Translational Oncology
www.transonc.com

Volume 10 Number xx Month 2017 pp. 396–409 396

Chromosomal Alterations and Gene Expression Changes Associated with the Progression of Leukoplakia to Advanced Gingivobuccal Cancer

Priyanka G. Bhosale[*,†,1], Simona Cristea[‡,§,1], 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.

Translational Oncology (2017) 10, 396–409

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

Address all correspondence to: 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.
E-mail: manojbmahimkar@gmail.com
[1]These authors contributed equally to this work.
Received 1 February 2017; Revised 21 March 2017; Accepted 21 March 2017

1936-5233
http://dx.doi.org/10.1016/j.tranon.2017.03.008

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

Table 1. The Clinicopathologic and Demographic Characteristic of the Study Patients.

Patient Characteristics	Total Study Samples	aCGH & GE Study (n = 121)#		qRT-PCR (n = 207)#		IHC & FISH (n = 370)#	
	Normal, Leukoplakia and OSCC	Leukoplakia	OSCC	Leukoplakia	OSCC	Leukoplakia	OSCC
	n = 481	aCGH: n = 24 GE: n = 15	aCGH: n = 91 GE: n = 61	n = 37	n = 138	n = 108	n = 185
Age at diagnosis							
Median age	49	42	50	41	50	44	50
Range (IQR)*	40–59	38–50	43–61	33–53	42–59	32–57	42–60
Gender							
Male	299 (76.3%)	21 (87.5%)	70 (76.9%)	33 (89.2%)	102 (75.6%)	96 (90.6%)	140 (75.7%)
Female	93 (23.7%)	3 (12.5%)	21 (23.1%)	4 (10.8%)	33 (24.4%)	10 (9.4%)	45 (24.3%)
Pathological stage							
Stage 1 and 2 (Early stage OSCC)	82 (35.5%)	NA	32 (35.2%)	NA	56 (41.5%)	NA	67 (36.2%)
Stage 3 and 4 (Advanced stage OSCC)	149 (64.5%)	NA	59 (64.8%)	NA	79 (58.5%)	NA	118 (63.8%)
Pathological T classification							
T1	31 (13.4%)	NA	7 (7.7%)	NA	25 (18.5%)	NA	24 (13%)
T2	100 (43.3%)	NA	40 (44%)	NA	64 (47.4%)	NA	80 (43.2%)
T3	10 (4.3%)	NA	4 (4.4%)	NA	4 (3%)	NA	8 (4.3%)
T4	90 (39%)	NA	40 (44%)	NA	42 (31.1%)	NA	73 (39.5%)
Pathological cervical lymph node involvement							
Node negative (N0)	133 (57.6%)	NA	55 (60.4%)	NA	79 (58.5%)	NA	112 (60.5%)
Node positive (N+)	98 (42.4%)	NA	36 (39.6%)	NA	56 (41.5%)	NA	73 (39.5%)
Pathological grade							
Well	27 (7.9%)	NA	8 (8.8%)	NA	12 (8.9%)	NA	23 (12.5%)
Moderate	139 (40.9%)	NA	55 (60.4%)	NA	87 (64.4%)	NA	106 (57.6%)
Poor	64 (18.8%)	NA	28 (30.8%)	NA	36 (26.7%)	NA	55 (29.9%)
Hyperplasia	89 (26.2%)	21 (87.5%)	NA	31 (86.1%)	NA	80 (80.8%)	NA
Mild dysplasia	11 (3.2%)	3 (12.5%)	NA	3 (8.3%)	NA	9 (9.1%)	NA
Moderate dysplasia	8 (2.4%)	NA	NA	2 (5.6%)	NA	8 (8.1%)	NA
Severe Dysplasia	2 (0.6%)	NA	NA	NA	NA	2 (2%)	NA
Habit profile							
No Habit	9 (3.1%)	9 (45%)	NA	NA	3 (2.6%)	NA	8 (5.3%)
Exclusive tobacco users	157 (54.5%)	3 (15%)	63 (77.8%)	13 (41.9%)	79 (70%)	30 (33%)	98 (64.5%)
Exclusive smoker	18 (6.3%)	NA	2 (2.5%)	5 (16.2%)	2 (2%)	12 (13.2%)	4 (2.6%)
Exclusive alcohol drinker	1 (0.3%)	NA	NA	NA	1 (0.8%)	NA	1 (0.7%)
Mixed habit**	103 (35.8%)	8 (40%)	16 (19.8%)	13 (41.9%)	28 (24.7%)	49 (53.8%)	41 (27%)

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 size; 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.

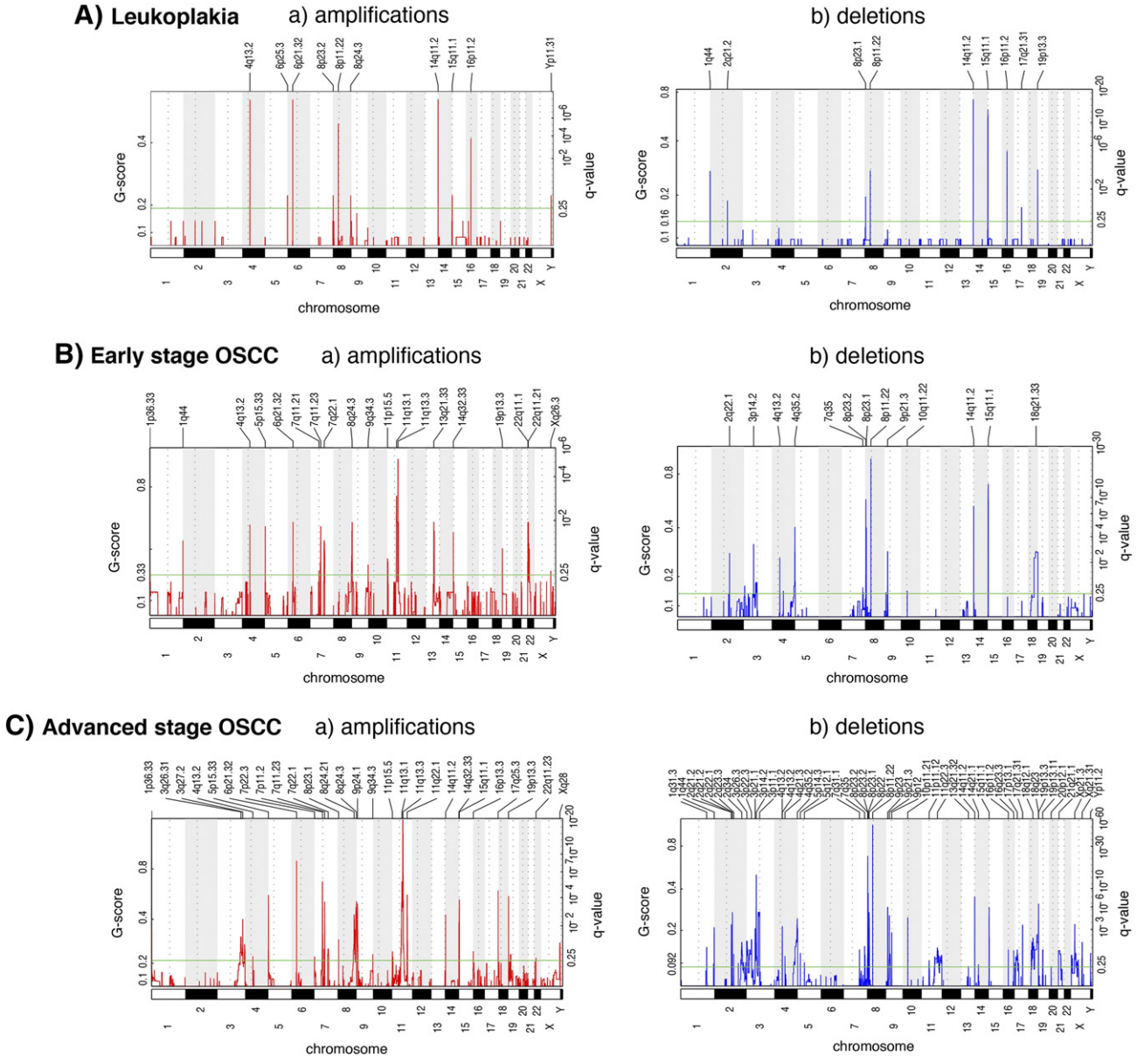

Figure 1. All amplifications (a) and deletions (b) inferred by GISTIC 2.0 in leukoplakia (A), early-stage OSCC (B), and advanced-stage OSCC (C) samples. Each alteration is assigned a G-score (left axis) by GISTIC 2.0, which considers the amplitude of the alteration, as well as its occurrence across samples.

