Bhosale, P. G., Cristea, S., Ambatipudi, S., Desai, R. S., Kumar, R., Patil, A., ... Mahimkar, M. B. (2017). Chromosomal Alterations and Gene Expression Changes Associated with the Progression of Leukoplakia to Advanced Gingivobuccal Cancer. *Translational Oncology, 10*(3), 396-409. https://doi.org/10.1016/j.tranon.2017.03.008

Publisher's PDF, also known as Version of record
License (if available):
CC BY-NC-ND
Link to published version (if available):
10.1016/j.tranon.2017.03.008

Link to publication record in Explore Bristol Research
PDF-document

This is the final published version of the article (version of record). It first appeared online via Elsevier at https://doi.org/10.1016/j.tranon.2017.03.008. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights
This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/about/ebr-terms
Chromosomal Alterations and Gene Expression Changes Associated with the Progression of Leukoplakia to Advanced Gingivobuccal Cancer

Priyanka G. Bhosale*,†, Simona Cristea‡,§, Srikant Ambatipudi*,¶, Rajiv S. Desai#, Rajiv Kumar**, Asawari Patil**, Shubhada Kane**, Anita M. Borges††, Alejandro A. Schäffer‡‡, Niko Beerenwinkel‡,§ and Manoj B. Mahimkar*†

*Cancer Research Institute (CRI), Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre (TMC), Navi Mumbai, 410210, India; †Homi Bhabha National Institute, Training School Complex, Anushakti Nagar, Mumbai, 400085, India; ‡Department of Biosystems Science and Engineering, ETH Zurich, 4058, Basel, Switzerland; §SIB Swiss Institute of Bioinformatics, 4058, Basel, Switzerland; ¶MRC Integrative Epidemiology Unit, University of Bristol, BS8 1TH, UK; #Department of Oral Pathology & Microbiology, Nair Hospital Dental College, Mumbai, 400 008, India; ‡‡Department of Pathology, Tata Memorial Hospital, Tata Memorial Centre (TMC), Parel, Mumbai, 400012, India; **Department of Pathology & Laboratory Medicine, S. L. Raheja Hospital, Mumbai, 400016, India; ††Computational Biology Branch, National Center for Biotechnology Information, National Institute of Health (NIH), Department of Health and Human Services (DHHS), Bethesda, MD, 20894, USA

Abstract
We present an integrative genome-wide analysis that can be used to predict the risk of progression from leukoplakia to oral squamous cell carcinoma (OSCC) arising in the gingivobuccal complex (GBC). We find that the genomic and transcriptomic profiles of leukoplakia resemble those observed in later stages of OSCC and that several changes are associated with this progression, including amplification of 8q24.3, deletion of 8p23.2, and dysregulation of DERL3, EIF5A2, ECT2, HOXC9, HOXC13, MAL, MFAP5 and NELL2. Comparing copy number profiles of primary tumors with and without lymph-node metastasis, we identify alterations associated with metastasis, including amplifications of 3p26.3, 8q24.21, 11q22.1, 11q22.3 and deletion of 8p23.2. Integrative analysis reveals several biomarkers that have never or rarely been reported in previous OSCC studies, including amplifications of 1p36.33 (attributable to MXRA8), 3q26.31 (EIF5A2), 9p24.1 (CD274), and 12q13.2 (HOXC9 and HOXC13). Additionally, we find that amplifications of 1p36.33 and 11q22.1 are strongly correlated with poor clinical outcome. Overall, our findings delineate genomic changes that can be used in treatment management for patients with potentially malignant leukoplakia and OSCC patients with higher risk of lymph-node metastasis.

Introduction
Oral cancer starts with an oral pre-invasive lesion (OPL) that progresses from hyperplasia through dysplasia, and finally develops into invasive oral squamous cell carcinoma (OSCC). Leukoplakia is the most predominant pre-invasive lesion [1–4], however the ability to predict the malignant potential from histopathological data is limited. Moreover, the 5-year overall survival in OSCC has not
substantially improved in recent decades [5], and early diagnosis and primary prevention remain the best approaches for OSCC management. To this end, the major challenge in early diagnosis is identifying pre-invasive lesions that are at high risk of malignant transformation [6,7]. However, OSCC is frequently diagnosed in advanced stages, which negatively influences prognosis. The most important prognostic factors that determine mortality and morbidity in OSCC patients are lymph node involvement and locoregional recurrence.

The histopathological evaluation of oral cancers is often not sufficient to predict disease aggressiveness and clinical outcome [8]. Multiple genetic and epigenetic events occur before tissue changes are microscopically detectable. The number of acquired genetic alterations increases with disease advancement from squamous hyperplasia through dysplasia to invasive carcinoma [5]. It is known that copy number alterations (CNAs) ranging from a small number of specific genes to entire chromosomes are significantly associated with OSCC development and progression [9,10]. These alterations are presumed to alter the expression level of single genes or gene clusters mapping within CNA regions [11]. Therefore, analyses that integrate CNA data with gene expression (GE) data may identify predictive DNA-based markers applicable in clinical prognosis [12].

Molecular profiles of oral cancers are largely influenced by the site of tumor development and associated etiological agents, implying divergent pathways for oral cancer development [13–17]. India is an interesting location to study the genomics of tobacco-associated OSCC, due to the fact that in India there is a high incidence of oral cancers associated with the abuse of smokeless tobacco, most of which are negative for human papilloma virus (HPV) [18].

This is the first comprehensive study combining genomic profiling and integrative analyses of HPV-negative gingivobuccal complex (GBC) leukoplakia and OSCC of different stages from a large set of Indian patients. We identified signatures associated with the progression of pre-invasive lesions to invasive OSCC and found candidate driver alterations unique to primary tumors with lymph node metastasis and related to patient survival.

