05 was used to identify a statistically significant analysis. The hub proteins were identified using the Cytoscape v3.6.0. plugin called 'Network Analyzer'⁹⁰. The hub proteins were evaluated by analyzing the highest closeness centrality (CC), betweenness centrality (BC), and the node degree. Genes with degree ≥ 10 were defined as hub genes in the present study. In addition, the Molecular Complex Detection (MCODE) plugin of Cytoscape software was also employed to identify functionally related and highly interconnected clusters from the PPI network with a degree cutoff of 2, node score cutoff of 0.3, *k*-core of 4, and maximum depth of 100 (<http://baderlab.org/Software/MCODE>)⁹¹. Significant modules were identified with MCODE score ≥ 4 and nodes ≥ 6 . Subsequently, based on modules selected from the PPI network, functional enrichment analysis was performed using DAVID with the criterion of $P < 0.05$. The protein-protein interactions PPI was visualised using Search Tool for the Retrieval of Interacting Genes (STRING) online database (v10.5) (www.string-db.org) for network construction⁹².

qPCR. Total RNA was extracted from the tissue samples, homogenising them in liquid nitrogen, using RNeasy mini kit (Qiagen, Hilden, Germany) using standard manufacturer protocol. Total RNA ($2\ \mu\text{g}$) was converted to cDNA using High capacity cDNA reverse transcription kit (Applied Biosystems, California, U.S.A.) as per instructions of the manufacturer and stored at -40°C . Primers used for the study are listed in Supplementary Table S7. The quantitative real-time RT-PCR was performed using FastStart Universal SYBR Green Master (Rox) (Roche, Basel, Switzerland) in a $20\ \mu\text{l}$ reaction mix according to the manufacturer's instructions on a 7500 Real Time PCR System (Applied Biosystems, California, U.S.A.). The thermal cycling conditions used were as mentioned before²³. Triplicates were performed for each gene and average expression value was computed for subsequent analysis. The relative expression level of the genes was calculated using the $2^{-\Delta\Delta\text{Ct}}$ method. We defined the cut-off value for over-expression of genes based on the median log RQ values for each of the marker. Patients with pathological stages I and II disease were grouped as 'early stage' tumors and patients with pathological stages III and IV were grouped as 'advanced stage' tumors. We used *t* test to identify the markers whose expression varied most among the staging, node positive and outcome subgroups.

Immunohistochemistry. The IHC methodology is mentioned previously²³. The IHC detection of MMP9, LAMC2, ECAD, was performed on five-micron sections of FFPE tissues. The sections were deparaffinised in xylene, rehydrated in absolute ethanol and endogenous peroxidase activity was blocked by incubation in 0.3% hydrogen peroxide in Phosphate-buffered Saline (PBS) for 30 min. The sections were then subjected to heat-induced epitope retrieval according to the conditions listed in Supplementary Table S8. Briefly, the sections were blocked in 2% Bovine Serum Albumin (BSA) for 30 min and then incubated with primary monoclonal antibody overnight at 4°C . The marker expression was observed using the SuperSensitive Polymer-HRP IHC Detection System (BioGenex Laboratories, San Ramon, CA) as per the instructions of the manufacturer. Primary antibody was replaced with 2% BSA in negative control and suitable positive control specific to each of the markers was used. Immunostaining of the sections was reviewed along with corresponding Haematoxylin and Eosin stained sections.

Statistical analysis. Overall Survival (OS) was calculated as time in months from the date of primary treatment to the date of death due to the disease. Disease free survival (DFS) was calculated as time in months from the date of primary treatment to the date of objective tumor recurrence (local, nodal, locoregional, and distant metastases). The samples were blinded with respect to clinical data during the molecular testing. All statistical analysis was done using SPSS, version 16 (IBM).