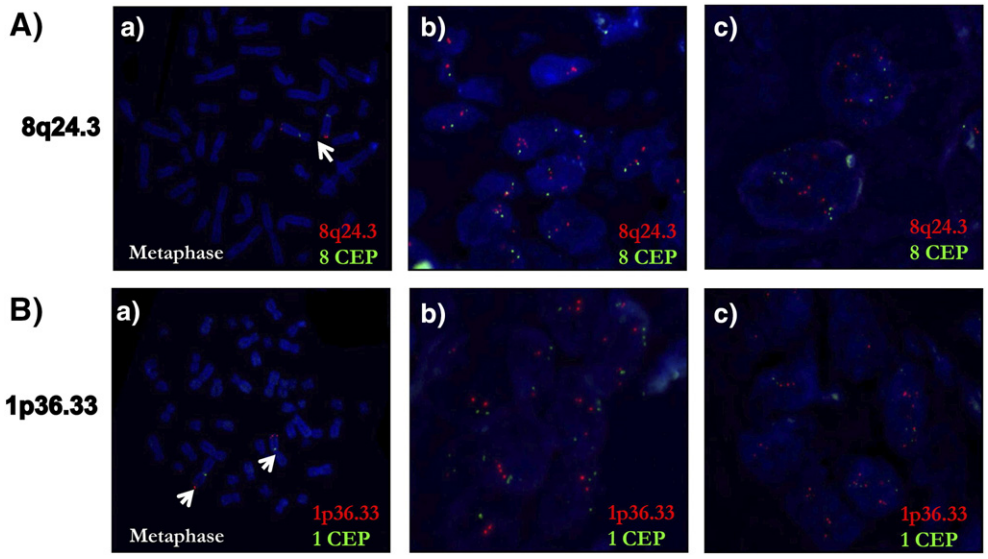

Figure 2. Nuc ish (FISH) for validating the amplifications of the regions 8q24.3 and 1p36.33. A) nuc ish for 8q24.3 locus in leukoplakia (b) and OSCC (c). B) nuc ish for 1p36.33 locus with weak amplification (b) and strong amplification (c) in OSCC. The specificities of the 8q24.3 locus probe (red)/chr 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.

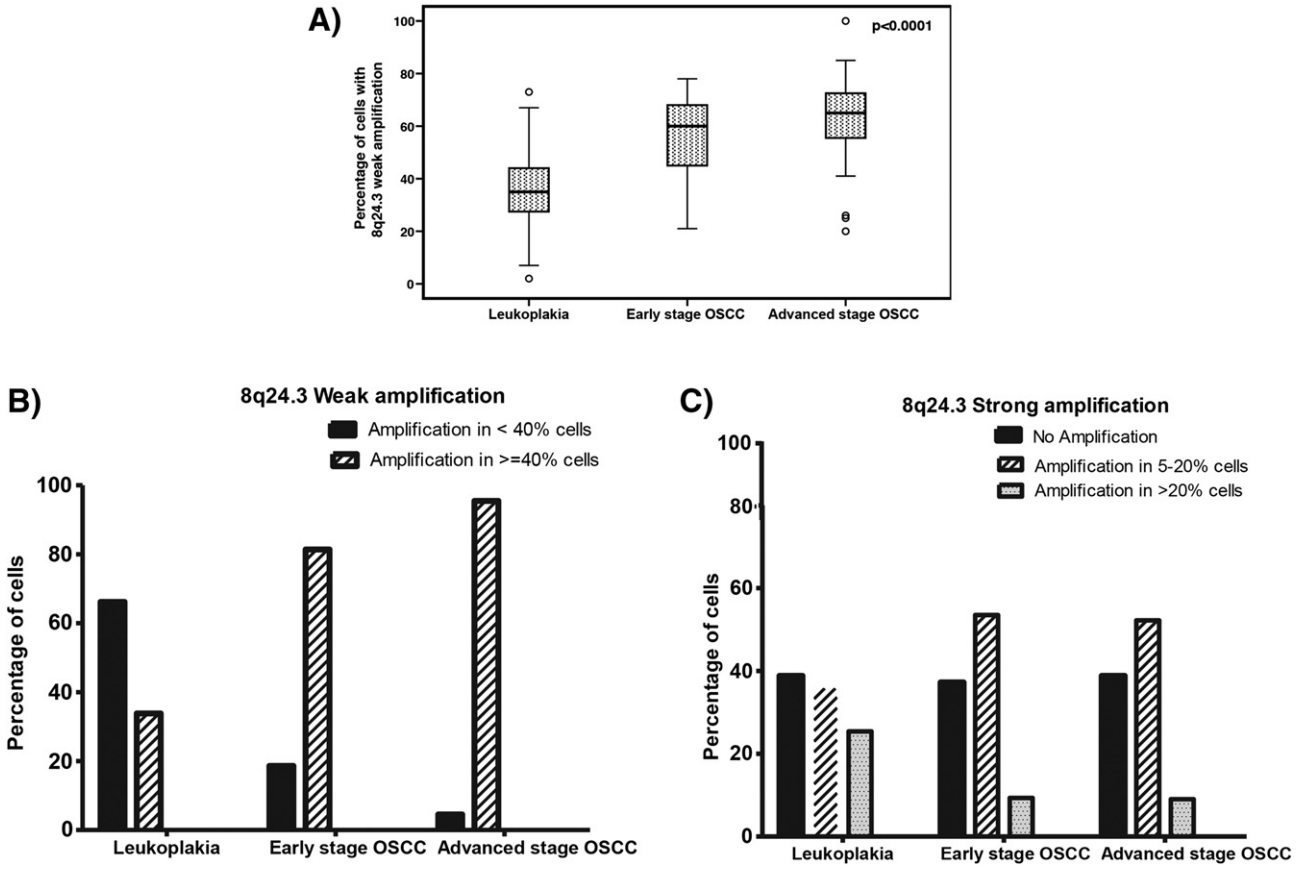

Figure 3. Correlation between 8q24.3 amplification and oral cancer progression. A) Percentage of cells with 8q24.3 weak amplification across different groups, correlated with disease progression (*P*-value calculated using Spearman correlation). B) and C) The increased percentage of tumor cells with 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 OSCCs (Table S6). The 10 most frequently amplified regions were 11q13.1 (70% of all patients), 8q24.3 (69%), 11p15.5 (60%), 1p36.33 (59%), 9q34.3 (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 in 3q26-qter, 8q24-qter, and 11q13.2-q13.4 to be common in high grade dysplasia, and amplifications in 3q, 7p, 8q, 11q, together with deletions in 3p and 8p to be most common in OSCCs [19].

CNAs associated with disease progression. Alterations associated with disease progression were identified by comparison between the following groups: 1) leukoplakia vs. early-stage OSCC, 2) leukoplakia vs. advanced-stage OSCC, and 3) early-stage OSCC vs. advanced-stage OSCC (Figure S2). Three amplified regions (4q13.2, 6p21.32, 8q24.3) and four deleted regions (8p23.1, 8p11.22, 14q11.2, 15q11.1) are common among the leukoplakia, early-stage OSCC, and advanced-stage OSCC samples (Figures 1 and S2), suggesting their involvement in disease progression. Interestingly, the amplification of 8q24.3 (harboring the candidate genes *FBXL6, GPR172A, PTP4A3*) was found in almost 60% of the analyzed leukoplakia cases and also in almost 70% of the tumors (Table S6).