**Materials and Methods**

**Tissue Specimen Collection**

The study was approved by the Institutional Local Ethics Committee of Tata Memorial Hospital (TMH) and Nair Hospital Dental College, Mumbai, India. Written informed consent was obtained from all the study participants. Paraffin blocks and frozen tissue samples of leukoplakia, neo-primary oral tumor tissues, and non-inflamed gingivobuccal mucosa tissues from clinically healthy individuals with no previous personal history of cancer were recruited from Nair Hospital and TMH, respectively. Patients received neither radiation nor chemotherapy before surgery. Histopathologically confirmed leukoplakia and tumor tissues were subjected to DNA–RNA extraction as detailed in Supplementary Information. Screening for the presence of HPV was done as described in [18]. Details regarding the numbers of samples used in the test and validation sets, as well as the

---

| Patient Characteristics | Total Study Samples | aCGH & GE Study (n = 121)# | qRT-PCR (n = 207)# | IHC & FISH (n = 370)# |
|------------------------|---------------------|-----------------------------|-------------------|-----------------------|
| Age at diagnosis        |                      |                             |                   |                       |
| Median age              | 49                  | 42                          | 50                | 41                    |
| Range (IQR)*            | 40–59               | 38–50                       | 43–61             | 33–53                 |
| Gender                  |                      |                             |                   |                       |
| Male                    | 299 (76.3%)         | 21 (87.5%)                  | 79 (76.9%)        | 33 (89.2%)            |
| Female                  | 93 (23.7%)          | 3 (12.5%)                   | 21 (23.1%)        | 4 (10.8%)             |
| Pathological stage      |                      |                             |                   |                       |
| Stage 1 and 2 (Early stage OSCC) | 82 (35.5%) | 32 (35.2%)                   | NA                | 56 (41.5%)            |
| Stage 3 and 4 (Advanced stage OSCC) | 149 (64.5%) | NA                          | 59 (64.8%)        | 79 (58.5%)            |
| Pathological T classification |              |                             |                   |                       |
| T1                      | 31 (13.4%)          | NA                          | 7 (7.7%)          | 25 (18.5%)            |
| T2                      | 100 (43.3%)         | 40 (44%)                    | 40 (44%)          | 64 (47.4%)            |
| T3                      | 10 (4.3%)           | 4 (4.4%)                    | 4 (4.4%)          | 4 (4.3%)              |
| T4                      | 90 (39%)            | NA                          | 40 (44%)          | 42 (31.1%)            |
| Pathological cervical lymph node involvement | |                             |                   |                       |
| Node negative (N0)      | 133 (57.6%)         | NA                          | 55 (60.4%)        | 79 (58.5%)            |
| Node positive (N+)      | 98 (42.4%)          | NA                          | 36 (39.6%)        | 56 (41.5%)            |
| Pathological grade      |                      |                             |                   |                       |
| Well                    | 27 (7.9%)           | NA                          | 8 (8.8%)          | 12 (8.9%)             |
| Moderate                | 139 (40.9%)         | NA                          | 55 (60.4%)        | 87 (64.4%)            |
| Poor                    | 64 (18.8%)          | NA                          | 28 (30.8%)        | 36 (26.7%)            |
| Hyperplasia             | 89 (26.2%)          | NA                          | NA                | 31 (86.1%)            |
| Mild dysplasia          | 11 (3.2%)           | 3 (12.5%)                   | NA                | 9 (9.1%)              |
| Moderate dysplasia      | 8 (2.4%)            | NA                          | 2 (5.6%)          | 8 (8.1%)              |
| Severe Dysplasia        | 2 (0.6%)            | NA                          | NA                | 2 (2%)                |
| Habit profile           |                      |                             |                   |                       |
| No Habit                | 9 (3.1%)            | 9 (45%)                     | NA                | 3 (2.6%)              |
| Exclusive tobacco users | 157 (54.5%)         | 3 (15%)                     | 63 (77.8%)        | 13 (41.9%)            |
| Exclusive smoker        | 18 (6.3%)           | NA                          | 2 (2.5%)          | 5 (16.2%)             |
| Exclusive alcohol user  | 1 (0.3%)            | NA                          | NA                | 1 (0.8%)              |
| Mixed habit**           | 103 (35.8%)         | 8 (40%)                     | 16 (19.8%)        | 13 (41.9%)            |

# Represents total number of samples, including Buccal Mucosa (BM) Normals: n = 6 (GE), n = 32 (qRT-PCR) and n = 77 (IHC); all samples belonged to the gingivobuccal complex region of the oral cavity; T: Tumor classification based on site; N: Tumor classification based on lymph node metastasis; * IQR: Inter quartile range; **Mixed Habit: Tobacco chewing along with bidi/cigarette smoking and/or alcohol users.
clinicopathologic and demographic characteristics of patients, are provided in Table 1 and Figure S1.

**Array CGH and Gene Expression Profiling**

Whole-genome copy number and gene expression profiling was performed on 2x105K CGH oligonucleotide arrays and Whole Human Genome Microarray 4x44K (Agilent Technologies, USA) respectively. Hybridization and detailed analysis are described in Supplementary Information. The raw aCGH data have been submitted to the Gene Expression Omnibus (GEO) with accession numbers GSE85514 and GSE23831 and accession numbers for GE raw data are GSE85195 and GSE23558.

**Validation of Targets**

The copy number status of the targets was evaluated by fluorescence in situ hybridization (FISH/nuc ish). Immunohistochemical analysis (IHC) and quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) were performed for selected candidate genes found significantly deregulated. Detailed protocol and analysis is provided in Supplementary Information. Details regarding FISH probes, fluorescent TaqMan probes and antibodies used are listed in Tables S1, S2, S3.