References

1. Elango, J. K. *et al.* Trends of head and neck cancers in urban and rural India. *Asian Pac. J. Cancer Prev.* **7**, 108–112 (2006).
2. Joseph, L. J. *et al.* Racial disparities in squamous cell carcinoma of the oral tongue among women: a SEER data analysis. *Oral Oncol.* **51**, 586–592 (2015).
3. Annertz, K., Anderson, H., Palmér, K. & Wennerberg, J. The increase in incidence of cancer of the tongue in the Nordic countries continues into the twenty-first century. *Acta Otolaryngol.* **132**, 552–557 (2012).
4. Malik, A. *et al.* Trends of oral cancer with regard to age, gender, and subsite over 16 years at a tertiary cancer center in India. *Indian J. Med. Paediatr. Oncol.* **39**, 297 (2018).
5. Mukdad, L. *et al.* Oral tongue squamous cell carcinoma survival as stratified by age and sex: a surveillance, epidemiology, and end results analysis. *Laryngoscope* **129**, 2076–2081 (2018).
6. Patel, S. C. *et al.* Increasing incidence of oral tongue squamous cell carcinoma in young white women, age 18 to 44 years. *J. Clin. Oncol.* **29**, 1488–1494 (2011).
7. Iyer, N. G. *et al.* Changing epidemiology of oral squamous cell carcinoma of the tongue: a global study. *Head Neck* **39**, 297–304 (2017).
8. Bello, I. O. *et al.* Prognostic evaluation of oral tongue cancer: means, markers and perspectives (I). *Oral Oncol.* **46**, 630–635 (2010).
9. D'Cruz, A. K. *et al.* Elective versus therapeutic neck dissection in node-negative oral cancer. *N. Engl. J. Med.* **373**, 521–529 (2015).
10. Faisal, M. *et al.* Depth of invasion (DOI) as a predictor of cervical nodal metastasis and local recurrence in early stage squamous cell carcinoma of oral tongue (ESSCOT). *PLoS ONE* **13**, e0202632 (2018).
11. Tam, S. *et al.* Depth of invasion as a predictor of nodal disease and survival in patients with oral tongue squamous cell carcinoma. *Head Neck* **41**, 177–184 (2018).
12. Sharma, K. *et al.* Prognostic factors, failure patterns and survival analysis in patients with resectable oral squamous cell carcinoma of the tongue. *Radiat. Oncol. J.* **37**, 73–81 (2019).
13. Okuyemi, O. T. *et al.* TNM staging compared with a new clinicopathological model in predicting oral tongue squamous cell carcinoma survival. *Head Neck* **36**, 1481–1489 (2014).
14. Sharma, P. *et al.* A prospective study of prognostic factors for recurrence in early oral tongue cancer. *J. Clin. Diagn. Res* **7**, 2559–2562 (2013).
15. Almangush, A. *et al.* Prognostic biomarkers for oral tongue squamous cell carcinoma: a systematic review and meta-analysis. *Br. J. Cancer* **117**, 856–866 (2017).
16. Khan, H. *et al.* Correlation between expressions of Cyclin-D1, EGFR and p53 with chemoradiation response in patients of locally advanced oral squamous cell carcinoma. *BBA Clin.* **3**, 11–17 (2015).
17. Wangsa, D. *et al.* Ki-67 expression predicts locoregional recurrence in stage I oral tongue carcinoma. *Br. J. Cancer* **99**, 1121 (2008).
18. Kauppila, J. *et al.* Toll-like receptor 5 (TLR5) expression is a novel predictive marker for recurrence and survival in squamous cell carcinoma of the tongue. *Br. J. Cancer* **108**, 638–643 (2013).
19. Wen, H. *et al.* Decreased expression of BATF2 is significantly associated with poor prognosis in oral tongue squamous cell carcinoma. *Oncol.* **31**, 169–174 (2014).
20. Braakhuis, B. J. M. *et al.* Treatment choice for locally advanced head and neck cancers on the basis of risk factors: biological risk factors. *Ann. Oncol.* **23**, x173–x177 (2012).