We validated the 8q24.3 amplification by nuc ish using a disjoint validation set of 108 leukoplakia and 185 OSCC samples (Figure 2*A* and Table 1). The 8q24.3 locus-specific probe and a centromere 8 probe hybridized to their target loci and showed no cross reactivity (Figure 2*A*). The percentage of cells with 8q24.3 amplification increased as disease progressed from leukoplakia to OSCC, and the amplification was associated with disease progression (P <0.001) (Figure 3*A*). Almost 95% of the advanced-stage OSCC samples used for validation had a weak amplification of 8q24.3 in more than 40% of the tumor cells (Figure 3*B*). Moreover, this amplification was significantly associated with tumor grade (P =.046) and lymph node metastasis (P =. 0.007). The strong amplification of 8q24.3 in leukoplakia and tumors was only present in 5–20% of cells (Figure 3*C*). Our validation results confirm that 8q24.3 amplification is an important early event associated with OSCC progression.

CNAs associated with clinical outcome. Chromosomal alterations were also analyzed for their associations with clinicopathologic parameters, including nodal status, grade, and survival. We identified 64 recurrent chromosomal alterations in lymph node metastasis-negative OSCCs and 46 recurrent alterations in lymph node metastasis-positive OSCCs (q-value <0.25, Figure S3). In addition, we found 13 alterations unique to the primary tumors that metastasized to the nodes (Figure S2C and D), including amplifications of 8p23.1, 8q24.21 (harboring the candidate gene *MYC*), 11q22.1 (*MMPs, BIRC2, BIRC3*), and deletions of 2q23.3, 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 (q-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.

Cytoband	Alteration	Disease Specific Survival		Recurrence-Free Survival	
		BH Corrected *P*-Value	CPH Coef.	BH Corrected *P*-Value	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	0.0655	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	0.7098	0.2113	0.4591
9p23	Deletion	0.0969	0.711	0.1339	0.5327
11q22.3	Deletion	0.0829	0.7902	0.099	0.6467
6p21.1	Amplification	0.2198	0.4606	-	-
11q22.1	Amplification	0.1543	0.5204	-	-
18p11.31	Amplification	0.2227	0.5128	-	-
22q11.21	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	-	-	0.1986	–0.4272
9p21.3	Deletion	-	-	0.2007	0.4494
14q11.2	Deletion	-	-	0.1116	0.5358
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 4*A*. We validated the amplification of 1p36.33 by nuc ish (Figure 2*B*), and we confirmed that the amplification of 1p36.33 is associated with poor survival in an independent OSCC cohort (Figure 4*B*). 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

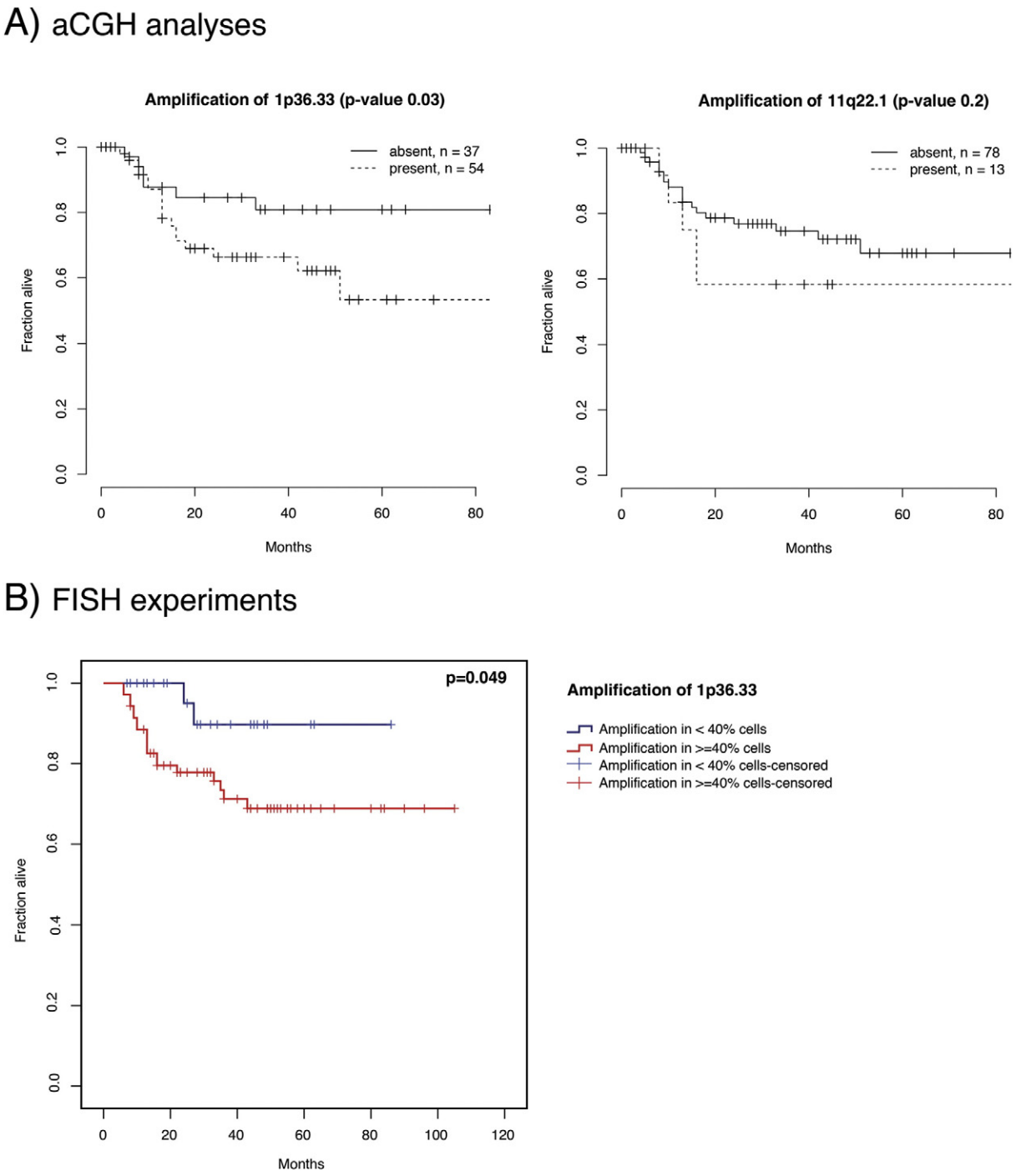

Figure 4. Kaplan–Meier plots of disease specific patient survival for selected chromosomal alterations based on aCGH analysis (A, 1p36.33 and 11q22.1) and nuc ish validation experiments (B, 1p36.33). Survival in months (X axis) is plotted against the fraction of patients alive (Y axis).