**Literature Mining**

We updated our existing literature-based list of genes related to oral cancer from 277 genes [10] to 562 genes (as of May 2015). The list (Table S4) includes genes that were previously found to be either differentially expressed or copy number altered in oral cancers. The purposes were 1) to place our new results in the context of previous knowledge and 2) to determine the novelty of any gene expression change or CNA that we would choose for targeted validation.

![Figure 1. All amplifications (a) and deletions (b) inferred by GISTIC 2.0 in leukoplakia (A), early-stage OSCC (B), and advanced-stage OSCC (C) samples. Each alteration is assigned a G-score (left axis) by GISTIC 2.0, which considers the amplitude of the alteration, as well as its occurrence across samples.](image-url)
Figure 2. Nuc ish (FISH) for validating the amplifications of the regions 8q24.3 and 1p36.33. A) nuc ish for 8q24.3 locus in leukoplakia (b) and OSCC (c). B) nuc ish for 1p36.33 locus with weak amplification (b) and strong amplification (c) in OSCC. The specificities of the 8q24.3 locus probe (red)/chr 8 centromere (CEP) (green) and 1p36.33 (red) locus probe/chr 1 centromere (CEP) (green) were confirmed on the metaphase spreads as represented in A(a) and B(a). In all cases, the magnification was 630X.

Figure 3. Correlation between 8q24.3 amplification and oral cancer progression. A) Percentage of cells with 8q24.3 weak amplification across different groups, correlated with disease progression (P-value calculated using Spearman correlation). B) and C) The increased percentage of tumor cells with 8q24.3 amplification in leukoplakia and OSCC.
**Results**

**Clinicopathological and Demographic Characteristics of the Study Cohort**

Clinicopathological and demographic characteristics of all 481 leukoplakia and OSCC patients, together with follow-up data, are summarized in Tables 1 and S5, while Figure S1 shows how many samples were used in each phase of the study. Most patients were smokeless tobacco users while many had mixed habits (~35%) (chewing, bidi/cigarette smoking or consuming alcohol) and were negative for high-risk HPV [18]. The histopathology of the leukoplakia samples was either hyperplastic (89 samples) or mild dysplastic (11 samples) and 50% of the lesions analyzed for aCGH and GE study either transformed to OSCC or recurred after primary treatment. One hundred forty-nine patients (~65%) had advanced-stage OSCC and 82 patients (~35%) had early-stage OSCC. Approximately 60% cases were negative for lymph node metastasis.

**Genome-Wide copy Number Alterations**

 Genome-wide analysis of CNA was carried out in 24 leukoplakia, 32 early-stage OSCC, and 59 advanced-stage OSCC cases and revealed recurrent focal regions of amplification and deletion (Figures 1 and S2). We identified 19 alterations in leukoplakia, 32 alterations in early-stage OSCCs and 69 alterations in advanced-stage OSCGs (Table S6). The 10 most frequently amplified regions were 11q13.1 (70% of all patients), 8q24.3 (69%), 11p15.5 (60%), 1p36.33 (59%), 9p24.3 (59%), 8q24.21 (55%), 7q22.1 (54%), 7q11.23 (53%), 16p13.3 (53%) and 3q27.2 (52%). The 10 most frequently deleted regions were 8p23.2 (66% of all patients), 8p11.22 (65%), 3p14.2 (56%), 3p21.1 (54%), 8p22 (54%), 3p11.1 (53%), 3p22.3 (53%), 3p26.3 (53%), 8p23.1 (51%) and 15q11.1 (40%). Previous reports have proposed amplifications 8p23.1, 8q24.21 (harboring the candidate gene MYC), 11q22.1 (MMPS, BIRC2, BIRC3), and deletions of 4q23.3, 3p26.3 (CHL1), 3p12.2, 4q21.3, 7q31.1, 8p23.2 (CSMD1), 9p12, 11q22.3 (ATM, H2AFX), as well as 18q12.1. We hypothesize that these alterations are potential predictive biomarkers of lymph node metastasis. According to previous studies, the amplification of 8q24.21 and the deletion of 3p26.3 are associated with metastasis, invasion, and therapy resistance [20,21].

We found 25 CNAs associated with recurrence-free survival and 26 CNAs associated with disease specific survival (p-value < 0.25, Table 2). For example, the amplifications of 1p36.33, 11q13.3, 11q22.1 and 16p11.2 were associated with poor clinical outcome, whereas the

**Table 2. Univariate Cox Proportional Hazards Regression Analysis of Single Predictors for Recurrence-Free and Disease Specific Survival.**