21. Da Fonseca, A. J. *et al.* Comparison of three human papillomavirus DNA detection methods: next generation sequencing, multiplex-PCR and nested-PCR followed by Sanger based sequencing. *J. Med. Virol.* **88**, 888–894 (2015).
22. Thangaraj, S. V. *et al.* Molecular portrait of oral tongue squamous cell carcinoma shown by integrative meta-analysis of expression profiles with validations. *PLoS ONE* **11**, e0156582 (2016).
23. Marsh, D. *et al.* Stromal features are predictive of disease mortality in oral cancer patients. *J Pathol* **223**, 470–481 (2011).
24. Kelner, *et al.* Activin A immunorexpression as predictor of occult lymph node metastasis and overall survival in oral tongue squamous cell carcinoma. *Head Neck* **37**, 479–486 (2015).
25. Zhao, J. *et al.* Signal transduction by focal adhesion kinase in cancer. *Cancer Metastasis Rev.* **28**, 35–49 (2009).
26. Zhang, H. *et al.* Identification of key genes and pathways in tongue squamous cell carcinoma using bioinformatics analysis. *Med. Sci. Monit.* **23**, 5924–5932 (2017).
27. Huan, J. *et al.* Insights into significant pathways and gene interaction networks underlying breast cancer cell line MCF-7 treated with 17 β -Estradiol (E2). *Gene* **533**, 346–355 (2014).
28. Fang, Z.-Q. *et al.* Gene expression profile and enrichment pathways in different stages of bladder cancer. *Genet. Mol. Res.* **12**, 1479–1489 (2013).
29. Horvath, S. *et al.* Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc. Natl. Acad. Sci. USA* **103**, 17402–17407 (2006).
30. Ko, S.-Y. *et al.* Increased expression of amyloid precursor protein in oral squamous cell carcinoma. *Int. J. Cancer* **111**, 727–732 (2004).
31. Hu, S. *et al.* Salivary proteomics for oral cancer biomarker discovery. *Clin. Cancer Res.* **14**, 6246–6252 (2008).
32. Chaudhary, N. *et al.* Incidence of occult metastasis in clinically N0 oral tongue squamous cell carcinoma and its association with tumor staging, thickness, and differentiation. *J. Head Neck Phys. Surg.* **5**, 75 (2017).
33. Ho, C. M. *et al.* Occult lymph node metastasis in small oral tongue cancers. *Head Neck* **14**, 359–363 (1992).
34. Weiss, M. H. *et al.* Use of decision analysis in planning a management strategy for the stage NO neck. *Arch. Otolaryngol. Head Neck Surg.* **120**, 699–702 (1994).
35. Fakhri, A. R. *et al.* Prophylactic neck dissection in squamous cell carcinoma of oral tongue: a prospective randomized study. *Semin. Surg. Oncol.* **5**, 327–330 (1989).
36. Greenberg, J. S. *et al.* Disparity in pathologic and clinical lymph node staging in oral tongue carcinoma implications for therapeutic decision making. *Cancer* **98**, 508–515 (2003).
37. Moilanen, M. *et al.* Podoplanin, a potential therapeutic target for nasopharyngeal carcinoma. *Biomed Res. Int.* <https://doi.org/10.1155/2019/7457013> (2019).
38. Rosenthal, E. L. & Matrisian, L. M. Matrix metalloproteinases in head and neck cancer. *Head Neck* **28**, 639–648 (2006).
39. Fingleton, B. Matrix metalloproteinases: roles in cancer and metastasis. *Front. Biosci.* **11**, 479–491 (2006).
40. Zhou, X. *et al.* Global expression-based classification of lymph node metastasis and extracapsular spread of oral tongue squamous cell carcinoma. *Neoplasia* **8**, 925–932 (2006).
41. Kuratomi, Y. *et al.* Laminin gamma2 chain expression in squamous cell carcinomas of the tongue and its clinical relevance. *Nihon Jibiinkoka Gakkai Kaiho* **107**, 653–657 (2004).
42. Katoh, K. *et al.* Correlation between laminin-5 γ 2 chain expression and epidermal growth factor receptor expression and its clinicopathological significance in squamous cell carcinoma of the tongue. *Oncology* **62**, 318–326 (2002).