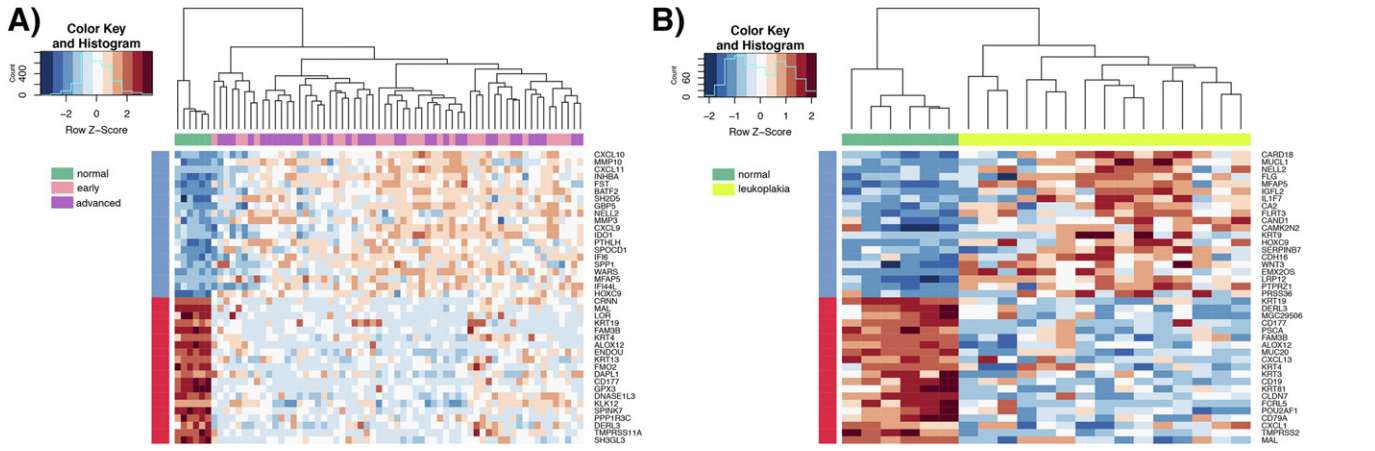

Figure 5. Heatmap of the top 20 differentially expressed genes in leukoplakia and OSCC. A) Differentially expressed genes in early-stage OSCC (pink) and advanced-stage OSCC (magenta) when compared to normal (green). B) Differentially expressed genes in leukoplakia (chartreuse) when compared to normal (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, GOLIM4, KLHL6, CLDN11, EIF4G1, ECT2*, GPR160, EIF5A2*, AP2M1*	*CHRD, ECT2*, EIF5A2*, KLHL6, NCEH1, GPR160*
4q13.2	Amplification	0.00112		
6p21.1	Amplification	0.11293	*YIPF3, NFKBIE*	*MDFI, VEGFA, C6orf132, PTK7, YIPF3*
7p11.2	Amplification	3.31E–08		*EGFR*
7q11.23	Amplification	3.66E–07		*SRRM3, SRCRB4D*
7q22.1	Amplification	0.0012685	*PCOLCE, GNB2*	*PCOLCE*
8q24.21	Amplification	0.0019006		*FAM49B*
8q24.3*	Amplification	6.67E–07		*FBXL6, GPR172A, ADCK5*
9p24.1	Amplification	5.90E–05		*CD274*
9q34.3	Amplification	0.01951		
11p15.5	Amplification	0.002778	*CHID1*	*SLC25A22, TMEM80, LSP1*
11p11.2	Amplification	0.043015	*PACSIN3, ARFGAP2*	
11q13.1	Amplification	9.71E–10	*RNASEH2C*	
11q13.3	Amplification	1.09E–27		*ANO1*
12q13.2	Amplification	0.17411	*MUCL1, HOXC9*, HOXC13*, ERBB3, ZNF385A, HOXC10, IKZF4*	*HOXC9*, MUCL1, HOXC13*, ERBB3, ITGA7, ITGA5, ZNF385A*
14q11.2	Amplification	0.00033729		*SALL2*
14q32.33	Amplification	2.63E–05	*INF2*	
16p11.2	Amplification	0.14231	*CD19, PRSS36, IL27, SPN, SEZ6L2, FUS, CORO1A*	*CD19, GDPD3, NUPR1, SPN, YPEL3, PRRT2*
17p13.1	Amplification	0.17411	*CLDN7, CD68, TMEM88, PHF23*	*CLDN7, CLEC10A, ATP1B2, EFNB3, TMEM88*
19p13.3	Amplification	4.12E–07		*PPAP2C*
20q11.22	Amplification	0.19131	*SPAG4, ERGIC3*	*SPAG4*
22q11.21	Amplification	0.0012607		*CLDN5, CDC45*
22q11.23	Amplification	0.0049669	*DERL3*, C22orf43*	*DERL3*, MMP11, C22orf43*
Xq28	Amplification	0.0063493	*SSR4, ATP6AP1, IRAK1*	*SSR4, FLNA, ARHGAP4*
1q44	Deletion	2.09E–05		*CNST*
2q34	Deletion	0.00096669	*SPAG16*	*IKZF2, SPAG16*
3p26.3	Deletion	0.0013306	*BHLHE40, C3orf32*	*CRBN, CAV3*
3p21.1	Deletion	1.13E–05		*CACNA2D3, SELK*
3p14.2	Deletion	4.30E–20		
4q13.2	Deletion	3.84E–07		
9p21.3	Deletion	5.36E–11		
10p11.21	Deletion	1.57E–05		*PARD3*
11q22.3	Deletion	0.042974	*CARD18, POU2AF1, CTSC, ST3GAL4, THY1, TRIM29, EI24, AMOTL1, PVRL1, ZC3H12C, CHEK1, DLAT*	*MMP10, MMP3, CRYAB, POU2AF1, EXPH5, FEZ1, CADM1, CTSC, TMEM25, VWA5A, CHEK1, HSPB2, PANX1, AMICA1, CARD17, ABCG4, ARHGAP32, UPK2, MPZL2, ETS1, FLI1, SCN4B, TRIM29, ME3, C11orf52, C11orf54, THY1, FZD4, OAF, PAFAH1B2, NLRX1, LOC283143, AMOTL1*
17p13.1	Deletion	0.030986	*ALOX12, CLDN7, VMO1, HS3ST3A1, CD68, GPR172B, SLC25A11, C17orf59, TMEM88, ATP2A3, UBE2G1, PHF23*	*ALOX12, SPNS2, CLDN7, XAF1, GPR172B, GAS7, ALOX12B, UBE2G1, CLEC10A, PMP22, USP43, ATP1B2, C17orf59, EFNB3, AURKB, GGT6, TMEM88, ATP2A3*
17q21.31	Deletion	2.56E–05		*ETV4*
21q21.3	Deletion	0.017484	*ADAMTS5*	*JAM2, NRIP1, CYYR1*

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.