| Chromosomal Alteration | Disease Specific Survival | Recurrence-Free Survival |
|-------------------------|---------------------------|--------------------------|
| **Cytoband**            | **BH Corrected** | **CPH Coef.** | **BH Corrected** | **CPH Coef.** |
| 1p36.33* | Amplification | 0.0327 | 1.0292 | 0.013 | 0.9185 |
| 1q23.2 | Amplification | 0.0269 | 0.8763 | 0.0934 | 0.574 |
| 3q27.2 | Amplification | 0.2467 | 0.436 | 0.2171 | 0.3823 |
| 11q13.3 | Amplification | 0.1232 | 0.4237 | 0.2142 | 0.2804 |
| 16p13.3 | Amplification | 0.0653 | 0.6641 | 0.0658 | 0.5375 |
| 16p11.2 | Amplification | 0.0497 | 0.8665 | 0.0549 | 0.6726 |
| 16q12.2 | Amplification | 0.0607 | 0.6662 | 0.0308 | 0.6497 |
| 2p11.2 | Deletion | 0.0473 | 1.0051 | 0.0844 | 0.777 |
| 2q22.1 | Deletion | 0.0147 | 1.0214 | 0.0623 | 0.6329 |
| 2q34 | Deletion | 0.0644 | 0.7772 | 0.0281 | 0.745 |
| 3p14.2 | Deletion | 0.1976 | 0.5665 | 0.2202 | 0.4236 |
| 4q13.2 | Deletion | 0.0582 | 0.8036 | 0.1775 | 0.4836 |
| 4q22.1 | Deletion | 0.0998 | 0.7998 | 0.2113 | 0.4591 |
| 9p23 | Deletion | 0.0969 | 0.711 | 0.1339 | 0.5327 |
| 15q22.3 | Deletion | 0.0829 | 0.7902 | 0.099 | 0.6467 |
| 6p21.1 | Amplification | 0.2198 | 0.4606 | - | - |
| 11q22.1 | Amplification | 0.1543 | 0.5204 | - | - |
| 1p11.31 | Amplification | 0.2227 | 0.5128 | - | - |
| 22q11.1 | Amplification | 0.1832 | 0.6038 | - | - |
| 1q31.3 | Deletion | 0.2371 | 0.5055 | - | - |
| 3p11.1 | Deletion | 0.0809 | 0.7957 | - | - |
| 4q13.2 | Deletion | 0.09 | 0.7457 | - | - |
| 5p14.3 | Deletion | 0.112 | 0.7583 | - | - |
| 7q31.1 | Deletion | 0.1624 | 0.598 | - | - |
| 13q21.32 | Deletion | 0.2166 | 0.5221 | - | - |
| 21q21.3 | Deletion | 0.1385 | 0.6243 | - | - |
| 2q37.3 | Amplification | - | - | 0.2185 | 0.3978 |
| 5p15.33 | Amplification | - | - | 0.223 | 0.3284 |
| 11p15.5 | Amplification | - | - | 0.1793 | 0.4287 |
| 13q21.33 | Amplification | - | - | 0.2436 | 0.7795 |
| 16q21 | Amplification | - | - | 0.1975 | 0.4048 |
| 19p13.3 | Amplification | - | - | 0.1465 | 0.3602 |
| 22q11.23 | Amplification | - | - | 0.1386 | 0.4272 |
| 9p23.1 | Deletion | - | - | 0.2007 | 0.6494 |
| 14q11.2 | Deletion | - | - | 0.1116 | 0.5538 |
| 17p13.1 | Deletion | - | - | 0.0346 | 0.7927 |

BH: Benjamini-Hochberg multiple testing correction method; CPH coef.: Cox Proportional Hazard coefficient. A positive regression coefficient means that the hazard is higher, thus the prognosis is worse; * Targets selected for validation; — represents not applicable.
amplification of 22q11.21 was associated with better survival. Additionally, a poor clinical outcome was also associated with the deletions of 2p11.2, 3p14.2, 4q13.2, 9p23 and 11q22.3. Kaplan–Meier survival curves for 1p36.33 and 11q22.1 are shown in Figure 4A. We validated the amplification of 1p36.33 by nuc ish (Figure 2B), and we confirmed that the amplification of 1p36.33 is associated with poor survival in an independent OSCC cohort (Figure 4B). Moreover, we found a strong association between the amplification of 1p36.33 and lymph node metastasis ($P < .001$).

**Gene Expression and Integrative Analyses**

Transcriptome-wide analysis was performed on 6 buccal mucosa normal tissues, 15 leukoplakia, 27 early-stage OSCC, and 34 advanced-stage OSCC. Principal component analysis of 3805 differentially expressed genes (log fold change of 2 and q-value ≤ 0.01) revealed two separate clusters of normal and OSCC samples, while the leukoplakia samples displayed changes overlapping with both these groups (Figure S4). We identified 849 genes differentially expressed (395 up-regulated and 454 down-regulated) in leukoplakia, 1813 (805

**Figure 4.** Kaplan–Meier plots of disease specific patient survival for selected chromosomal alterations based on aCGH analysis (A, 1p36.33 and 11q22.1) and nuc ish validation experiments (B, 1p36.33). Survival in months (X axis) is plotted against the fraction of patients alive (Y axis).
Figure 5. Heatmap of the top 20 differentially expressed genes in leukoplakia and OSCC. A) Differentially expressed genes in early-stage OSCC (pink) and advanced-stage OSCC (magenta) when compared to normal (green). B) Differentially expressed genes in leukoplakia (chartreuse) when compared to normal (green). Blue genes are up-regulated, while down-regulation is represented in red. All expression values were scaled across samples.

Table 3. Associations Between DNA Copy Number Alterations and Differentially Expressed Transcripts in Leukoplakia and OSCC.