43. Stoltzfus, P. *et al.* The gamma2 chain of laminin-5 as an indicator of increased risk for recurrence in T1 stage tongue cancer. *Anti-cancer Res.* **24**, 3109–3114 (2004).
44. Nguyen, C. T. K. *et al.* LAMC2 is a predictive marker for the malignant progression of leukoplakia. *J. Oral Pathol. Med.* **46**, 223–231 (2017).
45. Han, C.-P. *et al.* Desmoglein-2 overexpression predicts poor prognosis in hepatocellular carcinoma patients. *Eur. Rev. Med. Pharmacol. Sci.* **22**, 5481–5489 (2018).
46. Yashiro, M. *et al.* Decreased expression of the adhesion molecule desmoglein-2 is associated with diffuse-type gastric carcinoma. *Eur. J. Cancer* **42**, 2397–2403 (2006).
47. Hundsdorfer, B. *et al.* Tumour-associated urokinase-type plasminogen activator (uPA) and its inhibitor PAI-1 in normal and neoplastic tissues of patients with squamous cell cancer of the oral cavity—clinical relevance and prognostic value. *J. Cranio-Maxillofacial Surg.* **33**, 191–196 (2005).
48. Baker, E. A. *et al.* Plasminogen activator system in oral squamous cell carcinoma. *Br. J. Oral Maxillofac. Surg.* **45**, 623–627 (2007).
49. Okada, K. *et al.* Overexpression of forkhead box M1 transcription factor (FOXM1) is a potential prognostic marker and enhances chemoresistance for docetaxel in gastric cancer. *Ann. Surg. Oncol.* **20**, 1035–1043 (2013).
50. Li, X. *et al.* Forkhead box transcription factor 1 expression in gastric cancer: FOXM1 is a poor prognostic factor and mediates resistance to docetaxel. *J. Transl. Med.* **11**, 204 (2013).
51. Rickman, D. S. *et al.* Distinctive molecular profiles of high-grade and low-grade gliomas based on oligonucleotide microarray analysis. *Cancer Res.* **61**, 6885–6891 (2001).
52. Xu, N. *et al.* FoxM1 mediated resistance to gefitinib in non-smallcell lung cancer cells. *Acta Pharmacol. Sin.* **33**, 675–681 (2012).
53. Harada, K., Ferdous, T., Minami, H. & Mishima, K. Prognostic significance of FOXM1 in oral squamous cell carcinoma patients treated by docetaxel-containing regimens. *Mol. Clin. Oncol.* **10**, 29–36 (2019).
54. Wessels, D., Murray, J., Jung, G., Hammer, J. A. & Soll, D. R. Myosin IB null mutants of Dictyostelium exhibit abnormalities in motility. *Cell Motil. Cytoskeleton* **20**, 301–315 (1991).
55. Yamada, Y. *et al.* Passenger strand of miR-145-3p acts as a tumor-suppressor by targeting MYO1B in head and neck squamous cell carcinoma. *Int. J. Oncol.* **52**, 166–178 (2018).
56. Ohmura, G. *et al.* Aberrant myosin 1b expression promotes cell migration and lymph node metastasis of HNSCC. *Mol. Cancer Res.* **13**, 721–731 (2015).
57. Liu, P. *et al.* SOX2 Promotes cell proliferation and metastasis in triple negative breast cancer. *Front. Pharmacol.* **9**, 942 (2018).
58. Piva, M. *et al.* Sox2 promotes tamoxifen resistance in breast cancer cells. *EMBO Mol. Med.* **6**, 66–79 (2014).
59. Han, S. *et al.* Prognostic value of CD133 and SOX2 in advanced cancer. *J. Oncol.* **2019**, 3905817 (2019).
60. Leis, O. *et al.* Sox2 expression in breast tumours and activation in breast cancer stem cells. *Oncogene* **31**, 1354–1365 (2012).
61. Liu, X. *et al.* Sox2 promotes tumor aggressiveness and epithelial–mesenchymal transition in tongue squamous cell carcinoma. *Int. J. Mol. Med.* **42**, 1418–1426 (2018).
62. Behrem, S. *et al.* Distribution pattern of tenascin-C in glioblastoma: correlation with angiogenesis and tumor cell proliferation. *Pathol. Oncol. Res.* **11**, 229–235 (2005).
63. Yang, Z. *et al.* Tenascin C is a prognostic determinant and potential cancer-associated fibroblasts marker for breast ductal carcinoma. *Exp. Mol. Pathol.* **102**, 262–267 (2017).
64. Yang, Z.-T. *et al.* Tenascin-C is a diagnostic marker for cervical cancer and a potential marker for cancer-associated fibroblasts in cervical carcinoma. *Int. J. Clin. Exp. Pathol.* **10**, 139–147 (2017).
65. Sundquist, E. The role of tumor microenvironment on oral tongue cancer invasion and prognosis (2018).
66. Nie, S. *et al.* Tenascin-C: a novel candidate marker for cancer stem cells in glioblastoma identified by tissue microarrays. *J. Proteome Res.* **14**, 814–822 (2015).
67. Astarita, J. L. *et al.* Podoplanin: emerging functions in development, the immune system, and cancer. *Front. Immunol.* **3**, 283 (2012).
68. Retzbach, E. P. *et al.* Podoplanin emerges as a functionally relevant oral cancer biomarker and therapeutic target. *Oral Oncol.* **78**, 126–136 (2018).