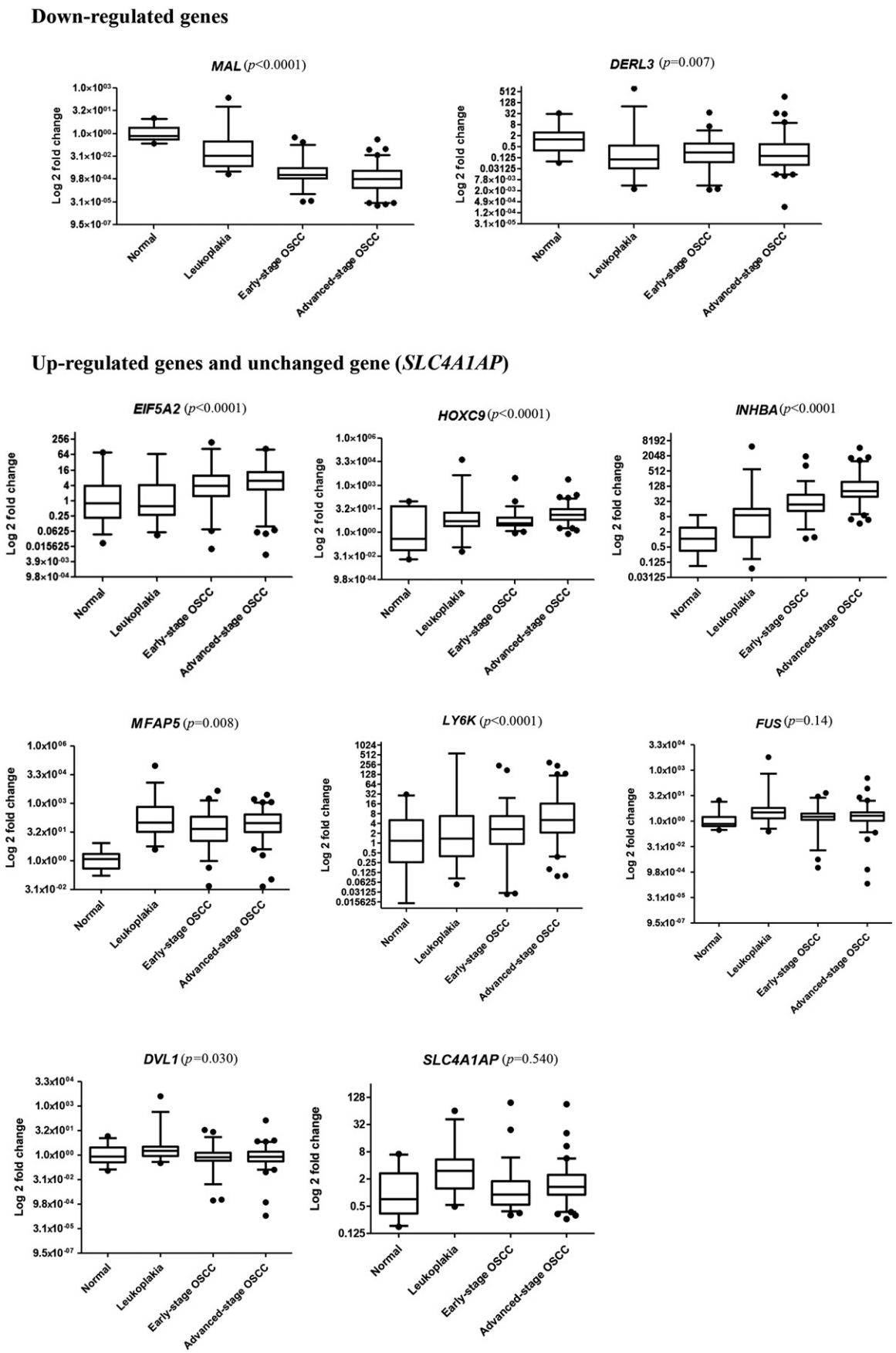

Figure 6. qRT-PCR expression changes of the 10 target genes (2 down-regulated, 7 up-regulated and 1 unchanged) in normal, leukoplakia and early and advanced-stage OSCC samples. The *P*-value (*P*) was calculated using the Spearman correlation test.

Table 4. Correlation of qRT-PCR Validation Targets with clinicopathological parameters.

Targets (Genes) →			*EIF5A2* ↑	*HOXC9* ↑	*INHBA* ↑	*MFAP5* ↑	*LY6K* ↑	*FUS* ↑	*DVL1* ↑	*SLC4A1AP* ↕	*DERL3* ↓	*MAL* ↓
Clinicopathological Parameters ↓		N	*P*-value (Coef*)	*P*-Value (Coef*)	*P*-Value (Coef*)	*P*-Value (Coef*)	*P*-Value (Coef*)	*P*-Value (Coef*)	*P*-Value (Coef*)	*P*-Value (Coef*)	*P*-Value (Coef*)	*P*-Value (Coef*)
Pre invasive stage	Normal Oral Mucosa	32	0.849 (–0.02)	**0.015 (0.31)**	**0.05 (0.39)**	**<0.001 (0.72)**	0.665 (0.58)	**<0.001 (0.46)**	**0.018 (0.28)**	**0.001 (0.43)**	**<0.001 (–0.48)**	**<0.001 (–0.51)**
	Leukoplakia	37										
Invasive OSCC	Normal Oral Mucosa	32	**<0.001 (0.31)**	**<0.001 (0.27)**	**<0.001 (0.45)**	**<0.001 (0.42)**	**0.003 (0.23)**	**<0.001 (0.27)**	0.620 (–0.03)	0.288 (0.08)	**<0.001 (–0.31)**	**<0.001 (–0.59)**
	OSCC	138										
OSCC progression	Normal Oral Mucosa	32	**<0.001 (0.39)**	**<0.001 (0.35)**	**<0.001 (0.68)**	**0.008 (0.19)**	**<0.001 (0.36)**	0.141 (0.10)	**0.030 (–0.15)**	0.540 (–0.04)	**0.007 (–0.19)**	**<0.001 (–0.68)**
	Leukoplakia	37										
	OSCC	138										
Tumor stage	Stage I/II	56	**0.044 (0.17)**	**<0.001 (0.42)**	**<0.001 (0.48)**	**0.041 (0.18)**	**0.001 (0.28)**	0.702 (0.03)	0.817 (0.09)	0.130 (0.21)	0.444 (–0.06)	**0.007 (–0.23)**
	Stage III/IV	79										
Tumor grade	Well	12	0.181 (0.12)	0.250 (–0.10)	0.062 (0.16)	0.620 (0.04)	0.500 (0.06)	0.815 (0.02)	0.144 (–0.12)	0.980 (–0.02)	0.418 (0.07)	0.629 (–0.04)
	Moderate	87										
	Poor	36										
Node metastasis	Node negative	79	**0.018 (0.20)**	**0.015 (0.21)**	**<0.001 (0.39)**	0.676 (0.03)	0.133 (0.13)	**0.029 (0.19)**	**0.014 (0.21)**	0.003 (0.25)	0.369 (0.08)	0.461 (–0.06)
	Node positive	56										

coef*: Spearman correlation coefficient. Bold font represents significant changes at a significance level of 0.05, ↑: up-regulated gene, ↓: down-regulated gene, ↕: unchanged gene; The association with OSCC progression was computed by comparing expression changes between leukoplakia, early stage OSCC and advance stage OSCC.

up-regulated and 1008 down-regulated) in early-stage OSCC, and 1924 (798 up-regulated and 1056 down-regulated) in advanced-stage OSCC (Figure S5). Up-regulation of *NELL2, MFAP5, CA2, FLRT3, HOXC9, CDH16, LRP12, PTPRZ1, TNNT1, WDR66, NEXN, EGR2, HOXC13, E2F7, ECT2, EIF5A2* and down-regulation of *KRT19, DERL3, CD177, PSCA, FAM3B, ALOX12, MUC20, CXCL13, KRT4, KRT3, CD19, KRT81, CLDN7, FCRL5, POU2AF1, CD79A, TMPRSS2, MAL, TNFRSF17, FCN1, PNOC, CXCL17, CEACAM1, FUT6, CLCA4, PITX1, DACT2, MEI1, GPX3* were observed 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, IFIT3, SERPINE1, GALNT6, FOXL2, PDPN, ITGA3, VEGFC, STAT1, LY6K, KLF7, 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, PTHLH, KRT4, KRT76* and *MAL*[10,22–26]. The novel oral cancer driving genes identified here include *DERL3, EIF5A2, HOXC9, HOXC13, MFAP5, NELL2, CD274, DHRS2, FST* and *GPX3*. 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 21q21.3 were identified as regions of copy number-dependent gene down-regulation (Table 3). In particular, 3q26.31 and 12q13.2, spanning the genes *ECT2, EIF5A2, HOXC9, HOXC13* and *MUCL1*, showed a strong correlation between copy number amplifications and gene over-expression. The deletion of 11q22.33 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 *DERL3* (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 *MFAP5*, 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, MFAP5,* 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 ($P<.0001$ and $P=.03$, 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