| Cytoband | Alteration | q–Value | DE Genes in Leukoplakia vs. Normal | DE Genes in OSCC vs. Normal |
|----------|------------|---------|----------------------------------|----------------------------|
| 1p36.33* | Amplification | 0.00037596 | GLTPD1, LOC148413 | MXRA8 |
| 3q26.31 | Amplification | 0.01951 | CAMK2N2, GOI1M4, KLHL6, CLDN11, EIF4G1, ECT2*, GPR160, EIF5A2*, AP2M1 |
| 4q13.2 | Amplification | 0.0011293 | YIPF3, NFKBIE |
| 7p11.2 | Amplification | 3.31E–08 | |
| 7q11.23 | Amplification | 3.66E–07 | |
| 7q22.1 | Amplification | 0.00012685 | PCOLCE, GNB2 |
| 8q24.21 | Amplification | 0.00019006 | |
| 8q24.3* | Amplification | 6.67E–07 | |
| 9p24.1 | Amplification | 5.90E–05 | |
| 9q43.4 | Amplification | 0.01951 | |
| 11p15.5 | Amplification | 0.0002778 | CHD1 |
| 11p11.2 | Amplification | 0.043015 | PACSIN3, ARFGAP2 |
| 11q13.1 | Amplification | 9.71E–10 | RNASEH2C |
| 11q13.3 | Amplification | 1.09E–27 | |
| 12p13.2 | Amplification | 0.17411 | MUC11, HOXC9*, HOXC13*, ERBB3, ZNF385A, HOXC10, IKZF4 |
| 14q11.2 | Amplification | 0.0003729 | |
| 14q32.33 | Amplification | 2.63E–05 | INF2 |
| 16p11.2 | Amplification | 0.14231 | CD19, GBP3, P53L2, IL17, SPN, SEZ6L2, FUS, CORO1A |
| 17p13.1 | Amplification | 0.17411 | CLDN7, CD68, TME88, PHF23 |
| 17p13.3 | Amplification | 4.12E–07 | |
| 20q11.22 | Amplification | 0.19131 | SPAG4, ERGIC3 |
| 22q11.2 | Amplification | 0.0001267 | |
| 22q11.23 | Amplification | 0.0049669 | DERL3*, CC2orf43 |
| Xq28 | Amplification | 0.0063493 | SS8, AP2A1P1, IRAK1 |
| 1q44 | Deletion | 2.09E–05 | |
| 2q44 | Deletion | 0.00096669 | SPAG16 |
| 3p26.3 | Deletion | 0.0013306 | BHLHE40, C30orf2 |
| 3p21.1 | Deletion | 1.13E–05 | |
| 3p14.2 | Deletion | 4.30E–20 | |
| 4q13.2 | Deletion | 3.84E–07 | |
| 9p21.3 | Deletion | 5.36E–11 | |
| 10p11.21 | Deletion | 1.57E–05 | |
| 11q22.3 | Deletion | 0.042974 | CARD18, POU2AF1, CTSC, ST5GAL4, THY1, TRIM29, ED24, AMOTL1, PVRL1, ZC3H12C, CHEK1, DLAT |
| 17p13.1 | Deletion | 0.030986 | ALOX12, CLDN7, VM01, HES3ST3A1, CD68, GPR172B, SLC25A11, C17orf59, TME88, ATP2A3, UBE2G1, PHF23 |
| 17q21.3 | Deletion | 2.50E–05 | |
| 21q21.3 | Deletion | 0.017484 | ADAMT5 |

DE: Differential expression; Red font indicates amplified loci or up-regulated genes, and Blue font indicates deleted loci or down-regulated genes. The majority of genes showed consistent changes on both the DNA and the RNA levels, and are depicted here in black font (up-regulated genes in amplified regions and down-regulated genes in deleted regions). Few genes showed opposite direction expression changes and are colored respectively (genes in blue font were down-regulated, but located in amplified regions, while genes in red font were up-regulated, but located in deleted regions).

* Targets selected for validation.
Figure 6. qRT-PCR expression changes of the 10 target genes (2 down-regulated, 7 up-regulated and 1 unchanged) in normal, leukoplakia and early and advanced-stage OSCC samples. The $P$-value ($P$) was calculated using the Spearman correlation test.
Table 4. Correlation of qRT-PCR Validation Targets with clinicopathological parameters.

| Clinicopathological Parameters | EIF5A2 | HOXC9 | INHBA | MFAP5 | DVL1 | SLC4A1AP |
|-------------------------------|--------|-------|-------|-------|------|----------|
| Oral or head and neck cancers |        |       |       |       |      |          |
| Stage I/II                    | 0.015 (0.31) | 0.238 (0.08) | 0.414 (0.15) | 0.591 (0.54) | 0.007 (0.23) | 0.069 (0.06) |
| Stage III/IV                  | 0.008 (0.19) | 0.141 (0.10) | 0.877 (0.09) | 0.530 (0.21) | 0.072 (0.21) | 0.069 (0.06) |
| Node metastasis               | -0.001 (0.39) | -0.001 (0.36) | -0.001 (0.36) | -0.001 (0.36) | -0.001 (0.36) | -0.001 (0.36) |
| Node negative                 | -0.001 (0.39) | -0.001 (0.36) | -0.001 (0.36) | -0.001 (0.36) | -0.001 (0.36) | -0.001 (0.36) |

Specifically, EIF5A2, HOXC9, and INHBA were found in all three groups, revealing their role in disease progression from pre-invasive to cancerous lesions. A higher number of genes were dysregulated in OSCC compared to leukoplakia, including CXCL10, MMP10, INHBA, GBP5, CXCL11, MMP3, FST, BATF2, SPP1, SH2D5, CXCL9, IFT73, SERPINE1, GALNT6, FOXL2, PDPN, ITGA3, VEGFC, STAT1, LY6K, KL7, SOX9 and CD274. Among all the differentially expressed genes identified in this study, 61 have been previously reported to be involved in leukoplakia and 188 in oral or head and neck cancers, including ECT2, INHBA, SERPINE1, GBP5, MMP10, MMP3, LY6K, SPP1, PDL1, PTTH1, KRT4, KRT76 and MAL[10,22–26]. The novel oral cancer driving genes identified here include DERL3, EIF5A2, HOXC9, HOXC13, MAFAP5, NELL2, CD274, DHR52, FST and GIPX3. The top dysregulated genes in leukoplakia and OSCC are represented in Figure 5 and listed in Table S7.