69. Kaneko, M. K. *et al.* ChLpMab-23: cancer-specific human-mouse chimeric anti-podoplanin antibody exhibits antitumor activity via antibody-dependent cellular cytotoxicity. *Monoclon. Antib. Immunodiagn. Immunother.* **36**, 104–112 (2017).
70. Ochoa-Alvarez, J. A. *et al.* Antibody and lectin target podoplanin to inhibit oral squamous carcinoma cell migration and viability by distinct mechanisms. *Oncotarget* **6**, 9045–9060 (2015).
71. Tsuneki, M. *et al.* Podoplanin-mediated cell adhesion through extracellular matrix in oral squamous cell carcinoma. *Lab. Investig.* **93**, 921–932 (2013).
72. Hsu, Y.-B. *et al.* Podoplanin, a potential therapeutic target for nasopharyngeal carcinoma. *Biomed Res. Int.* **2019**, 1–9 (2019).
73. Krishnamurthy, A. Current status and future perspectives of sentinel lymph node biopsy in oral cancers. *Indian J. Dent. Res.* **28**, 239–240 (2017).
74. Hyslop, T. *et al.* Occult tumor burden predicts disease recurrence in lymph node-negative colorectal cancer. *Clin. Cancer Res.* **17**, 3293–3303 (2011).
75. Pagliarulo, V. *et al.* Detection of occult lymph node metastases in locally advanced node-negative prostate cancer. *J. Clin. Oncol.* **24**, 2735–2742 (2006).
76. Zhi, M. & Iganej, S. Outcomes of stage I–II squamous cell carcinoma of the oral tongue (OTSCC) managed with surgery including an elective nodal dissection: potential indications for postoperative radiation therapy (PORT). *Int. J. Radiat. Oncol. Biol. Phys.* **99**, E386–E387 (2017).
77. Marra, A. *et al.* Long-term disease-free survival in surgically-resected oral tongue cancer: a 10-year retrospective study. *Acta Otorhinolaryngol. Ital.* **39**, 84–91 (2019).
78. Murthy, V. *et al.* Analysis of prognostic factors in 1180 patients with oral cavity primary cancer treated with definitive or adjuvant radiotherapy. *J. Cancer Res. Ther.* **6**, 282 (2010).
79. Oh, H. J. *et al.* Overexpression of POSTN in tumor stroma is a poor prognostic indicator of colorectal cancer. *J. Pathol. Transl. Med.* **51**, 306–313 (2017).
80. González-González, L. & Alonso, J. Periostin: a matricellular protein with multiple functions in cancer development and progression. *Front. Oncol.* **8**, 225 (2018).
81. Park, S. Y., Piao, Y., Jeong, K. J., Dong, J. & de Groot, J. F. Periostin (POSTN) regulates tumor resistance to antiangiogenic therapy in glioma models. *Mol. Cancer Ther.* **15**, 2187–2197 (2016).
82. Choi, P. *et al.* Examination of oral cancer biomarkers by tissue microarray analysis. *Arch. Otolaryngol. Head Neck Surg.* **134**, 539–546 (2008).
83. Al-Omer, H., Gawish, G.E.-H. & Sheler, C. OI0249 role of POSTN gene in regulating tumorigenesis of oral squamous cell carcinomas. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **117** e345 (2014).
84. Yu, B. *et al.* Periostin secreted by cancer-associated fibroblasts promotes cancer stemness in head and neck cancer by activating protein tyrosine kinase 7. *Cell Death Dis.* **9**, 1082 (2018).
85. Yong-Deok, K. *et al.* Molecular genetic study of novel biomarkers for early diagnosis of oral squamous cell carcinoma. *Med. Oral Patol. Oral Cir. Bucal.* **20**, e167–179 (2015).
86. Almangush, *et al.* Prognostic biomarkers for oral tongue squamous cell carcinoma: systematic review and meta-analysis. *Br. J. Cancer* **117**, 856–866 (2017).

87. Mohan, V. *et al.* Comprehensive genomic analysis identifies pathogenic variants in maturity-onset diabetes of the young (MODY) patients in South India. *BMC Med. Genet.* **19**, 22 (2018).
88. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44 (2008).
89. Ogata, H. *et al.* KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
90. Cline, M. S. *et al.* Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* **2**, 2366–2382 (2007).
91. Bader, G. D. & Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinf.* **4**, 2 (2003).
92. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).

Author contributions

V.R. conceptualized the study design, analyzed the results and wrote the manuscript text, S.V.T. analyzed the qPCR results, summarised the results, prepared the figures, tables, and supplementary tables, V.R.S. analyzed the immunohistochemistry results, A.K. managed the clinical data of patients. All the authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-78624-4>.

Correspondence and requests for materials should be addressed to V.R.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021