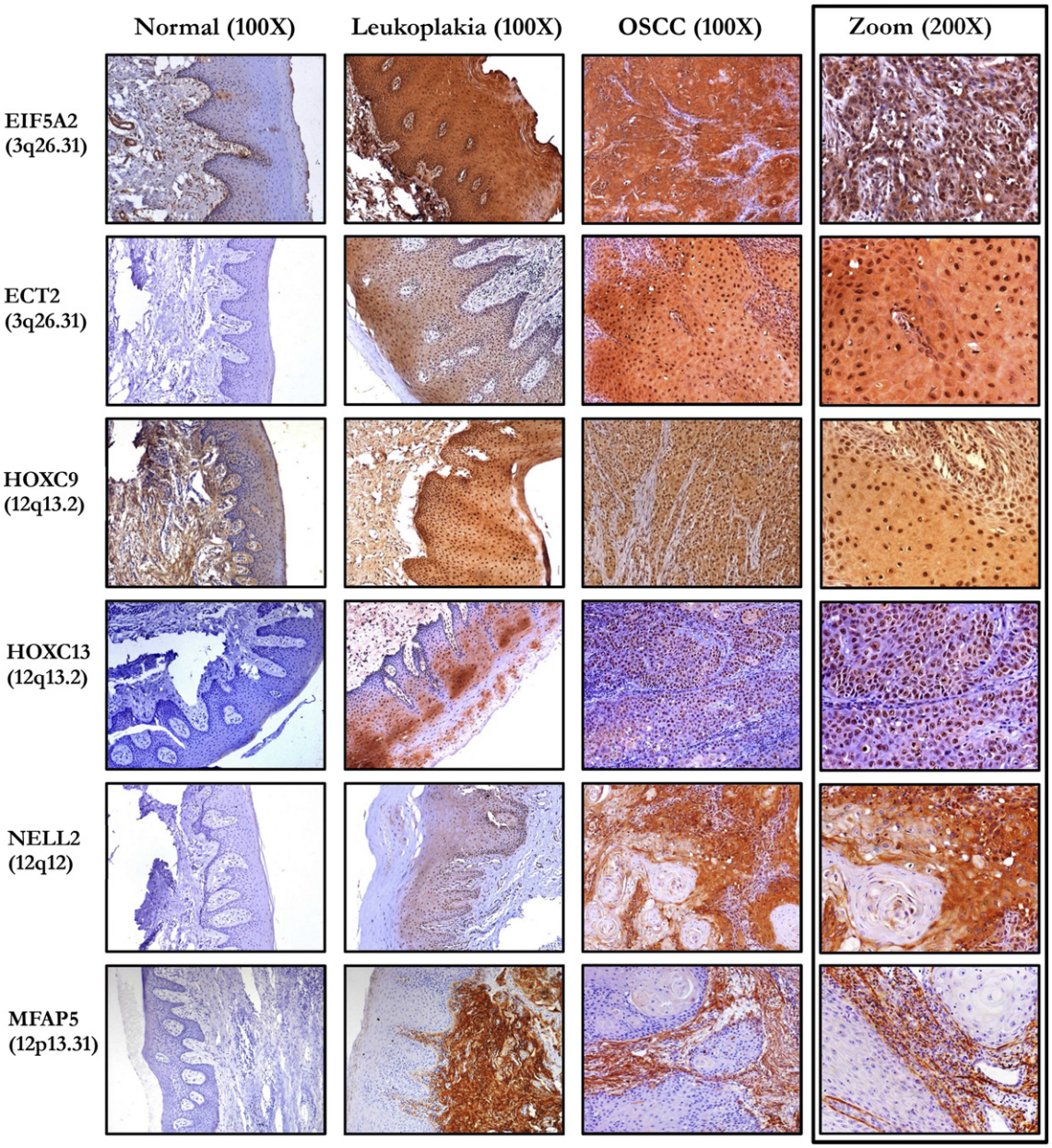

Figure 7. IHC analysis for target validation in normal, leukoplakia, and OSCC samples. NELL2 showed cytoplasmic staining, EIF5A2, ECT2, HOXC9, HOXC13 showed cytoplasmic and nuclear expression, while MFAP5 was localized predominantly into the matrix. Respective isotype controls for all the cases had no 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.

Table 5. Representative Biological Processes Significantly Altered in Leukoplakia and OSCC.

GO:BP:ID	*P*-Value	Odds ratio	Count	Size	Biological process
Leukoplakia					
GO:0044691	6.57E-05	433.0789474	2	3	tooth eruption
GO:0035116	0.000373976	24.34222222	3	30	embryonic hindlimb morphogenesis
GO:0021983	0.000879749	17.75243243	3	40	pituitary gland development
GO:0030199	0.000879749	17.75243243	3	40	collagen fibril organization
GO:0001568	0.001192229	3.794054527	9	556	blood vessel development
GO:0071230	0.001591348	14.27130435	3	49	cellular response to amino acid stimulus
GO:0043206	0.001656565	39.34688995	2	13	extracellular fibril organization
GO:0040012	0.002562823	11.97495573	3	62	regulation of locomotion
GO:0044259	0.002720644	7.421757892	4	123	multicellular organismal macromolecule metabolic process
GO:0009888	0.002854111	2.692398599	13	1260	tissue development
OSCC					
GO:0008544	1.03E-09	5.089295677	24	307	epidermis development
GO:2,000,145	1.90E-08	3.377582317	33	628	regulation of cell motility
GO:0051674	3.78E-07	2.447114154	47	1237	localization of cell
GO:0051272	4.34E-07	3.851913001	22	361	positive regulation of cellular component movement
GO:0040017	6.57E-07	3.750171556	22	370	positive regulation of locomotion
GO:0043207	1.47E-06	2.622472611	36	869	response to external biotic stimulus
GO:0060337	2.07E-06	7.805019305	10	84	type I interferon signaling pathway
GO:0034340	2.31E-06	7.70047619	10	85	response to type I interferon
GO:0018149	5.56E-06	9.573286052	8	56	peptide cross-linking
GO:0032496	8.14E-06	5.32031185	12	145	response to lipopolysaccharide

The top 10 dysregulated pathways (ordered by corrected *P*-value in Leukoplakia and OSCC), as identified by the Bioconductor package GOStats.

dysregulated processes. However, we identified a higher number of OSCC-related genes as part of dysregulated pathways as compared to genes related to pre-invasive lesions. Among them, we note genes associated with apoptotic processes (*FADD, BAX, BAK1, CASP7, INHBA, BIRC5*), developmental processes (*HOXC9, HOXC13, NELL2, CD274, ETS, STAT1*), biological regulation (*BATF, HOXC9, TERC*), and metabolic processes (*RAD51, HOXC13, E2F1, HOXB6, CDK1*). The top representative biological processes significantly altered in leukoplakia and OSCC are listed in Table 5.

Discussion

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, 4q21.3, 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. *PTK2, LY6K,* and *MYC* are prominent candidate oncogenes on 8q [10,32–34]. 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, KLHL6, GPR160*) and 12q13.2 (*HOXC9, HOXC13, ERBB3, MUCL1*) in both

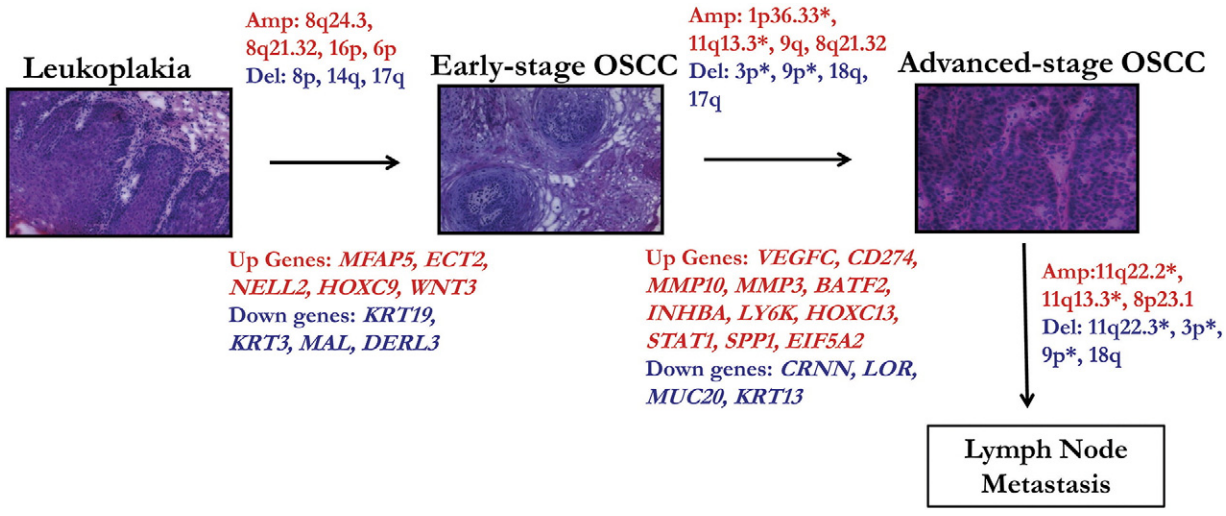

Figure 8. Summary of the predictive genomic and transcriptomic signatures associated with gingivobuccal cancer progression from pre-invasive lesions (leukoplakia) to cancer and lymph node metastasis. (*) represents alterations associated with disease specific survival in OSCC patients. Amp: Amplification (red); Del: Deletion (blue); Up: genes up-regulated (red), Down: genes down-regulated (blue).

leukoplakia and OSCC. Amplifications at 9p24.1 (*CD274*), 11q13.3 (*ANO1*), 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 *CHL1*, 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 *NRAS* (1p13.2), 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 $\alpha_5\beta_3$ 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 JJ, Hong WK, Hittelman WN, Mao L, Lotan R, Shin DM, Benner SE, Xu XC, Lee JS, and Papadimitrakopoulou 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.