Integrative analysis of gene expression and CNAs. We integrated the GE and CNA datasets to identify genes whose expression and copy number status were correlated. We found 3q26.31, 6p21.1, 7p11.2, 8q24.21, 8q24.3, 9p24.1, 11q13.3, 12q13.2, 16q24.2 and 17p13.1 as chromosomal hotspots for copy number-dependent gene overexpression, while 1q44, 2q34, 3p26.3, 3p21.1, 10p11.21, 11q22.3, 17p13.1 and 21q12.3 were identified as regions of copy number-dependent gene down-regulation (Figure 6). In early-stage OSCC, we observed a strong correlation between copy number amplifications and gene over-expression. The deletion of 11q22.3 was correlated with a few genes with significant copy number-dependent underexpression, including CRYAB, POU2AF1, EXPH5, MPZL2 and ARHGAP32. The opposite direction of expression change (amplification of down-regulated genes and deletion of up-regulated genes) was observed for a few genes, including MMP3, MMP10, FEZ1, CTSC, CHEK1, PANX1 and PAFAH1B2. Interestingly, 16p11.2, 17p13.1 and 22q11.23 were significantly amplified, however, the majority of the genes located in these three regions were down-regulated, e.g., CD19, GDPD3, NUPR1, SPN, CLDN7 and DELR3 (Table 3), potentially a consequence of epigenetic regulation.

Validation of dysregulated transcripts. To confirm the results of the GE analysis, real-time qRT-PCR (TaqMan assays) was performed in 32 normal, 37 leukoplakia, and 138 OSCC samples. We selected 10 genes for validation based on either novelty or on published studies implicating these genes in OSCC development (Table S4): seven up-regulated genes DVL1, EIF5A2, FUS, HOXC9, INHBA, LY6K, and MAFAP5, two down-regulated genes DERL3 and MAL, and the unchanged gene SLC4A1AP, along with RNA18S5 as endogenous control. All the validation targets that were found to be differentially expressed in the GE analysis were confirmed to display significant differences in expression between normal, leukoplakia, and tumors (Figure 6), and no significant difference was found in the expression of the unchanged gene SLC4A1AP. Specifically, HOXC9, MAFAP5, and INHBA showed very high expression changes in leukoplakia, early and advanced tumors versus normal (P<.01), while EIF5A2 and LY6K
were significantly overexpressed only in early and advanced-stage OSCCs ($P < 0.0001$ and $P = 0.03$, respectively). The log$_2$ fold change in expression of DVL1 and FUS was approximately 1 across the three groups, consistent with the microarray data.

We analyzed the associations between the validated target genes and clinicopathologic parameters (Table 4). Most targets (EIF5A2, HOXC9, MFAP5, LY6K, INHBA and DVL1) showed a positive correlation between their expression changes and OSCC progression from pre-invasive lesions to cancer. The expressions of DERL3 and MAL, which are down-regulated, were negatively correlated with disease progression. EIF5A2, HOXC9, INHBA, and MFAP5 were associated with disease advancement from early-stage OSCC to advanced-stage OSCC, and EIF5A2, HOXC9, INHBA, FUS and DVL1 were significantly associated with lymph node metastasis. IHC was performed to validate the protein overexpression of EIF5A2, ECT2, HOXC9, HOXC13, MFAP5, and NELL2. IHC analysis revealed strong protein expression of all the six targets in leukoplakia ($n = 108$) and OSCC ($n = 185$) versus normal ($n = 77$) (Figure 7). Further analyses reinforced the associations of these markers with disease progression, except NELL2 (Figure S6 and Table S8).

**Pathway Analyses**

To interpret the association of dysregulated genes with biological processes, we used the PANTHER (Protein ANalysis THrough Evolutionary Relationships) classification system [27,28] (Figure S7). Both leukoplakia and OSCC samples shared a large number of

![Figure 7](image-url)  
Figure 7. IHC analysis for target validation in normal, leukoplakia, and OSCC samples. NELL2 showed cytoplasmic staining, EIF5A2, ECT2, HOXC9, HOXC13 showed cytoplasmic and nuclear expression, while MFAP5 was localized predominantly into the matrix. Respective isotype controls for all the cases had no staining (images not shown). The rightmost panel shows differential localization of each marker. The magnification was 100X for the leftmost three panels, and 200X for the rightmost panel.
We have presented the first comprehensive analysis of genomic and transcriptomic profiles of a large set of tobacco-associated, HPV-negative gingivobuccal leukoplakia and OSCC patients from India. Our main goals were threefold: 1) to identify novel driver events associated with the transformation of pre-invasive lesions to high risk malignant OSCCs, as well as with patient survival; 2) to identify unique driver alterations found in primary tumors with lymph node metastasis; and 3) to identify driver genes with correlated CNA and gene expression profiles. Therefore, our study contributes to a genetic progression model of oral carcinogenesis (Figure 8).

The CNA landscape of gingivobuccal cancers is dominated by amplifications of the chromosomal regions 1p36.33, 3q26.31, 6p21.32, 7p11.2, 8q24.21, 8q24.3, 9q34.3, 11q13.1, 11q13.3, 11q22.1, 12q13.2, 16p11.2, and deletions of 3p21.1, 3p14.2, 4q13.1, 8p23.2, 8p11.22, 9p23, 9p21.3, 17p13.1. The amplifications of 3q, 7p, 8q, 9q, 11q, and 12q, as well as the deletions of 3p, 4q, 8p, and 9p were reported at least three times among 12 aCGH studies on primary OSCC tumors [19], with amplifications of 3q26, 11q13 and 11q22.2 being the most reported CNAs in advanced-stage OSCCs [9,29,30]. An extensive review by Gollin outlines established associations of most of these alterations in HNSCC [31]. In our data, at the whole-arm level, the amplification of 8q is the most common amplification associated with OSCC progression. At the sub-band level, the amplification of 8q24.3 was observed in 58% of the leukoplakia samples, as well as in 69% of the OSCC samples, while the region 8q24.21 was amplified in 55% of the OSCC samples.

We observed strong correlations between gene expression and amplifications at 3q26.31 (including the genes ECT2, EIF5A2, KHLH6, GPR160) and 12q13.2 (HOXC9, HOXC13, ERBB3, MUC11) in both OSCC and leukoplakia. In addition, we observed deletions of multiple regions on 3p (3p26.3, 3p22.3, 3p21.1, 3p14.2, 3p11.1) and 8p (8p23.2, 8p23.1, 8p22, 8p11.22) with high frequency (>52% in OSCCs), in line with literature reports in oral pre-invasive lesions [35,36]. These alterations can therefore be considered as important events associated with OSCC progression [36].

We observed strong correlations between gene expression and amplifications at 3q26.31 (including the genes ECT2, EIF5A2, KHLH6, GPR160) and 12q13.2 (HOXC9, HOXC13, ERBB3, MUC11) in both OSCC and leukoplakia. In addition, we observed deletions of multiple regions on 3p (3p26.3, 3p22.3, 3p21.1, 3p14.2, 3p11.1) and 8p (8p23.2, 8p23.1, 8p22, 8p11.22) with high frequency (>52% in OSCCs), in line with literature reports in oral pre-invasive lesions [35,36]. These alterations can therefore be considered as important events associated with OSCC progression [36].

We observed strong correlations between gene expression and amplifications at 3q26.31 (including the genes ECT2, EIF5A2, KHLH6, GPR160) and 12q13.2 (HOXC9, HOXC13, ERBB3, MUC11) in both OSCC and leukoplakia. In addition, we observed deletions of multiple regions on 3p (3p26.3, 3p22.3, 3p21.1, 3p14.2, 3p11.1) and 8p (8p23.2, 8p23.1, 8p22, 8p11.22) with high frequency (>52% in OSCCs), in line with literature reports in oral pre-invasive lesions [35,36]. These alterations can therefore be considered as important events associated with OSCC progression [36].

Figure 8. Summary of the predictive genomic and transcriptomic signatures associated with gingivobuccal cancer progression from pre-invasive lesions (leukoplakia) to cancer and lymph node metastasis. (*) represents alterations associated with disease specific survival in OSCC patients. Amp: Amplification (red); Del: Deletion (blue); Up: genes up-regulated (red), Down: genes down-regulated (blue).
leukoplakia and OSCC. Amplifications at 9p24.1 (CD274), 11q13.3 (AN01), and 7p11.2 (EGFR) were only identified in OSCCs, indicating their role in disease advancement, rather than their appearance at pre-invasive stages. Additionally, CD274 and its ligand PD1 are important targets of immunotherapy in various cancers, including OSCC [37–40].

A second hotspot for CNA-dependent gene over-expression was observed on 3q26.31, with ECT2 and the oncogene EIF5A2 over-expressed. Overexpression of EIF5A2 has not been previously reported in leukoplakia or OSCC, even though it has been proposed as a prognosis biomarker and potential therapeutic target for various other human tumors [41–44]. ECT2 has been previously found to be overexpressed in oral cancers [22], and also be involved in metastasis and angiogenesis of solid tumors [43,45–47].

A third interval of interest for amplifications and gene overexpression is 12q13.2, comprising of HOXC9 and HOXC13, genes associated with disease progression in OSCC. The HOX transcriptional regulators family is involved in pattern formation and organogenesis during embryo development [48] and potentially in the maintenance and regulation of cancer stem cells [49]. In particular, HOXC9 has been linked with cell cycle exit and cell invasion in breast cancer and neuroblastoma [48,50–52], and HOXC13 plays an important role in maintaining skin homeostasis and in regulating the transcription of cytokeratins genes [53,54]. Kasiri et al. [55] and Cantile et al. [56] showed that HOXC13 is a key player in tumor cell growth and viability in various human cancers.

Pathare et al. [57] and Bhattacharya et al. [58] demonstrated that specific CNAs are associated with lymph node metastasis. Here, the most frequent such alteration, specific to the lymph node metastatic tumors, was the amplified region 8q24.21 (57%), which includes the gene MYC, whose over-expression is postulated to activate various hallmarks of cancer, such as metastasis, invasion, and therapy resistance [20,59]. A highly recurrent deletion identified was 3p26.3 (57%), including the gene CHLI, alteration previously reported as a predictor of survival and lymph node metastasis in OSCC, along with loss of 3p14.2 (FHIT) [21,60]. The deletion of 8p23.2 was the most frequent in our study (68%). Genes in this region have been reported to be involved in lung, head and neck, breast, and skin cancers [61], but further studies are required to delineate its functional role in OSCC progression.

We separately analyzed early-stage and advanced-stage OSCCs to identify distinguishable CNAs with respect to recurrence and survival. For the first time, we report a recurrent amplification on 1p36.33 as significantly associated with clinical outcomes. Literature evidence supports that genes located on 1p, including JUN (1p32–31), TP73 (1p36.3), CASP9 (1p36.21), and NCOA5 (1p31.2), are important in the initiation and progression of several cancer types [62,63]. Genes of interest on the 1p36.33 amplicon include MXRA8 and DVL1. Here, we report the copy number dependent up-regulation of MXRA8, previously shown to function in tumor stroma by aiding the recovery of angiogenesis in capillaries [64,65]. DVL1 belongs to the Wnt signaling pathway known to be involved in growth, progression, and metastasis of various cancer types [66].

Additionally, we report copy number independent up-regulation of INHBA, MFAP5 and NELL2 in both leukoplakia and OSCC samples. MFAP5 is a secretory stromal protein overexpressed in leukoplakia and OSCC, possibly playing a role in malignant transformation and as a potential serum biomarker of cancer progression. Reports on ovarian cancer suggest that MFAP5 promotes tumor cell survival and angiogenesis through αvβ3 integrin-mediated signaling [67–69]. We identified few genes in copy number-altered regions that had a significant expression change in the direction opposite to what would be expected (e.g., a down-regulated gene in an amplified region), possibly following epigenetic regulation. One example is the down-regulation of DERL3, located at the 22q11.23 amplicon. Further studies are needed to confirm the significance of the DERL3 in oral tumorigenesis and to understand its gene regulation.