[8] Westra WH and Sidransky D (2006). Fluorescence visualization in oral neoplasia: shedding light on an old problem. *Clin Cancer Res* **12**(22), 6594–6597.

[9] Ambatipudi S, Gerstung M, Gowda R, Pai P, Borges AM, Schaffer AA, Beerenwinkel N, and Mahimkar MB (2011). Genomic profiling of advanced-stage oral cancers reveals chromosome 11q alterations as markers of poor clinical outcome. *PLoS One* **6**(2), e17250.

[10] Ambatipudi S, Gerstung M, Pandey M, Samant T, Patil A, Kane S, Desai RS, Schaffer AA, Beerenwinkel N, and Mahimkar MB (2012). Genome-wide expression and copy number analysis identifies driver genes in gingivobuccal cancers. *Genes Chromosomes Cancer* **51**(2), 161–173.

[11] Tang YC and Amon A (2013). Gene copy-number alterations: a cost-benefit analysis. *Cell* **152**(3), 394–405.

[12] Costa JL, Meijer G, Ylstra B, and Caldas C (2008). Array comparative genomic hybridization copy number profiling: a new tool for translational research in solid malignancies. *Semin Radiat Oncol* **18**(2), 98–104.

[13] Estilo CL, O-charoenrat P, Talbot S, Socci ND, Carlson DL, Ghossein R, Williams T, Yonekawa Y, Ramanathan Y, and Boyle JO, et al (2009). Oral tongue cancer gene expression profiling: Identification of novel potential prognosticators by oligonucleotide microarray analysis. *BMC Cancer* **9**, 11.

[14] Boffetta P, Hecht S, Gray N, Gupta P, and Straif K (2008). Smokeless tobacco and cancer. *Lancet Oncol* **9**(7), 667–675.

[15] Michaud DS, Langevin SM, Eliot M, Nelson HH, Pawlita M, McClean MD, and Kelsey KT (2014). High-risk HPV types and head and neck cancer. *Int J Cancer* **135**(7), 1653–1661.

[16] Braakhuis BJ, Snijders PJ, Keune WJ, Meijer CJ, Ruijter-Schippers HJ, Leemans CR, and Brakenhoff RH (2004). Genetic patterns in head and neck cancers that contain or lack transcriptionally active human papillomavirus. *J Natl Cancer Inst* **96**(13), 998–1006.

[17] Klussmann JP, Mooren JJ, Lehnen M, Claessen SM, Stenner M, Huebbers CU, Weissenborn SJ, Wedemeyer I, Preuss SF, and Straetmans JM, et al (2009). Genetic signatures of HPV-related and unrelated oropharyngeal carcinoma and their prognostic implications. *Clin Cancer Res* **15**(5), 1779–1786.

[18] Bhosale PG, Pandey M, Desai RS, Patil A, Kane S, Prabhash K, and Mahimkar MB (2016). Low prevalence of transcriptionally active human papilloma virus in Indian patients with HNSCC and leukoplakia. *Oral Surg Oral Med Oral Pathol Oral Radiol* **122**(5), 609–618.

[19] Salahshourifar I, Vincent-Chong VK, Kallarakkal TG, and Zain RB (2014). Genomic DNA copy number alterations from precursor oral lesions to oral squamous cell carcinoma. *Oral Oncol* **50**(5), 404–412.

[20] Prochownik EV (2008). c-Myc: linking transformation and genomic instability. *Curr Mol Med* **8**(6), 446–458.

[21] Uchida K, Oga A, Nakao M, Mano T, Mihara M, Kawauchi S, Furuya T, Ueyama Y, and Sasaki K (2011). Loss of 3p26.3 is an independent prognostic factor in patients with oral squamous cell carcinoma. *Oncol Rep* **26**(2), 463–469.

[22] Iyoda M, Kasamatsu A, Ishigami T, Nakashima D, Endo-Sakamoto Y, Ogawara K, Shiiba M, Tanzawa H, and Uzawa K (2010). Epithelial cell transforming sequence 2 in human oral cancer. *PLoS One* **5**(11), e14082.

[23] Chung CH, Parker JS, Ely K, Carter J, Yi Y, Murphy BA, Ang KK, El-Naggar AK, Zanation AM, and Cmelak AJ, et al (2006). Gene expression profiles identify epithelial-to-mesenchymal transition and activation of nuclear factor-kappaB signaling as characteristics of a high-risk head and neck squamous cell carcinoma. *Cancer Res* **66**(16), 8210–8218.

[24] Mendez E, Houck JR, Doody DR, Fan W, Lohavanichbutr P, Rue TC, Yueh B, Futran ND, Upton MP, and Farwell DG, et al (2009). A genetic expression profile associated with oral cancer identifies a group of patients at high risk of poor survival. *Clin Cancer Res* **15**(4), 1353–1361.

[25] Chen C, Mendez E, Houck J, Fan W, Lohavanichbutr P, Doody D, Yueh B, Futran ND, Upton M, and Farwell DG, et al (2008). Gene expression profiling identifies genes predictive of oral squamous cell carcinoma. *Cancer Epidemiol Biomark Prev* **17**(8), 2152–2162.

[26] Ziober AF, Patel KR, Alawi F, Gimotty P, Weber RS, Feldman MM, Chalian AA, Weinstein GS, Hunt J, and Ziober BL (2006). Identification of a gene signature for rapid screening of oral squamous cell carcinoma. *Clin Cancer Res* **12**(20 Pt 1), 5960–5971.

[27] Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, and Narechania A (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* **13**(9), 2129–2141.

[28] Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, Guo N, Muruganujan A, Doremieux O, and Campbell MJ, et al (2005). The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res* **33**, D284–D288 [Database issue].

[29] Sugahara K, Michikawa Y, Ishikawa K, Shoji Y, Iwakawa M, Shibahara T, and Imai T (2011). Combination effects of distinct cores in 11q13 amplification region on cervical lymph node metastasis of oral squamous cell carcinoma. *Int J Oncol* **39**(4), 761–769.

[30] Liu CJ, Lin SC, Chen YJ, Chang KM, and Chang KW (2006). Array-comparative genomic hybridization to detect genomewide changes in microdissected primary and metastatic oral squamous cell carcinomas. *Mol Carcinog* **45**(10), 721–731.

[31] Gollin SM (2014). Cytogenetic alterations and their molecular genetic correlates in head and neck squamous cell carcinoma: a next generation window to the biology of disease. *Genes Chromosomes Cancer* **53**(12), 972–990.

[32] Skinner HD, Giri U, Yang L, Woo SH, Story MD, Pickering CR, Byers LA, Williams MD, El-Naggar A, and Wang J, et al (2016). Proteomic profiling identifies PTK2/FAK as a driver of radioresistance in HPV negative head and neck cancer. *Clin Cancer Res* **22**(18), 4643–4650

[33] Roy-Luzarraga M and Hodivala-Dilke K (2016). Molecular Pathways: Endothelial Cell FAK-A Target for Cancer Treatment. *Clin Cancer Res* **22**(15), 3718–3724.

[34] Matsuda R, Enokida H, Chiyomaru T, Kikkawa N, Sugimoto T, Kawakami K, Tatarano S, Yoshino H, Toki K, and Uchida Y, et al (2011). LY6K is a novel molecular target in bladder cancer on basis of integrate genome-wide profiling. *Br J Cancer* **104**(2), 376–386.

[35] Garnis C, Chari R, Buys TP, Zhang L, Ng RT, Rosin MP, and Lam WL (2009). Genomic imbalances in precancerous tissues signal oral cancer risk. *Mol Cancer* **8**, 50.