In sum, our study identifies CNAs and gene expression changes related to oral cancer progression. Alterations shared between leukoplakia and OSCC can be considered as important early events that are essential for initial cell transformation and progression. Integrative analysis of CNA and gene expression allows us to identify various novel drivers in oral cancer pathogenesis.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.tranon.2017.03.008.

Conflict of Interest
The authors affirm that they have no conflict of interest.

Grant Support
The work was supported by Grants from the Council of Scientific and Industrial Research (CSIR Scheme No. 27(0271)/12/EMR-II); Department of Biotechnology (DBT – BT/PR3317/MED/12/524/2011) and Annual Scientific Funds (ASF), ACTREC, Tata Memorial Centre. This research was supported in part by the Intramural Research Program of the National Institutes of Health, NLM, and by the Swiss National Science Foundation (Sinergia project 136,247).

Author’s Contribution
Conceived and designed the experiments: PGB, MBM. Performed the experiments: PGB, SA. Analyzed the data: PGB, SC, AAS, NB, MBM. Contributed reagents/materials/analysis tools: MBM, NB, AAS, SK, RSD. Wrote the paper: PGB, SC, AAS, NB, MBM. Assessment of clinical annotation, histopathological evaluation, and IHC grading: AMB, AP, RK.

Acknowledgments
The authors thank all participants of the study. ICMR National Tumor Tissue Repository, Tata Memorial Hospital; ACTREC Biorepository and Department of Pathology, Tata Memorial Centre is acknowledged for providing tumor tissue samples. We also thank Mr. P. Chavan and Mr. V. Sakpal from Histology Section, ACTREC, Tata Memorial Centre for their help. We thank Ms. Anjali Manaktala and Ms. Mayuri Inchanalkar for their help in real time PCRs.

References

[1] Lee J, Hong WK, Hittelman WN, Mao L, Lotan R, Shin DM, Benner SE, Xu XC, Lee JS, and Papadimitrakopoulos VM, et al (2000). Predicting cancer development in oral leukoplakia: ten years of translational research. Clin Cancer Res 6(5), 1702–1710.

[2] Amagasa T, Yamashiro M, and Uzawa N (2011). Oral premalignant lesions: from a clinical perspective. Int J Clin Oncol 16(1), 5–14.

[3] Lind PO (1987). Malignant transformation in oral leukoplakia. Scand J Dent Res 95(6), 449–455.

[4] Speight PM (2007). Update on oral epithelial dysplasia and progression to cancer. Head Neck Pathol 1(1), 61–66.

[5] Rivera C (2015). Essentials of oral cancer. Int J Clin Exp Pathol 8(9), 11884–11894.

[6] Neville BW and Day TA (2002). Oral cancer and precancerous lesions. CA Cancer J Clin 52(4), 195–215.

[7] Bouquot JE, Weiland LH, and Kurland LT (1988). Leukoplakia and carcinoma in situ synchronously associated with invasive oral/oropharyngeal carcinoma in Rochester, Minn., 1935–1984. Oral Surg Oral Med Oral Pathol 65(2), 199–207.
[50] Wang X, Choi JH, Ding J, Yang L, Ngoka LC, Lee EJ, Zha Y, Mao L, Jin B, and Ren M, et al (2013). HOX9 directly regulates distinct sets of genes to coordinate diverse cellular processes during neuronal differentiation. *BMC Genomics* 14, 830.

[51] Mao L, Ding J, Zha Y, Yang L, McCarthy BA, King W, Cui H, and Ding HF (2011). HOX9 links cell-cycle exit and neuronal differentiation and is a prognostic marker in neuroblastoma. *Cancer Res* 71(12), 4314–4324.

[52] Hassan NM, Hamada J, Murai T, Seino A, Takahashi Y, Tada M, Zhang X, Kashiwazaki H, Yamaraki Y, and Inoue N, et al (2006). Aberrant expression of HOX genes in oral dysplasia and squamous cell carcinoma tissues. *Onco Res* 16(5), 217–224.

[53] Joo YH, Park SW, Jung SH, Lee YS, Nam IC, Cho KJ, Park JO, Chung YJ, and Kim MS (2013). Recurrent loss of the FHIT gene and its impact on lymphatic metastasis in early oral squamous cell carcinoma. *Acta Otolaryngol* 133(9), 992–999.

[54] Kasiri S, Ansari KI, Hussain I, Bhan A, and Mandal SS (2013). Antisense oligonucleotide mediated knockdown of HOXC15 affects cell growth and induces apoptosis in tumor cells and over expression of HOXC13 induces 3D-colony formation. *RSC Adv* 3(10), 3260–3269.

[55] Cantile M, Scognamiglio G, Anniciello A, Farina M, Gentilcore G, Santonastaso C, Fulciniti F, Cillo C, Franco R, and Ascierto PA, et al (2012). Increased HOX C13 expression in metastatic melanoma progression. *J Biol Chem* 277(5), 3718–3726.

[56] Awgulewitsch A (2003). Hox in hair growth and development. *Naturwissenschaften* 90(5), 193–211.

[57] Kasiri S, Ansari KI, Hussain I, Bhan A, and Mandal SS (2013). Antisense oligonucleotide mediated knockdown of HOXC15 affects cell growth and induces apoptosis in tumor cells and over expression of HOXC13 induces 3D-colony formation. *RSC Adv* 3(10), 3260–3269.