[36] Tsui IF, Rosin MP, Zhang L, Ng RT, and Lam WL (2008). Multiple aberrations of chromosome 3p detected in oral premalignant lesions. *Cancer Prev Res (Phila)* **1**(6), 424–429.

[37] Cho YA, Yoon HJ, Lee JI, Hong SP, and Hong SD (2011). Relationship between the expressions of PD-L1 and tumor-infiltrating lymphocytes in oral squamous cell carcinoma. *Oral Oncol* **47**(12), 1148–1153.

[38] Lyford-Pike S, Peng S, Young GD, Taube JM, Westra WH, Akpeng B, Bruno TC, Richmon JD, Wang H, and Bishop JA, et al (2013). Evidence for a role of the PD-1:PD-L1 pathway in immune resistance of HPV-associated head and neck squamous cell carcinoma. *Cancer Res* **73**(6), 1733–1741.

[39] Ritprajak P and Azuma M (2015). Intrinsic and extrinsic control of expression of the immunoregulatory molecule PD-L1 in epithelial cells and squamous cell carcinoma. *Oral Oncol* **51**(3), 221–228.

[40] Zandberg DP and Strome SE (2014). The role of the PD-L1:PD-1 pathway in squamous cell carcinoma of the head and neck. *Oral Oncol* **50**(7), 627–632.

[41] Huff LP, Decristo MJ, Trembath D, Kuan PF, Yim M, Liu J, Cook DR, Miller CR, Der CJ, and Cox AD (2013). The Role of Ect2 Nuclear RhoGEF Activity in Ovarian Cancer Cell Transformation. *Genes Cancer* **4**(11–12), 460–475.

[42] Guan XY, Sham JS, Tang TC, Fang Y, Huo KK, and Yang JM (2001). Isolation of a novel candidate oncogene within a frequently amplified region at 3q26 in ovarian cancer. *Cancer Res* **61**(9), 3806–3809.

[43] Li Y, Fu L, Li JB, Qin Y, Zeng TT, Zhou J, Zeng ZL, Chen J, Cao TT, and Ban X, et al (2014). Increased expression of EIF5A2, via hypoxia or gene amplification, contributes to metastasis and angiogenesis of esophageal squamous cell carcinoma. *Gastroenterology* **146**(7), 1701–1713 [e1709].

[44] Liu Y, Du F, Chen W, Yao M, Lv K, and Fu P (2015). EIF5A2 is a novel chemoresistance gene in breast cancer. *Breast Cancer* **22**(6), 602–607.

[45] Wang FW, Cai MY, Mai SJ, Chen JW, Bai HY, Li Y, Liao YJ, Li CP, Tian XP, and Kung HF, et al (2014). Ablation of EIF5A2 induces tumor vasculature remodeling and improves tumor response to chemotherapy via regulation of matrix metalloproteinase 2 expression. *Oncotarget* **5**(16), 6716–6733.

[46] Meng QB, Kang WM, Yu JC, Liu YQ, Ma ZQ, Zhou L, Cui QC, and Zhou WX (2015). Overexpression of eukaryotic translation initiation factor 5A2 (EIF5A2) correlates with cell aggressiveness and poor survival in gastric cancer. *PLoS One* **10**(3), e0119229.

[47] Wei JH, Cao JZ, Zhang D, Liao B, Zhong WM, Lu J, Zhao HW, Zhang JX, Tong ZT, and Fan S, et al (2014). EIF5A2 predicts outcome in localised invasive bladder cancer and promotes bladder cancer cell aggressiveness in vitro and in vivo. *Br J Cancer* **110**(7), 1767–1777.

[48] Hur H, Lee JY, Yang S, Kim JM, Park AE, and Kim MH (2016). HOXC9 Induces Phenotypic Switching between Proliferation and Invasion in Breast Cancer Cells. *J Cancer* **7**(7), 768–773.

[49] Bhatlekar S, Fields JZ, and Boman BM (2014). HOX genes and their role in the development of human cancers. *J Mol Med* **92**(8), 811–823.

[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). HOXC9 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). HOXC9 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, Yamazaki Y, and Inoue N, et al (2006). Aberrant expression of HOX genes in oral dysplasia and squamous cell carcinoma tissues. *Oncol Res* **16**(5), 217–224.

[53] Jave-Suarez LF, Winter H, Langbein L, Rogers MA, and Schweizer J (2002). HOXC13 is involved in the regulation of human hair keratin gene expression. *J Biol Chem* **277**(5), 3718–3726.

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

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

[56] 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 Transl Med* **10**, 91.

[57] Pathare S, Schaffer AA, Beerenwinkel N, and Mahimkar M (2009). Construction of oncogenetic tree models reveals multiple pathways of oral cancer progression. *Int J Cancer* **124**(12), 2864–2871.

[58] Bhattacharya A, Roy R, Snijders AM, Hamilton G, Paquette J, Tokuyasu T, Bengtsson H, Jordan RC, Olshen AB, and Pinkel D, et al (2011). Two distinct routes to oral cancer differing in genome instability and risk for cervical node metastasis. *Clin Cancer Res* **17**(22), 7024–7034.

[59] Hanahan D and Weinberg RA (2011). Hallmarks of cancer: the next generation. *Cell* **144**(5), 646–674.

[60] 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.

[61] Ma C, Quesnelle KM, Sparano A, Rao S, Park MS, Cohen MA, Wang Y, Samanta M, Kumar MS, and Aziz MU, et al (2009). Characterization CSMD1 in a large set of primary lung, head and neck, breast and skin cancer tissues. *Cancer Biol Ther* **8**(10), 907–916.

[62] Herath NI, Kew MC, Whitehall VL, Walsh MD, Jass JR, Khanna KK, Young J, Powell LW, Leggett BA, and Macdonald GA (2000). p73 is up-regulated in a subset of hepatocellular carcinomas. *Hepatology* **31**(3), 601–605.

[63] Ma Y, Li Q, Cui W, Miao N, Liu X, Zhang W, Zhang C, and Wang J (2012). Expression of c-Jun, p73, Casp9, and N-ras in thymic epithelial tumors: relationship with the current WHO classification systems. *Diagn Pathol* **7**, 120.

[64] Kiflemariam S, Ljungstrom V, Ponten F, and Sjoblom T (2015). Tumor vessel up-regulation of INSR revealed by single-cell expression analysis of the tyrosine kinome and phosphatome in human cancers. *Am J Pathol* **185**(6), 1600–1609.

[65] Yonezawa T, Ohtsuka A, Yoshitaka T, Hirano S, Nomoto H, Yamamoto K, and Ninomiya Y (2003). Limitrin, a novel immunoglobulin superfamily protein localized to glia limitans formed by astrocyte endfeet. *Glia* **44**(3), 190–204.

[66] Zhang K, Song H, Yang P, Dai X, Li Y, Wang L, Du J, Pan K, and Zhang T (2015). Silencing dishevelled-1 sensitizes paclitaxel-resistant human ovarian cancer cells via AKT/GSK-3beta/beta-catenin signalling. *Cell Prolif* **48**(2), 249–258.

[67] Spivey KA and Banyard J (2010). A prognostic gene signature in advanced ovarian cancer reveals a microfibril-associated protein (MAGP2) as a promoter of tumor cell survival and angiogenesis. *Cell Adhes Migr* **4**(2), 169–171.

[68] Milwid JM, Elman JS, Li M, Shen K, Manrai A, Gabow A, Yarmush J, Jiao Y, Fletcher A, and Lee J, et al (2014). Enriched protein screening of human bone marrow mesenchymal stromal cell secretions reveals MFAP5 and PENK as novel IL-10 modulators. *Mol Ther* **22**(5), 999–1007.

[69] Leung CS, Yeung TL, Yip KP, Pradeep S, Balasubramanian L, Liu J, Wong KK, Mangala LS, Armaiz-Pena GN, and Lopez-Berestein G, et al (2014). Calcium-dependent FAK/CREB/TNNC1 signalling mediates the effect of stromal MFAP5 on ovarian cancer metastatic potential. *Nat Commun* **5**, 5092.